BUSINESS FLUCTUATIONS

SECOND EDITION

GORDON

BUSINESS FLUCTUATIONS

SECOND EDITION

Robert Aaron Gordon
PROFESSOR OF ECONOMICS
UNIVERSITY OF CALIFORNIA,
BERKELEY

Harper & Brothers
PUBLISHERS, NEW YORK

1961
[1952]

CONTENTS

FIGURES

TABLES

PREFACE TO THE SECOND EDITION

A REVISION of the first edition of *Business Fluctuations* has been overdue for some time. In the field of economic instability and growth, neither the scientific literature nor events stand still. A great deal has happened in the decade since the first edition of this book was completed.

The present revision is a complete one. All data have been brought up to date; most chapters have been largely or completely rewritten; a new historical chapter has been inserted on the decade of the 1950's; another chapter has been added on the control of inflation; entirely new material has been included on the determinants of aggregate supply and on the way aggregate demand and supply interact to determine the level of output and prices. Reflecting the interest of the times, I have given a great deal more attention to problems of growth and inflation than I did in the first edition.

For what I believe is the first time anywhere, Chapter 10 presents for a large number of series separate average reference-cycle patterns for the interwar and for the postwar periods. It seems to me that these patterns give a good deal of concreteness to what have thus far been vague impressions (and in some cases misconceptions) as to the ways in which the postwar business cycle has and has not differed from what we experienced before World War II. A good deal of our recollection of prewar business cycles is colored by the 1930's. Hence, I have tried in most cases to make separate comparisons of prewar (more accurately, interwar) *minor* cycles and the postwar cycles. The data permitted on averaging of three postwar cycles, through 1958, all of which fall into the minor category.

I must record my great debt to the National Bureau of Economic

Research for making available to me, through their computer program, the original data and reference cycle relatives on which the patterns in Chapter 10 are based. It has seemed to me that the Bureau's reference-cycle analysis deserves to be more widely used than it has been. In addition to its value in empirical research, it has pedogogical virtues which, I hope, are illustrated by its use in Chapter 10 to compare prewar and postwar cyclical behavior.

My thanks are due to the National Bureau on another count. They have been most generous in permitting me to make use of material from various of their publications.

To repeat what was said in the first edition, this book is intended for a one-semester course in business fluctuations (or aggregative analysis) at the junior or senior level. It does not assume that the student has already had a course in the theory of income and employment. Instead, the essential tools of aggregative analysis are developed in Part I, where their limitations for business-cycle analysis are also examined, and these tools are then used in the later chapters dealing with economic fluctuations and growth.

It is possible that more material has been included than some instructors will want to cover in a one-semester course. Depending on personal preference and the availability of time, some of the topics dealt with here can be omitted or treated very briefly. I hope that the book will be useful both in economics departments and in schools of business administration that put a good deal of emphasis on economic analysis.

Bert G. Hickman and David Smith were good enough to read parts of the manuscript of the revised edition, and I benefited greatly from their comments. Robert J. Gordon helped in clarifying the exposition at a number of points. I should also like to acknowledge again the help of those who read parts of the original manuscript. Their names were given in the preface of the first edition. Finally, my thanks go to Brian Van Arkadie, whose unstinting assistance has been invaluable at all stages of this revision.

R. A. Gordon

January, 1961

PART I

INCOME, SPENDING, OUTPUT, AND THE PRICE LEVEL

CHAPTER 1

WHAT THIS BOOK IS ABOUT

AMERICAN economic history records a marvelous story of technological progress, of startling increases in productive capacity, and of a rising standard of living that has been the envy of other peoples. But this progress has come in fits and starts. Periods of rapid growth and increasing prosperity have alternated with periods of declining output, rising unemployment, falling profits, and general economic distress. The American economy has thus been characterized by both *growth* (more rapid during some decades than others) and by *instability;* and this has, on the whole, also been true of other industrialized nations.

This instability extends to the prices of the things we buy. The great increase in prices since the 1930's is one of the outstanding characteristics of recent economic history, not only in the United States but in every country. Prices tend to show the same kind of short-run fluctuations as does business generally, but they have also moved through longer periods of rise and fall. Today's college students have lived their entire lives during a period of secularly (i.e., long-run) rising prices, and the belief is widely held that the "creeping inflation" of the last decade or more will continue indefinitely.

Some aspects of this record of growth and instability are traced out in Figure 1. Total output approximately quintupled in the first half of the century, and it has risen further since 1950. The serious depressions of the century can also be identified on the chart, and one can note the interruptions to growth created by the more frequent minor setbacks. The wide swings in prices in the twentieth century stand out clearly—above all, the tremendous rise in the price level that we have experienced since the 1930's.

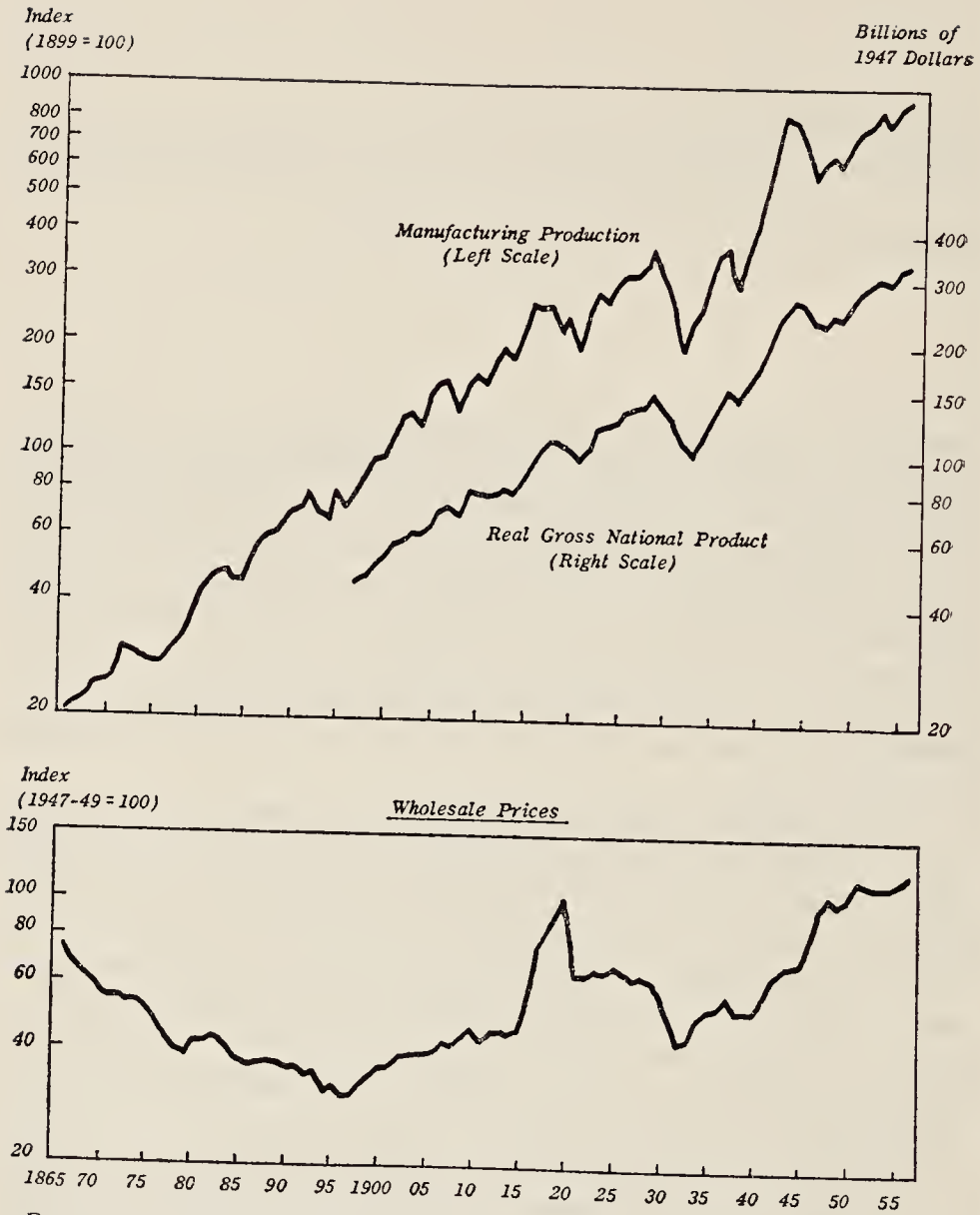


FIGURE 1. Manufacturing Production, Gross National Product, and Wholesale Prices, 1866–1956.

Reproduced with permission from B. G. Hickman, "Postwar Cyclical Experience and Economic Instability," *American Economic Review*, vol. 48, May, 1958, p. 122.

THE PROBLEM OF INSTABILITY

Since 1914, in about two generations, this country has experienced the catastrophic depression of the 1930's and two great, war-generated inflations. There have been other disturbances also—the violent deflation of 1920–1921 that was the aftermath of the inflationary boom following World War I, two mild recessions in the 1920's, the sharp decline of 1937–1938, and a series of mild but nonetheless disturbing business recessions since World War II.

Most of us have short memories, and the inflationary years since Pearl Harbor have helped to blur our impression of the 1930's. In mid-1929, there was little unemployment in the United States; production was setting new records; business generally was prosperous; the stock market was booming. Three years later, a quarter or more of our working force could not find jobs, and a substantial number of those working were on part time; industrial production had fallen by more than 50 percent; the national income had been cut in half; thousands of farm mortgages were being foreclosed; and our banking system was on the brink of complete collapse. The depression was world-wide, and the tensions it created helped to bring on the catastrophe of World War II.

It was a shock to all Americans that our powers of recovery after the low point of the Great Depression of the 1930's seemed to be so weak. Even at the peak of the boom in 1937 there were still seven or eight million men and women who could not find jobs. When Hitler marched his armies into Poland in 1939, there were still some nine million unemployed in the United States. It was not until after Pearl Harbor, when the productive machine was straining to meet the insatiable demands of total war, that the American economy was again able to provide jobs for all who wanted to work.

Since the beginning of World War II, inflation has succeeded depression as the kind of economic change immediately influencing our lives. Wholesale prices considerably more than doubled between 1940 and 1960, and the cost of living rose by a little more than 100 percent. While conditions have remained prosperous, on the whole, and output has expanded by 50 percent since 1946, growth has been uneven. We have continued to have business recessions, albeit mild ones. During most of the years since World War II, we have been able to sustain a high level of employment; but the economic ma-

chine still sputters and slows down more frequently than we like, and we have not yet learned how to cope with the marked price inflation that has beset the economy during the last 20 years. In short, economic instability is still with us.

THE STUDY OF BUSINESS FLUCTUATIONS

What makes the economy unstable, what determines how fast and how steadily it grows, and what can be done to make it more stable? Those are the questions with which this book is concerned. It deals with business (or economic) fluctuations and with economic growth.

The study of the alternating rise and fall of economic activity has been going on for many years. It is a difficult subject—one of the most difficult in economics—and it cannot be said that we yet fully know why booms and depressions occur. But we have learned a great deal, especially in recent years, and the chapters that follow will attempt to summarize what we do know about the causes of business fluctuations.

Although our knowledge of causes is imperfect, most economists agree that we know enough so that, with wise management, we can keep the American economy much more stable than it has been in the past. We are all agreed that we must never again have a Great Depression and that even less overwhelming catastrophes can and must be avoided. Virtually all governments of the world have put the maintenance of a high and stable level of employment close to the top of their list of domestic objectives. The American Employment Act of 1946 stated that "it is the continuing policy and responsibility of the Federal Government to use all practicable means . . . to promote maximum employment, production, and purchasing power." Similarly, under the United Nations Charter, the subscribing governments pledged themselves to take action to promote "higher standards of living, full employment, and conditions of economic and social progress and development."[1] In the later chapters of this book we shall examine the measures that are available to secure the largest possible degree of economic stability and see to what extent these measures are being used in this and other countries.

[1] Articles 55 and 56 of the United Nations Charter.

IS THE BUSINESS CYCLE OBSOLETE?

Occasionally the question is raised: "Isn't the business cycle obsolete?" It is more than 20 years since this country suffered a serious business recession. The business declines in the United States since World War II have been relatively brief and mild. There is much more government intervention in economic life in nearly all countries than there was 20 years ago, and we have noted that the maintenance of a high level of employment is an avowed objective of government policy in most countries. Besides, in these years of the cold war, large military expenditures provide an important and continuing support to the level of business activity. In view of all this, are we likely to continue to have the sort of business fluctuations that we call business cycles? In particular, are not depressions a thing of the past?

There is no doubt that we have greatly increased our ability to cope with business fluctuations; and, if we manage our affairs well, we ought to be able to avoid the extremes of both rapid inflation and severe depression. But it is by no means certain that we will manage things well. Our productive capacity will continue to expand, not smoothly, but in fits and starts; and, without wise management, the danger of too rapid expansion, followed by more or less severe depression, is a real one.

A leading authority in the field of business-cycle research has put the problem in words that are worth repeating.

> For well over a century business cycles have run an unceasing round. They have persisted through vast economic and social changes; they have withstood countless experiments in industry, agriculture, banking, industrial relations, and public policy; they have confounded forecasters without number, belied repeated prophecies of a "new era of prosperity" and outlived repeated forebodings of "chronic depression." Men who wish to serve democracy faithfully must recognize that the roots of business cycles go deep in our economic organization, that the ability of government to control depressions adequately is not yet assured, that our power of forecasting is limited, and that true foresight requires policies for coping with numerous contingencies.[2]

[2] A. F. Burns, *Stepping Stones Towards the Future*, 27th Annual Report of the National Bureau of Economic Research, 1947, p. 27. Reprinted in A. F. Burns, *The Frontiers of Economic Knowledge*, 1954, p. 45.

As another writer has put it, "The business cycle is not dead." Our postwar experience with cyclical instability has been "sufficient to demonstrate that business cycle phenomena continue to exert a profound influence upon our economy."[3]

Concern about the business cycle also represents a concern over our ability to maintain a steady and satisfactory rate of growth. Apart from their other consequences, business recessions represent interruptions in the growth of total output of goods and services. We have come to take for granted a steady upward trend in total production. We rely on this upward trend not only to feed, clothe, and house a steadily growing population but also to provide an ever-expanding standard of living for everyone. We count on continued growth not only in *total* output but also in output *per capita*. Many are also concerned over the rate at which such growth occurs. Our ability to compete with Russia, to carry the burden of our rising military expenditures and international obligations, and to finance an ever-greater demand for a variety of public services requires that the national output continue to expand at a satisfactory rate. What is a "satisfactory" rate is a matter of debate. It is clear that frequent or severe depressions make for an unsatisfactory rate of growth. Yet a very rapid rate of growth may create its own problems, among them an acceleration of inflationary tendencies and the development of economic maladjustments that will lead to a future depression.

PLAN OF THE BOOK

The heart of the present book is Part II, where we examine in detail the nature and causes of business fluctuations. But first we must equip ourselves with some of the necessary analytical tools. This is the function of Part I. In particular, we must understand how the economy generates a circular flow of income and spending, how we measure national income and the various components of aggregate spending, and something about the factors that determine the level of spending on and the total supply of new goods and services. It is only within the last 20 years or so that these topics have come to be included in books on business cycles. They are part of the theory of income and employment with which the name of J. M. Keynes is associated. Since changes in business activity arise out of fluctuations in

[3] G. H. Moore, "The 1957–58 Business Contraction: New Model or Old?" *American Economic Review*, vol. 49, May, 1959, pp. 292, 307.

spending (by consumers, businessmen, and government), consideration of these topics is an essential preliminary to the detailed study of business fluctuations and growth.

Part II offers a fairly detailed analysis of the nature and causes of business fluctuations, with an appropriate blending of statistical, historical, and theoretical treatments. Forecasting and the critical questions of policy—what we can do to keep the economy more stable than it has been in the past—are dealt with in the final chapters which make up Part III.

CHAPTER 2

SPENDING AND INCOME

IN THE "mixed economies" of the Western world, both private business and government employ economic resources, pay incomes to the factors of production, and produce goods and services. While the role of government in economic affairs has expanded greatly during the past century, economic activity in the United States and in most other countries is still primarily private business activity—controlled by businessmen and engaged in for profit. Goods are produced and job opportunities are offered when the current relation between costs and prices promises satisfactory profits to business firms. In short, the economy as a whole operates at a level determined by the profit calculations of individual businessmen. Some nonprofit-making economic activity has always been carried on by government at the local, state, and national levels—and the part played by government in economic life has been steadily expanding, in the United States as well as in other countries.

THE DEMAND FOR CURRENT OUTPUT

The output of an individual firm or industry and the prices it charges for its products are determined by a set of forces which economists sum up in the expression *demand-supply relationships.* A similar statement can be made about the economy as a whole. If there is no offsetting change on the supply side, an increase in demand for a single product will lead to increased output; and an increase in the total demand for things in general will lead to a general increase in production and therefore to an expansion in business ac-

tivity. For an increase in demand (technically, an upward shift in the demand curve) to lead to increased production of either a particular commodity or the composite of goods and services that make up the total output of the economy, it is necessary that the relevant supply curve have some elasticity; that is, there must be, in the case of total output, unutilized resources capable of increasing production in response to improved profit expectations. Otherwise the increased demand results only in higher prices for the same output.

We may therefore speak of "aggregate demand" and "aggregate supply" in describing the forces that lead to changes in the total output of goods and services, and hence to changes in business activity. These aggregative concepts—of total output and the demand for and supply of that output—are by no means simple, clear-cut notions. Total output comprises all newly produced goods and services, from musical recitals to newly constructed office buildings; and measures of the production of these different things do not all necessarily change in the same direction and at the same rate. This raises certain problems of measurement which we shall discuss in a later chapter. There is, however, one vitally important characteristic that is common to the production of all goods and services, and this characteristic provides one common denominator for measuring the effect of changes in aggregate demand. All production, whatever the end product, requires the use of economic resources, particularly labor. Although the correlation is not perfect, total output and the level of employment do move up and down together. It is clear also that the demand by businessmen for the factors of production is a derived demand. It is derived from the anticipated demand for the goods and services that the productive factors are employed to produce.

THE CIRCULAR FLOW OF SPENDING AND INCOME

Businessmen produce goods, and give employment to labor and the other factors of production, because they anticipate that buyers will purchase their output at profitable prices. They will tend to push output to the point at which the sums of money offered by buyers for the total output make that output worth while but provide no incentive to produce more (or less). We may call this situation one of equilibrium between aggregate demand and aggregate sup-

ply.[1] If aggregate demand should now rise (if the aggregate demand schedule or curve shifts upward), buyers spend more than formerly for the same output, profits and prices rise, and output and employment will expand toward a new equilibrium position, which in fact is never reached since further changes continue to occur on either the demand or the supply side.

Demand is reflected in the amount that buyers are willing to spend for different amounts of output. Spending, therefore, is one of the key variables to be studied in any analysis of the factors determining the level of business activity. The anticipation of spending by their customers leads businessmen to produce, and thereby to create job opportunities; the spending provides the proceeds from which payments can be made to the factors of production (that is, the spending in effect generates incomes); and these incomes, in turn, provide the source for further spending.

We therefore arrive quickly at the idea of a circular flow of spending and income creation. A simple representation of this circular flow is given in Figure 2. Out of the box, which represents the productive system as a whole, a flow of goods is continuously emerging, as indicated by the dotted arrow. The width of the pipe marked "Output" measures the value of these goods, and the value of these goods is nothing more or less than the sum of all incomes earned in producing them (including profits). Therefore, we must show emerging from the top of the box a pipe labeled "Income" which is exactly as large as the "Output" pipe. These incomes then move around (counterclockwise in the diagram) to create the spending that is exactly sufficient to buy the entire output of the period in question. This section of the pipe is labeled "Spending."

[1] Precise definitions of aggregate demand and supply may be offered here, although they are not essential in order to follow the discussion in the text. Following Keynes, we may speak of the aggregate demand schedule or function as that showing the total amounts that would be spent on current output corresponding to different levels of employment. The aggregate supply schedule shows the amounts of total spending necessary to induce businessmen to produce different levels of output. (Cf. J. M. Keynes, *The General Theory of Employment, Interest, and Money*, 1936, p. 25.) In the text, we shall use the concept of aggregate demand a bit more loosely than the preceding definition requires. We shall generally mean by aggregate demand not a schedule of hypothetical alternatives but the total amount actually spent or that would be spent in a given period on current output, given all the circumstances prevailing in that period, including the level of employment and the level and distribution of prices. This corresponds roughly to Keynes' concept of "effective demand." (*Ibid.*, pp. 25, 55.)

The anticipation of the spending led to output; this quantity of output at the assumed level of prices created incomes of the same money value; these incomes, if all spent, absorb the output; and this circular flow goes on continuously. In real life, the width of the pipe at some point is continually changing, thereby causing the

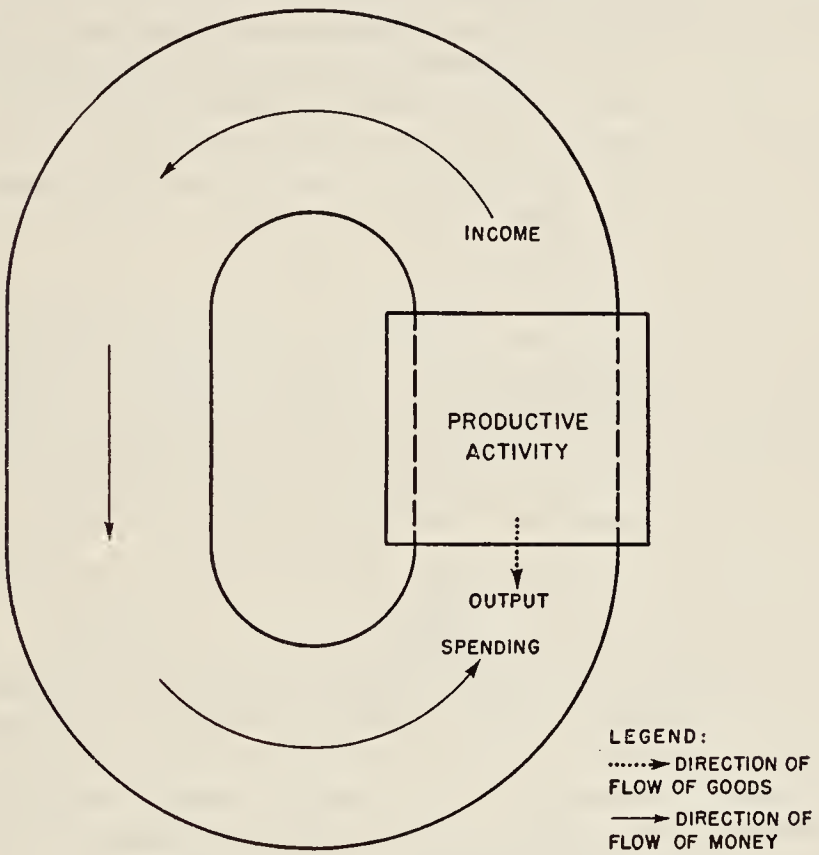


FIGURE 2. Circular Flow of Income and Spending.

pipe at all other points to expand or contract. We shall return to this issue in a moment.

The diagram in Figure 2 portrays a *flow* process; the size of the pipe measures the rate of flow over some period of time. The period may be a day or month or year. Thus in the year 1959 the total output of goods and services in the United States, and total spending on this output, amounted to 480 billion dollars. This was at the rate of 120 billions per quarter or 40 billions per month.

The circular flow process, as illustrated in Figure 2, describes certain relationships as they exist during a given period of time. The earning of income, the production of goods, and the spending of income on currently produced output are all portrayed as occurring simultaneously, as in fact they do. If incomes, output, and spending did not change with the passage of time, this would be all we needed to show. The circular flow diagram would then represent a cross-sectional view of what was happening in any short period of time. A given value of goods would be in the course of being produced; the same amount of incomes would be accruing to the productive factors; and, out of cash already in their possession, purchasers would be in the act of spending the same amount on currently produced goods.

In practice, the situation is much more complicated. Incomes, output, and spending constantly change in amount. While incomes are currently being earned but before they are actually received by the productive factors, spending on current output is taking place, with cash received in a previous period or with money borrowed for the purpose. The relation between past incomes and present spending may alter, causing changes in the valuation of current output and hence in the amount of incomes which correspond to the value of that output.

There is no completely satisfactory way of portraying these complex and changing relationships. Figure 3 attempts an approximation, which does have certain advantages over the first diagram by taking into account the passage of time. This is done through the introduction of a succession of "periods," thereby enabling us to picture more accurately the relations between production, the generation of income, and spending.

Figure 3 portrays the flow of income and spending through a series of hypothetical, short time-periods—say, days or weeks. In any one period, the volume of spending, the value of output, and the sum of currently accruing incomes are all equal. In any one period, the amount of spending determines the value of current output; they are different sides of the same set of transactions. Corresponding to this value of output there is an equal amount of income.

Still referring to Figure 3, let us now follow the flow from Period 1 to Period 2. In our illustration, the output-spending pipe in Pe-

riod 2 is greater than incomes earned and output in Period 1. We assume, which is approximately the case in practice, that incomes being earned in any relatively short period are available for spending only in the next period. This means that, although incomes rise in Period 2 because spending rises, the larger spending is not financed or caused by the larger incomes being earned in the same period. How, then, can spending in the second period be larger than the incomes earned in the first?

Spending can rise relative to receipts for either of two reasons.

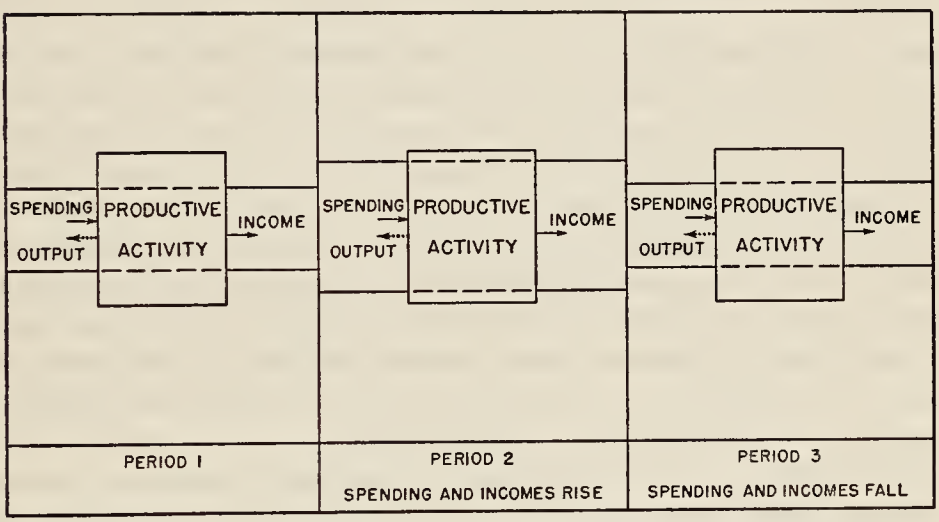


FIGURE 3. Flow of Income and Spending During Successive Time Periods.

Spenders may acquire additional cash by borrowing new money from the banks, in the form either of currency or of new deposits which are created for them. Or, without there being any increase in the supply of money, spenders may reduce the size of their cash balances, or, to put it not quite accurately, spend money that was formerly idle. In the terms used in the following section, spending can increase through an addition to the money supply (M) or through an increase in velocity (V).

Thus, to the incomes earned in Period 1 may be added borrowed new money or existing money that was idle. Spending in Period 2 is therefore greater than spending and incomes in Period 1; the value of output is greater; and hence also is income. The entire pipe in Period 2 is greater than that in Period 1.

If the increase in spending is unanticipated by businessmen, the physical volume of output may not rise, in which case the higher value of output is caused by higher prices.[2] The latter, if we assume costs have not risen, mean larger profits, and incomes rise by the same amount as spending and value of output.

Period 3 presents a case of a decline in spending. Despite the increase in incomes in Period 2, let us assume that individuals spend in Period 3 less than they earned in Period 2. Some individuals may have to repay bank loans, or they may choose to build up their cash balances in preference to spending. Thus the spending pipe in Period 3 is smaller than before; and value of output is correspondingly lower. If businessmen anticipate the decline, the physical volume of production will contract, there will be fewer jobs, and wages as well as profits will be lower. Total income will decline by the same amount as spending and value of output. If the decline in spending catches businessmen unawares, an unchanged physical volume of output will be sold at lower prices; the value of output still reflects the decline in spending; and so do incomes. In this case, it is profits alone that fall.

And thus the reduced incomes earned in Period 3 become available for spending in Period 4, which takes us beyond the limits of our diagram. Perhaps the reduced profits and lower prices of Period 3 lead to a cautious attitude by spenders. Some may prefer cash to goods, expecting prices to fall further; some, whose incomes have not yet been affected, may continue to save at their accustomed rate, but businessmen, with worsened profit expectations, choose not to borrow and invest these savings. For these and other reasons, spending in Period 4 may again be below the level of incomes in Period 3; and this means a corresponding decline in the value of output and in incomes.

This brief excursion into what economists call "period analysis," which is associated particularly with the name of D. H. Robertson, points up the fact that the size of the circular flow is subject to monetary influences. In particular, changes in the level of spend-

[2] Output may well lag behind spending, in which case we can speak of a "production lag." The adjustment needed to equate demand and supply in this case may take the form of either a change in prices or a change in inventories. The possibility of adjustment through changing inventories is purposely omitted at this point.

ing are influenced by changes in the supply of money and by the size of the cash balances that spenders desire to hold in relation to their current receipts. This carries us to the subject matter of the next section.

MONEY, VELOCITY, AND SPENDING

Spending is done through the use of money, either hand-to-hand currency or demand (i.e., checking) deposits. Total spending in any period—say, a year—may be several times the amount of money of all types (currency and deposits) actually available for spending during that period. Thus, in 1959, when total spending on currently produced output was 480 billions, the average amount of money in circulation (currency and demand deposits) was approximately 145 billions.[3] Actually each dollar had to do much more work than this implies, since many billions of dollars changed hands in the purchase of raw materials, securities, land, and many other things which are not included in the figure for final output. We shall return to this distinction in a moment.

We may call the average speed at which a dollar of the money supply makes a complete circuit—from one income receiver through the channels of production and distribution and back to another income receiver—the income velocity or circuit velocity of money. It can be measured by dividing total income (or spending on current final output) during a year by the average money supply during that year. Thus, in any period total spending and incomes may rise in association either with an increase in money supply, with no change in velocity, or with a change in velocity of a constant money supply.

Thus, in 1959 a money supply of 145 billion dollars was used to buy a current output of 480 billions.[4] Each dollar of money, on the

[3] This figure is an average of the figures at the beginning and end of the year for total demand deposits adjusted and currency outside banks as reported in the *Federal Reserve Bulletin*. Adjusted demand deposits exclude checks in process of collection, interbank items, and federal government deposits. We have excluded time deposits, although some economists would include them in measuring the money supply.

[4] As will become evident later, there is a distinction between gross income (or output) and net income (or output). The net national income in 1959 was 399 billion dollars, while the gross national product was 480 billions. The difference between net and gross output is explained in Chapter 3. Briefly, gross output, as here measured, includes depreciation and indirect business taxes. If we use the

average, bought more than three dollars worth of currently produced goods and services. As a dollar was paid out in, say, wages, it represented the distribution of a dollar of income. This dollar was eventually spent by the wage earner. It then passed to a retailer, becoming in part income to the storekeeper and his clerks; part was passed on to wholesalers, then to manufacturers, and so on. At the end of the year, our average dollar may have effected many dollars worth of transactions, and in the process it created about $3.30 worth of income (or alternatively bought this amount of final output corresponding to incomes earned).

Mention of money supply and velocity will undoubtedly bring to the reader's mind the familiar "equation of exchange": $MV = PT$. In this equation, M stands for the total supply of money, and V represents the velocity or rate of turnover of the money supply. Hence, MV is the same as total spending. On the right-hand side of the equation, T stands for the total physical volume of all sales transactions in which money changes hands, while P measures the average price of all the things involved in such transactions. Thus, PT represents the total value (price times quantity) of all transactions involving the use of money, and this is obviously equal to total spending (MV). This is why the equation is referred to as a truism; total spending must be equal to total sales.

MV represents a much larger volume of spending than that which we had in mind in referring to the aggregate demand that supports current production and employment. Total spending in the equation of exchange covers all monetary transactions, not merely purchases of the final output of new goods and services. It includes trading in paper wealth—stocks, bonds, and so on—as well as purchases of real goods; it includes every sale of an existing asset—land, old houses and other buildings, all sorts of secondhand goods—and not merely newly produced goods whose production generates incomes and employment in the current month or year; it includes, finally, every purchase and sale of raw material and semifinished goods as they progress through successive stages of

---

gross figure, income velocity was about 3.3; each dollar on the average bought $3.30 of current (gross) output. After deduction of depreciation and indirect business taxes, each dollar on the average, in the course of all the transactions in which it participated, resulted in $2.75 of net income to someone in the form of wages, rent, interest, or profits.

manufacture, as well as the sale of the finished goods to wholesalers and retailers, with the result that the value of the output of the earlier stages of production is counted more than once. It is clear, then, that total spending as defined by MV is much larger than the level of expenditures on new final products. The latter we may call "income-generating expenditures." It is this type of spending, not total MV, which directly determines the level of aggregate demand (for new goods and services) and thereby the level of employment, incomes, and business activity generally.

We have no complete record of all transactions included in MV (or PT). However, the great bulk of transactions in the United States is effected by check. Although we do not have a continuous record of all debits (that is, checks written) against demand deposits in all banks, we do have such data for bank debits and deposit turnover in all the important cities, which account for the great bulk of debits and deposits in the United States. In 1959, for example, the banks in the 344 cities included in this series reported total debits of $2,679 billion—a truly astronomical figure![5] To this we should add debits in cities not covered and all transactions effected through the use of currency. In contrast, income-generating expenditures (the gross national product) were only between a fifth and a sixth as large.

The ratio of total debits to the volume of bank deposits—the rate of turnover of these deposits—is a measure of the *transactions velocity* of that part of the money supply which consists of bank deposits. In 1959 the rate of turnover of demand deposits was more than 25 per year outside New York and more than 50 per year in New York City. (Deposit turnover is relatively high in New York because of the concentration of financial transactions in that city.) It is clear that the *transactions* velocity of money is much larger than *income* velocity, in the same measure that the more than two trillion dollars worth of total transactions in 1959 was greater than the 480 billion dollars worth of new goods and services produced. Transactions velocity (the V in the equation of exchange) measures the speed with which a dollar moves from one spender to the next; *any* type of spending tends to increase velocity

[5] This figure represents debits to total demand deposit accounts except interbank and federal government accounts and is reported regularly in the *Federal Reserve Bulletin.*

in this sense. *Income* or *circuit* velocity, however, measures the speed of the dollar in creating income (or purchasing final output); an expansion in intermediate transactions that merely transfer property rights or old goods increases V (M remaining constant) but does not lead to an increase in incomes or income velocity.

SPENDING THAT DOES AND DOES NOT CREATE INCOME

Our emphasis throughout this book will be on the aggregate demand for new goods and services and therefore on income-generating expenditures. However, spending even on securities or real estate can have repercussions on employment and income; and purely monetary changes, such as bank credit expansion (increase in M) or dishoarding of idle money (increase in V), may operate indirectly on incomes via initial changes in non-income-generating expenditures. (Of course, monetary changes may also act directly on the demand for new goods.) It is important, therefore, to see clearly the relations between MV and income-creating expenditures, and the relations between both and the supply and rate of use of money. We may express this relation algebraically as follows:

$$MV_F + MV_D + MV_Y = PT$$

The subscripts F, D, and Y relate to the three types of spending previously described. The first refers to essentially financial and speculative transactions, in securities, land, old buildings, etc.[6] The subscript D refers to "duplicative" transactions resulting from the fact that most goods are bought and sold more than once as they move through successive manufacturing stages and through wholesaling and retailing in their path to the final customer. Finally, MV_Y refers to expenditures on newly produced goods at their point of final use, or, alternatively, to the sum of incomes that result from the sale of these goods.[7] The M's in the equation are the

[6] This type of expenditure creates minor amounts of income—for example, in the form of brokerage fees—but this unimportant complication can be neglected, or MV_F can be interpreted as financial transactions net of expenses that represent someone's income.

[7] If MV_Y refers to the purchase of current output, actual payments of income to the factors of production cooperating in the production of these goods must be included in MV_D; otherwise there would be double counting. Let us say that a billion dollars is paid in wages, interest, etc., which eventually results in a flow

same, i.e., the total money supply; the V's, however, are different, each representing the type of velocity appropriate to the kind of expenditure in question.

This elaboration of the equation of exchange helps us to see why some writers have emphasized monetary factors in discussing the causes of changes in the general level of prices and of business activity. Obviously, however, the V's may change as well as M; a change in any of the V's involves human motivation—a change in spenders' attitudes toward holding money in preference to acquiring new goods and services or other assets. The spenders may be consumers, businessmen, or the government; and the changes in their rate of spending may affect the physical volume or the prices of the different kinds of things that are represented by the various MV's. The impetus to change, as a matter of fact, will frequently come from the right-hand side of the equation, which, like the left, is a sum of three sets of transactions. New investment opportunities opened up to business may lead to an increase in output (included in T) which at the same time expands MV_Y and may, through new activity in financial and raw-material markets, lead to further expansion in MV_F and MV_D (and to corresponding further changes in the right-hand side of the equation).

Let us consider a bit further the right-hand side of the equation of exchange. Corresponding to our breakdown of MV into three components, we can similarly divide PT, so that we have:

$$MV_F + MV_D + MV_Y = (PT)_F + (PT)_D + (PT)_Y = PT$$

Changes in $(PT)_F$ can be measured by multiplying an appropriate index of the prices of securities and old assets by an index of the physical volume of trading in such things, and similarly for $(PT)_D$, which refers to the value of goods (raw materials, shipments to wholesalers and to retailers, and so on) which are bought by businessmen for the purpose of passing them on to the final customer. $(PT)_D$ excludes the sale of the finished product to the final user, which is measured by $(PT)_Y$.

None of the (PT)'s is a simple homogeneous total, and in each case there are serious difficulties in trying to measure the P or T

---

of goods which are bought by ultimate users for a billion dollars. We count this expenditure on the goods as MV_Y but not both that expenditure and the payment of an equivalent amount to the factors of production.

in question. Thus in $(PT)_F$ we are talking about the prices of and volume of trading in stocks, bonds, land, old buildings, loans on mortgages, and so on. In the case of $(PT)_D$, the appropriate measure of prices comes close to an index of wholesale commodity prices, although some goods are sold at wholesale to the final users and therefore do not represent duplicative transactions.[8] The price element in $(PT)_Y$ refers chiefly to retail prices, since consumers are the most important group buying goods and services for final use. But $(PT)_Y$ also includes capital goods sold to businessmen for installation in their plants and shops; in addition, it includes the goods and services that government buys. As we noted previously, $(PT)_Y$, which is the same as MV_Y, represents the total spending on goods and services by the three main groups of final users of new output: consumers, business, and government.

THE NEED FOR DETAILED ANALYSIS OF TOTAL SPENDING

Although none of the magnitudes in our elaborated equation of exchange is simple conceptually or easy to measure, the equation serves a useful introductory purpose. It permits us to see the reciprocal two-way relationship between the monetary factors, M and V, and the results of spending as they are reflected in changes in prices and the physical volume of activity. It further provides a useful preliminary look at the possible relations between spending that directly affects incomes and employment and spending that affects these crucial variables only indirectly. And, finally, the discussion thus far points up an important truth which is sometimes forgotten when we concentrate on broad aggregates such as total spending. The clue to the causes of changes in broad aggregates is usually to be found in an analysis of the changes in the components which make up the totals. In recent years, economists have tended to neglect the study of total spending in the inclusive sense of $MV = PT$ in order to concentrate on aggregate demand and income-generating expenditures as reflected in $MV_Y = (PT)_Y$. Actually, we need to look at both, and in as much detail as the information will permit.

It is important to remember in this connection that aggregate demand as measured by MV_Y is merely the total of the demands

[8] As previously mentioned, the volume and prices of the factors of production are also included in $(PT)_D$.

for the products of many different industries, and the demand stems from different groups of buyers. Shifts in the direction of spending may lead to changes in the totals, and some types of spending may be more important than others in their repercussions on the level of business activity. Further, changes in some of the factors affecting the economy as a whole—for example, an increase in M resulting from an expansion of bank deposits—may affect one type of spending more than another and may lead in some cases to an increase in output and in other cases merely to a rise in prices.

It is important to remember, also, that spending, whether on new goods and services or on old assets, represents volitional action on the part of moderately rational human beings. Spending generally occurs for one of three reasons, corresponding to the three primary types of spenders. (1) Consumers spend in an attempt to maximize the satisfaction from their money incomes. They must choose first between consuming and not consuming; then they must distribute their consumption expenditures among the different things they wish to buy and direct their savings toward the forms of "investment" they prefer (securities, bank accounts, their own business, or perhaps currency in the mattress). It is important, therefore, to study the reasons for the spending behavior of consumers: the factors that determine how much of a change in income will be consumed or saved, why and how consumers distribute their expenditures on the kinds of goods available for purchase, the reasons for the disposition they make of their savings, and so on. (2) Government spending is made by officials to implement the policies of the state. Coming from a single source that can be turned on or off and directed by a few individuals in any chosen direction, government spending is the only one of the three types that, in a free-enterprise economy, is subject to immediate and direct control by a central authority in order to achieve desired ends of social policy. That is why, since the 1930's, so much has been heard of government intervention and government spending as instruments of public policy. This is a subject that will much concern us in later chapters. (3) In a private-enterprise economy it is business spending that most affects the level of employment and incomes. Business expenditures depend on profit expectations; favorable profit expectations depend primarily on profitable relation-

ships between costs and prices in the recent past and present. These crucial cost-price relationships are those which exist in different industries and firms; they cannot be expressed as a simple overall average ratio of all prices to all costs. Indeed, every cost to a buyer is a price to some seller. Thus, a study of the factors determining the level of business expenditures on final new output (what we call private investment or capital formation) must entail a study of the multitude of causes operating on profit expectations, and this leads inevitably into a study of the network of interrelationships which hold the economic system together. In short, a study of the crucial aggregates affecting business activity and incomes is but the first step in the process of getting at the causes of instability in our economic system.

We have already said that spending on final output comes from three classes of buyers: consumers, business firms, and government.[9] (The business spending included here is confined to expenditures on plant and equipment and on inventories. All other business spending is on goods and services—for example, raw materials or fuel—which become embodied in final products which are then sold to one of our three types of buyers; it therefore represents duplicative expenditures rather than spending on final output.)

Let us represent these three types of expenditures by, respectively, the symbols C, I, and G. The sum of the three equals the value of final output, for which, in the equation of exchange, we used the expression $(PT)_Y$. Thus we can write:

$$C + I + G = (PT)_Y = MV_Y$$

To put the same thing in words, changes in the value of final output come about because of changes in total spending on new goods and services (aggregate demand). The change in the value of final output may represent a change in the volume of goods and services (T) or merely a change in prices (P). An increase or decrease in total spending must come from one of three types of spenders, and for the spending of any of these groups to change there must be a corresponding change in the supply of money (M)

[9] Actually, there is also a fourth category of demand for final output, which we omit here in order to simplify the discussion. This is the *net* demand from abroad for the output of a given country—roughly, the excess of its exports over its imports.

or in income velocity (V_Y). Thus, in a sense, every change in aggregate demand has both a monetary and a nonmonetary aspect. Some group in the economy must have decided to change its spending, and monetary conditions must have been such as to permit the necessary change in M or V_Y. The reader can perhaps begin to see already why arguments so frequently arise—among economists as well as laymen—as to whether particular changes in output or the price level were "caused" by monetary conditions or by other factors.

A NOTE ON AGGREGATE SUPPLY

So far we have concentrated our attention on total spending and aggregate demand. The equation of exchange reminds us that we should also take a look at the supply side. If spending on final output rises but output itself does not change, then what we get is an increase in prices. This suggests at least one way of describing what we mean by *inflation*. It is what happens when aggregate demand rises faster than total output, so that prices rise.[10]

As long as there is unemployed labor and unused plant capacity, output can rise as fast as expenditures. But as we approach full employment and full capacity, the supply of new output becomes increasingly inelastic, and prices begin to rise. What happens is illustrated in Figure 4. Here we measure total expenditures on new output on the vertical axis and the physical volume of output on the horizontal axis.

Let us consider first the curve marked OS_1. This describes how output would behave as total expenditure (aggregate demand) rises, on the assumption that certain things are given and do not change. These given conditions are: (1) a certain amount of productive resources in the economy—labor, plant, etc., (2) the productivity of these resources—which depends on, among other things, the state of technology, and (3) the prices at which the factors of production, particularly labor, are willing to supply their services.[11] A change in any of these conditions will cause the

[10] This is not to suggest that inflation can occur only if it originates on the demand side. There can be a "cost-push" as well as a "demand-pull" inflation, or, in Professor Lerner's terms, a sellers' as well as a buyers' inflation. Figure 4 can be used to illustrate both types, but this would introduce complications that are unnecessary at this early stage of our analysis.

[11] It might be mentioned that the OS curves in Figure 4 correspond to aggregate supply functions as defined in footnote 1 of this chapter.

curve to shift, in ways that we shall describe later. For the time being, we need merely remember that the OS_1 curve assumes a given amount of labor and capital, a given state of technology, and a given level of wage rates.

The curves OS_2 and OS_3 are to be interpreted in a similar way. The fact that they are farther to the right implies that the produc-

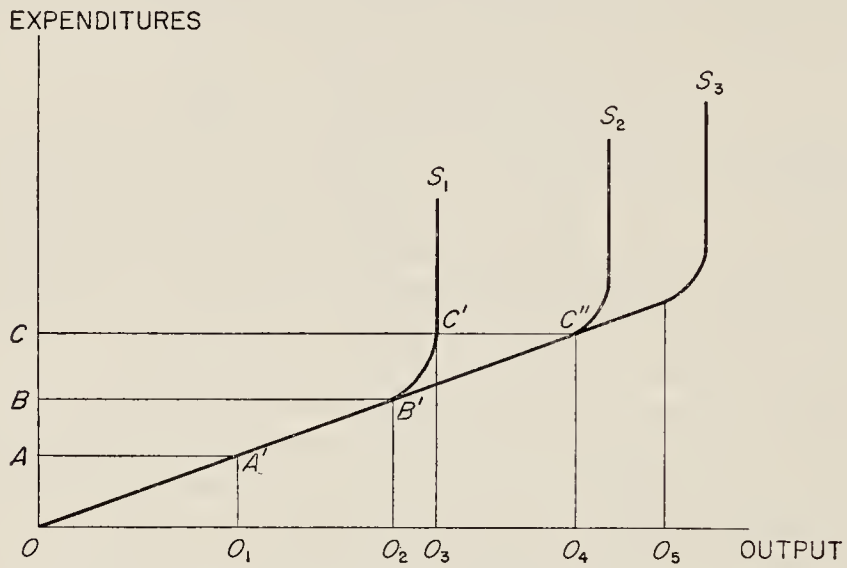


FIGURE 4. The Relation Between Total Expenditures and Output.

tive capacity of the economy is larger than in the first case, either because supplies of the factors of production are larger, or because factor productivity has increased, or for both reasons.[12]

Now let us assume that at the moment the level of aggregate demand, measured vertically in Figure 4, is at A. The horizontal line AA' meets the OS_1 curve at A'. Thus total expenditures are O_1A', and the level of output is OO_1. At this level of output, there are still unutilized resources, because expenditures could rise to B $(= O_2B')$ and bring forth a proportional increase in output. If expenditures were to rise above B—say, to C—output would rise much more slowly (only to O_3), *as long as we stay on the*

[12] The way that we have drawn these curves implies that if, for example, labor productivity increases from OS_1 to OS_2, wage rates rise correspondingly so that unit labor costs do not change, at least for the sections of the OS curves that are on the same straight line.

curve OS_1. If expenditures were to rise to C $(=O_3C')$ while output rose only to OO_3, prices clearly would have to rise.[13]

Let us translate this little piece of geometry into common-sense terms. How much the economy can produce is determined by the supply of the factors of production available and by the productivity (or efficiency) of these factors. As spending increases, there is a limit to how much output can increase unless the supply of productive resources, or the productivity of these resources, increases. This is what the shape of the curve OS_1 tells us.

To increase productive capacity takes time—time for the labor force to increase, for new plants to be built, for technological improvements to increase the efficiency of labor and capital. Thus, given enough time, the economy can move from curve OS_1, to OS_2, to OS_3 and so on through the years. But if we push up spending too rapidly—faster than the productive capacity of the economy expands—then beyond a certain point rising expenditures will evoke smaller and smaller increments of output, with the result that the increase in spending will be dissipated in rising prices rather than expanding production. If, however, aggregate demand rises more gradually, output may be able to expand as rapidly as spending, and in that case prices need not rise.

Let us turn back to Figure 4 for a moment. Suppose that spending jumps suddenly from B to C, while existing productive capacity, as reflected in the OS_1 curve, remains unchanged. Prices will rise, although there will also be some increase in output, from OO_2 to OO_3. If the same increase in spending takes place more gradually, while expanding productive capacity permits a shift to the OS_2 curve, there will be a much larger increase in output and no rise in prices.

The problem of matching demand and supply cuts both ways. Assume that, with a growing labor force and stock of capital, the economy moves from the aggregate supply function OS_1 to OS_3, while aggregate spending rises only to C. This level of spending is

[13] This is so because the price level of output is implicitly defined as the ratio of expenditures to output. The price level of the output OO_1 is $\frac{O_1A}{OO_1}$, which is equal to $\frac{O_2B'}{OO_2}$ if OB' is a straight line. Clearly, $\frac{O_3C'}{OO_3}$ is greater than either of the two preceding ratios.

not sufficient for full employment, and the economy will have idle resources. The full employment level of output for the curve OS_3 is in the neighborhood of OO_5, beyond which the curve begins to bend up rapidly, but the actual level of output is only OO_4.

Thus the combination of a satisfactory rate of growth of total output with full employment and no inflation requires (1) that the *OS* curves continue to shift to the right (i.e., that productive capacity continue to expand) at a satisfactory rate, and (2) that aggregate demand continues to rise at about the same rate, fast enough to absorb the increase in productive capacity but not so fast as to cause inflation.[14]

In this chapter we have been concerned with the interrelations among income, spending, and output for the economy as a whole. These are not merely theoretical concepts. They are economic variables which, fortunately, we can measure and the actual behavior of which we can observe in some detail. In the next chapter we shall see how in fact we do measure the national income and the components of aggregate demand. This will provide us with the opportunity to examine in some detail the rather complex interrelationships between income and spending in the United States.

[14] A further condition is that the *OS* curves do not shift vertically. This would happen if factor prices increased more than productivity, so that the increase in unit costs brought about a rise in prices even before the curved (full-employment) section of a particular *OS* curve was reached. This is a problem that we must defer to Chapter 4.

CHAPTER 3

NATIONAL INCOME AND GROSS NATIONAL PRODUCT

WE SAW in Chapter 2 that expenditures on new final output determine the level of current production; the level of these expenditures measures the value of current output and the amount of currently earned incomes; and these incomes are the source of continuing expenditures on new goods and services. Data on the national income and on expenditures on new goods and services, therefore, can be invaluable in studying the factors responsible for changes in the level of income, production, and employment—or of business activity measured in some other way. Hence it is not surprising that data on the national income are being more and more widely used, not only by economists but also by businessmen, government officials, newspaper editors, and even the man in the street.

As a result of the increasingly felt need for data in this field, remarkable progress has been made in the last 30 years in developing reasonably reliable measures of the components of national income and output. The pioneer in this area in the United States was the National Bureau of Economic Research. The National Bureau's methods of estimating the national income and, later, the gross national product were eventually taken over, with important modifications, by the federal government.[1] Techniques of

[1] Some of the National Bureau's work on national income dates back to the early 1920's. Since the early 1930's, its work on national income and product has been carried on chiefly by Simon Kuznets. See, for example, his *Commodity Flow and Capital Formation*, 1938; *National Income and Its Composition, 1919–38*, 1941; *National Income: A Summary of Findings*, 1946; and *National Product Since 1869*, 1946. The Department of Commerce began in the early 1930's to prepare estimates of the national income by type of payment and industrial

measurement and sources of information have been steadily improved, although the official estimates are still subject to varying, and in some cases disturbingly large, margins of error.

Today the United States Department of Commerce compiles quarterly and annual estimates of the gross national product and net national income, both shown in considerable detail. Current figures are published in the *Survey of Current Business* and are promptly reproduced in many other government and private publications. The Department substantially revised its estimates in 1947, at which time it first presented the data in the form of an integrated system of national accounts, and further important revisions have been made since.[2] Annual estimates are available back to 1929, and quarterly figures have been carried back to 1939. Since 1947, detailed breakdowns of the estimates, with corresponding figures for preceding years, have been presented annually in the July issue of the *Survey of Current Business.* Our discussion in the remainder of this chapter will be concerned with the estimates of national income and output as they are currently made by the Department of Commerce, on the basis of its most recently revised definitions and methods of estimation.

Estimating the national income has become an important statistical function of government in many other countries besides the United States. Indeed, as one group of experts observed, "The international spread of regular national income and product estimates in the decade after World War II was phenomenal. It is doubtful whether any equally important statistical innovation ever gained ground as rapidly on an international scale."[3] By 1958, more than 75 countries were regularly preparing national income

source. Beginning in the early 1940's, it began to publish estimates of the gross national product, by major groups of final products and type of final buyer, as well as continuing and improving its work on the income side.

[2] See, in particular, U.S. Department of Commerce, *U.S. Income and Output,* supplement to the *Survey of Current Business,* 1958, and *National Income,* supplement to the *Survey of Current Business,* 1954. For a brief history of the department's work in developing national income estimates, see the first of these two sources, pp. 47–49, and *The National Economic Accounts of the United States,* A Report of the National Accounts Review Committee of the National Bureau of Economic Research, National Bureau of Economic Research General Series No. 64, 1958, pp. 24–26.

[3] The *National Economic Accounts of the United States,* p. 28. Fairly detailed national income data for a large number of countries are published by the Statistical Office of the United Nations in *Yearbook of National Accounts Statistics.* The 1958 issue of the latter contained detailed estimates for 78 countries.

estimates. "National economic budgets," which contain detailed estimates of the components of the gross national product, provide the basis for much of the national economic planning in such countries as the United Kingdom, Sweden, Norway, and the Netherlands.[4]

SOME LIMITATIONS OF NATIONAL INCOME DATA

In examining the nature of these estimates of national income and gross national product, and their possible uses in business-cycle analysis, the reader should bear in mind the limitations of the data and of the concepts which underlie them. All estimates in the field of national income are arbitrary in the sense that discretionary judgment and the availability of data dictate what goods and services are included and how they are valued. The National Bureau of Economic Research, for example, always included the estimated rental value of owner-occupied homes in its totals; the Department of Commerce excluded this item until its 1947 revisions. The services of housewives are excluded, although the addition to the nation's welfare resulting from such services is both unmistakable and substantial. Similar questions of definition, and also of valuation, arise in the treatment of government activities. Before 1947, for example, the Department of Commerce included corporate income taxes in the gross national product but not in the national income; now they are included in both. Government interest payments are now excluded on the grounds that they are not a payment for a productive service; before 1947, they were included in both gross and net national income. Further arbitrariness arises in the classification of the national product. Residential construction, for example, is generally treated as a part of capital formation, whereas purchases of furniture and automobiles, which are also durable goods used by consumers, are treated as acts of consumption.

This does not exhaust the list of conceptual difficulties that arise in the field of national income estimation. Different estimators handle these problems differently, and data compiled by different organizations or for different countries must be examined carefully before comparisons are made. Fortunately, these difficulties are not too serious in analyzing changes over time if the data

[4] For a discussion of the Swedish case, see Erik Lundberg, *Business Cycles and Economic Policy*, 1957, esp. chap. 8.

compiled by the same estimator are used, and if the latter does not significantly alter his definitions, system of classification, or methods of compiling the figures. In this case, year-to-year changes in the data are likely to be significant, even though we may have some reservations as to the accuracy with which the absolute figures for any year measure the magnitudes they purport to represent.

The major difficulty in appraising the significance of changes reflected in, say, the Department of Commerce data is likely to come either when there are important movements in prices or when there is a major shift in the relative importance of different components of the national income. When prices change, we can never be sure to what extent the dollar totals reflect price movements and to what extent changes in physical volumes. No system of deflating the value figures by index numbers of prices will completely eliminate this difficulty, although existing deflating techniques do yield usable approximations. When there is a drastic shift in the relative magnitudes of some of the components, particularly when the government's share alters significantly, we again run into trouble in interpreting the figures. Thus, during World War II the role of government in American economic life expanded enormously. It has been argued that the Department of Commerce data exaggerated the rise in the national product during this period because the government's takings of goods and services were valued at a higher figure than would have resulted had the same economic resources been utilized under peacetime conditions. When the components of national output change in relative importance as much as they do in wartime, the meaningfulness of the estimates over time, particularly as measures of changes in real output, is open to serious question.[5]

These and similar conceptual problems cannot be simply resolved; and some of them cannot be resolved at all.[6] Our ap-

[5] See, for example, Simon Kuznets, *National Product in Wartime,* 1945, and a symposium by Milton Gilbert, Hans Staehle, W. S. Woytinsky, and Simon Kuznets, "National Product, War and Prewar," *Review of Economic Statistics,* vol. 26, August, 1944, pp. 109–135.

[6] In this connection, see the collection of papers in Conference on Research in Income and Wealth, *A Critique of the United States Income and Product Accounts,* Studies in Income and Wealth, vol. 22, 1958. This volume not only summarizes the more important issues currently being debated but also gives numerous references to the literature on national income accounting.

proach in this and other chapters of this book, however, makes these difficulties less serious than they might otherwise be. First of all, our interest in the national income and product stems primarily from our interest in the behavior of aggregate demand; we wish to measure the size and trace the direction of money flows. Though we are also obviously interested in the flow of goods, we are not particularly interested here in the precision with which any given estimate of the national income measures changes in some conception of economic welfare. The Department of Commerce estimates are particularly suitable for a study of changes in aggregate demand and of the disposition of the proceeds from the sale of the national product. In addition, we do not ordinarily have to cope with changes in the composition of the national product as violent as those that occur during war. Normally, the totals are satisfactorily comparable from quarter to quarter and from year to year. For other purposes, however, the Commerce data are less suitable; and these and other income estimates are subject to important limitations when used to measure changes in real output, particularly over considerable periods of time.[7]

GROSS AND NET NATIONAL PRODUCT AND NATIONAL INCOME

In 1958 the national income earned by the people of the United States was, according to the Department of Commerce, 366 billion dollars. This represents both the incomes earned by the factors of production and the net value at factor cost of the goods and services produced in 1958. The Department of Commerce reported, however, that the *gross* value of new goods and services sold to final users (excluding duplication of raw-material expenditures included in the MV_D of the elaborated equation of exchange) was 442 billion dollars. The gross national product was some 76 billions larger than the net national income. We need now to in-

[7] This brief discussion of some of the problems that arise in measuring the national income barely scratches the surface of the subject. Interested students should consult authoritative studies such as the volumes cited in the preceding footnotes, the numerous volumes of *Studies in Income and Wealth* published by the Conference on Research in Income and Wealth of the National Bureau of Economic Research, and the useful text by Richard and Nancy Ruggles, *National Income Accounts and Income Analysis,* 2nd. ed., 1956. See also United Nations, *A System of National Accounts and Supporting Tables,* 1952, and *Methods of National Income Estimation,* series F., no. 8, 1955.

vestigate the nature of this difference. The reconciliation, as made by the Department of Commerce, is shown in Table 1.

The difference between the gross national product (which we shall hereafter refer to simply as GNP) and the national income is accounted for chiefly by two items: capital consumption allowances (primarily depreciation on buildings and equipment) and indirect business taxes. The reason for the deduction of depreciation and similar items of capital consumption is simple. As we produce new output, a part of the nation's capital equipment is used up in the process. To this extent, new output does not represent income but is necessary to replace capital goods consumed in the process of production. Not to make this deduction would be to exaggerate the amount of net income; and, if the entire output were consumed, the nation's real capital would gradually be exhausted. At the same time, this item must be included in the GNP if we are to show the total expenditures made by final users in buying the national output.

TABLE 1. Relation of Gross National Product, Net National Product, and National Income, 1958
(In billions of dollars) [a]

Gross national product (representing total expenditures on new output by final users)	441.7
Less: Capital consumption allowances (chiefly depreciation but also accidental damage to capital equipment and capital outlays charged to current expense)	37.9
Equals: Net national product (which, however, is not the same as the total of incomes accruing to the factors of production)	403.8
Less: Indirect business tax and nontax liabilities (which are included in the sales price of the national product but do not accrue to the productive factors as income)	39.0
And less: Other adjustments [b]	−1.4
Equals: National income (which is the sum of all incomes, before income taxes, earned by the factors of production in producing new goods and services)	366.2

[a] Adapted from *Survey of Current Business*, July, 1959.

[b] Includes business transfer payments of 1.7 and statistical discrepancy of −2.1, which are deducted from net national product, and "subsides less current surplus of government enterprises" of 1.0, which is added. Since, in this case, a minus quantity is being deducted, the absolute amount of the net adjustment is added. In other years, the net adjustment might be positive, in which case the absolute amount would be subtracted.

The subtotal secured by subtracting capital consumption allow-

ances from the GNP is called the net national product.[8] We use the word *net* because depreciation has been subtracted. The net national product is valued at the market prices paid by the final purchasers, and at least one element in the market price cannot be passed on to any factor of production as income. This is the total of sales and other indirect business taxes which are paid to the government. With minor exceptions, the rest of the net national product accrues as income—in the form of wages, rent, interest, and distributed and undistributed profits—to the various factors of production.

Let us look a little further at the reason for deducting indirect business taxes from net national product in order to arrive at the national income. These indirect business taxes (chiefly federal and state sales and excise taxes but including also property taxes) are included in the price paid for current output by consumers and other final purchasers. They are therefore included in the gross national product, which records expenditures on new goods and services, and also in the net national product. These taxes, however, are paid by business to the government, rather than directly to the factors of production.[9] They may be looked on as payments for government services by business, such services being incorporated in the final product sold by business. Since the government is not an ultimate factor of production, these taxes must be

[8] The Department of Commerce no longer emphasizes this subtotal, and it does not now, as it once did, show it separately in its basic summary table for each year. The subtotal is still presented, however, in a different table presenting data for past years. Compare Tables I and I–17 in the July issues of *Survey of Current Business*.

[9] When the government uses these tax receipts, say, to pay salaries, these salary payments are included in the GNP as a purchase of services by government. To illustrate, assume that $2,000,000 worth of cigarettes are produced and sold, $700,000 of this amount being taken by the government in taxes and the remainder paid out in profits and wages by the cigarette companies. Suppose also that the government immediately spends the $700,000 in paying the salaries of government workers. The national income statement would be as follows:

Purchases by consumers (cigarettes)	$2,000,000
Purchases by government (services of government workers)	700,000
Gross national product	2,700,000
Less: Business taxes	700,000
Net national income	2,000,000
which was earned as follows:	
Wages and profits in cigarette industry	1,300,000
Government salaries	700,000

deducted in order to arrive at the national income accruing to the factors of production. Before deducting these indirect taxes, we have the net national product valued at market prices. After the deduction, we have the net output valued at factor cost—i.e., in accordance with the incomes earned in producing this output—and this is what we call the national income.

For our purposes, the net national product is not an important subtotal, nor is it emphasized by the Department of Commerce. The nature and derivation of the GNP and national income, however, should be clearly understood and remembered.

Let us look briefly at the item entitled "other adjustments" in Table 1. The two most important adjustments are business transfer payments and a "statistical discrepancy." A small part of the gross receipts of business is used to make payments which are in the nature of gifts. They do not represent the purchase of productive services and hence are not considered part of the national income. They include such items as corporate gifts to nonprofit institutions, consumers' bad debts, cash prizes, and so on. These are similar in principle to the much more important category of government transfer payments, which we shall consider later. Both types of transfer payments represent transfers of purchasing power which do not correspond to incomes earned in producing new goods and services.

The statistical discrepancy, which is also included in the other adjustments, arises out of the fact that there are two ways of computing the GNP and national income, and the Department of Commerce uses both. The GNP is computed by the "final-products method," that is, by estimating directly the sales of final products to consumers, business, government, and foreigners. Then the national income is estimated entirely independently by adding the incomes earned by the factors of production in the form of wages, rent, interest, and distributed and undistributed profits. By definition, the national income is equal to the GNP minus capital consumption allowances, indirect business taxes, and the other minor deductions. Because of estimating errors, however, the figure for the national income obtained in this way will not be precisely equal to the national income computed directly by summing the incomes earned by the factors of production. Hence we have the "statistical discrepancy" as an adjusting item to maintain equality

between the "final-products" and "incomes-earned" methods of estimating the national income.

If we neglect the minor adjustment items, the gross national product is larger than the national income by the sum of capital consumption allowances (chiefly depreciation) and indirect business taxes. Our primary interest is in the GNP rather than the national income. Our concern is with the total demand for new goods and services, as measured by the GNP, and with the distribution and use of the gross sales proceeds by income receivers, business, and government.

THE INCOME SIDE OF THE NATIONAL INCOME ACCOUNT

We have seen that there are two ways of measuring the national income, either by adding the value of newly produced output sold to final users or by taking the sum of incomes earned in producing this output. We referred to these as the "final-products" and "incomes-earned" methods of measuring the national income.

These two ways of viewing the national product correspond to the two sides of a national income account, in which we record on one side the incomes earned in producing a given output and on the other side the expenditures made on that output. One of the great advances made in the measurement of national income in the last 15 years has been the setting up of a system of accounts for the economy as a whole, just as a business firm has a set of accounts in which to record its receipts and expenditures.[10] Such an overall account for 1958 is given in Table 2.

The left-hand side of Table 2 records the incomes earned in producing the national product in 1958. The sum of employees' compensation, proprietors' income, rents, corporate profits, and interest comes to the national income of 366 billions, the same figure given in Table 1. Then we have to make the adjustments previously described and add in capital consumption and indirect business taxes to get the GNP of 442 billions. Here we are making the same corrections to go from national income to GNP that

[10] Unfortunately, we do not have yet anything for the economy as a whole that corresponds to a business firm's balance sheet, i.e., that records the nation's assets (wealth) and liabilities. Raymond Goldsmith has been a pioneer in developing a usable set of wealth data for the United States. For a statement of the problem, see National Accounts Review Committee, *The National Economic Accounts of the United States*, chap. 14.

we made in moving from GNP to national income in our discussion of Table 1.

The right-hand side of Table 2 is self-explanatory. This is the "final-products" or expenditures side. It is standard practice to use four categories on this side of the national income account; and, as we shall see later, each category has an economic significance of its own. These are consumers' expenditures, expenditures on

TABLE 2. National Income and Product Account, 1958
(In billions of dollars) [a]

Compensation of employees	256.8	Personal consumption expenditures	293.0
Proprietors income	46.6	Gross private domestic investment	54.9
Rental income	11.8	Net exports of goods and services	1.2
Corporate profits and inventory valuation adjustment	36.7	Government purchases of goods and services	92.6
Net interest	14.3		
NATIONAL INCOME	*366.2*		
Business transfer payments	1.7		
Indirect business tax and nontax liability	39.0		
Current surplus of government enterprises less subsidies	−1.0		
Capital consumption allowances	37.9		
Statistical discrepancy	−2.1		
GROSS NATIONAL PRODUCT	441.7	GROSS NATIONAL PRODUCT	441.7

[a] From *Survey of Current Business*, July, 1959.

capital formation or investment (plant, equipment, increase in inventories, etc.), net purchases by foreigners (exports minus imports), and government purchases of goods and services. We shall treat each of these types of expenditures in some detail in the next section, but first we should say something more about the income side.

To repeat, the national income represents the incomes earned by the factors of production. (For this reason, the national income is sometimes described as the national product valued at factor cost—i.e., at cost in terms of the incomes, including profits, that accrue to all productive factors in producing this output.) Data on incomes earned can be classified in a number of ways—for example, by type of payment (wages, interest, etc.) or by the kind

of industry in which earned. These two ways are illustrated in Table 3. Thus, in 1958, of a national income of 366 billions, 19 billion was earned in agriculture, 5.3 in mining, and so on. These figures also measure the *net value added* to the total output of the American economy by each of these industries. Thus, while the

TABLE 3. National Income, by Industrial Origin and Distributive Shares, Selected Years, 1929–1958
(In billions of dollars) [a]

	1929	1933	1939	1958
Industrial Origin				
Agriculture, forestry, fisheries	8.3	3.7	5.9	19.0
Mining	2.0	0.6	1.6	5.3
Contract construction	3.8	0.8	2.3	19.9
Manufacturing	21.9	7.6	17.9	103.7
Wholesale and retail trade	13.4	5.5	12.5	60.5
Finance, insurance, real estate	12.7	5.7	7.9	36.6
Transportation	6.6	3.0	4.6	16.4
Communications and public utilities	2.9	2.0	2.9	14.4
Services	10.3	5.6	8.3	41.6
Government	5.1	5.3	8.5	46.7
Rest of the world	0.8	0.3	0.3	2.1
Total national income	87.8	40.2	72.8	366.2
Distributive Share				
Compensation of employees	51.1	29.5	48.1	256.8
Proprietors' income	14.8	5.6	11.6	46.6
Rental income	5.4	2.0	2.7	11.8
Corporate profits [b]	10.1	−2.0	5.7	36.7
Net interest	6.4	5.0	4.6	14.3
Total national income	87.8	40.2	72.8	366.2

[a] Figures for 1929, 1933, and 1939 are from *U.S. Income and Output*, Supplement to *Survey of Current Business;* those for 1958 are from *Survey of Current Business*, July, 1959.
[b] After inventory valuation adjustment.

total gross value of agricultural production was considerably more than 19 billion dollars, only this much corresponded to incomes earned *in* agriculture. The rest represented the cost of fertilizer, fuel, power, etc. and thus, ultimately, corresponded to incomes earned in some other industry. (Depreciation and indirect business taxes are also included in the gross value of product.) This is, perhaps, most clearly seen in the case of retail trade. The in-

comes earned (and therefore the value added) in retail trade are much smaller than the total of retail sales, the difference representing goods and services bought from wholesalers, manufacturers, and so on. This is a further example of the "duplicative transactions" mentioned in Chapter 2 in connection with our discussion of the equation of exchange.

The industries listed in Table 3 are not all affected by the business cycle in precisely the same way. The proportions of the national income accounted for by the different industrial groups change over the cycle and may also show a long-run tendency to change. Thus the proportion of the national income contributed by agriculture has declined markedly, and that contributed by manufacturing, trade, and service has risen during the past century.

The lower part of Table 3 classifies the national income by type of payment. (The same classification was also used on the left-hand side of Table 2.) As we should expect, wages and salaries are by far the most important form in which the national income is distributed. Here again there are important contrasts in the cyclical behavior of the various shares. We shall have occasion to refer to this fact at a later point.

TYPES OF EXPENDITURE ON THE GROSS NATIONAL PRODUCT

The entire current output of final goods and services is bought by many different purchasers, who are usually divided into four groups. These are consumers, business firms, government, and foreigners.

CONSUMERS' EXPENDITURES

In peacetime, consumers take by far the largest part of the current flow of newly produced goods and services.[11] In 1958, for example, consumers purchased new goods and services to the amount of 293 billion dollars, or about 66 percent of a total gross national product of some 442 billion dollars. Before World War II, con-

[11] Consumers' expenditures, as reported by the Department of Commerce, include certain items that do not involve current money payments by consumers. Of these, the most important is imputed rent on owner-occupied homes. Subsistence to members of the armed forces, payments in kind to employees, and goods withdrawn by proprietors (including farmers) for their own consumption are also included in consumption expenditures.

sumers' expenditures normally constituted about three quarters of the gross national product, but the fraction was considerably larger than this during the worst years of the Great Depression. During the last decade, the ratio of consumers' expenditures to GNP has been in the neighborhood of two thirds, but with some year-to-year variability. (See Table 5 on page 45.) We shall have to examine the reasons for these changes in a later chapter. The course of consumers' expenditures since 1929 is shown graphically in Figure 5.

The things consumers spend their money on can be classified in various ways. Economists have found it useful to classify the objects of consumers' expenditures according to their degree of durability. Thus, in Table 4 consumers' purchases are shown separately for nondurable goods (chiefly food and clothing), durable goods (for example, furniture, electric appliances, and, most important, automobiles), and services (rent, utilities, recreation, medical care, transportation, etc.). Nondurable goods have accounted for about half of total consumers' expenditures in recent years, but since World War II the percentage has been declining as the proportion spent on services has increased. The most volatile component is expenditures on durable goods, which fluctuate quite widely over the business cycle.

GROSS PRIVATE DOMESTIC INVESTMENT

A portion of the final output of goods and services is taken by business firms rather than by consumers. Such business purchases of current output are not the raw materials which eventually emerge as consumers' goods; these are not counted in the gross national product at all.[12] The part of the GNP sold to business includes only those goods that business firms themselves will use to replace or add to the stock of real capital held by business. For this reason, business expenditures on the GNP may be called "private capital formation" or "private investment." The objects of such expenditures are not consumed but replace or add to the nation's stock of income-yielding wealth.

Private investment may be either *gross* or *net,* depending on whether expenditures for replacement purposes are included or

[12] Except, as will be noted shortly, as they are retained by business in the form of increased inventories.

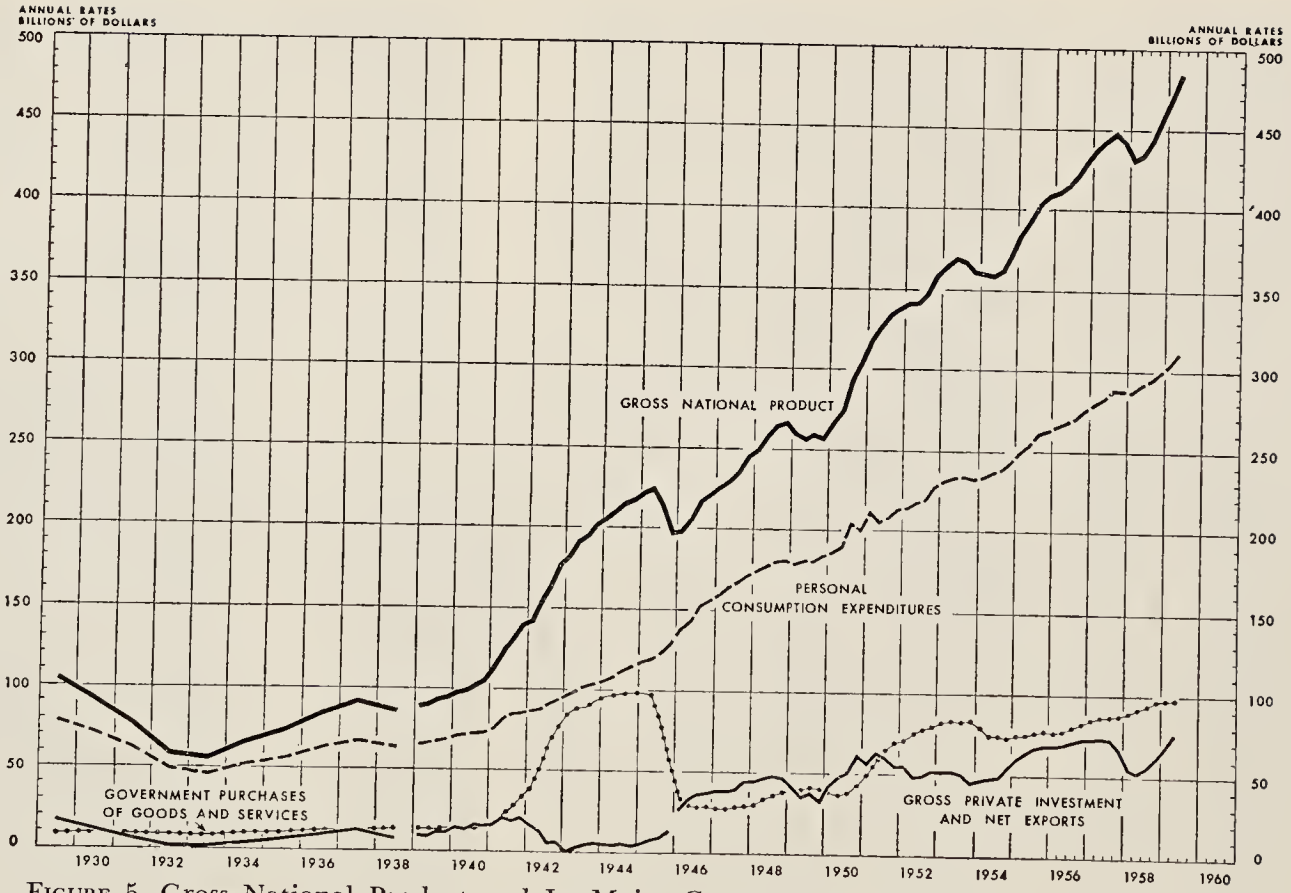


FIGURE 5. Gross National Product and Its Major Components, 1929–1959.

From *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics,* September, 1959. Based on Department of Commerce data.

TABLE 4. Gross National Product by Type of Expenditure, Selected Years, 1929–1959
(In billions of dollars) [a]

	1929	1933	1939	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
Gross national product	104.4	56.0	91.1	258.1	284.6	329.0	347.0	365.4	363.1	397.5	419.2	442.5	441.7	482.1
Personal consumption expenditures	79.0	46.4	67.6	181.2	195.0	209.8	219.8	232.6	238.0	256.9	269.9	284.8	293.0	313.8
Durable goods	9.2	3.5	6.7	24.6	30.4	29.5	29.1	32.9	32.4	39.6	38.5	40.3	37.6	43.4
Nondurable goods	37.7	22.3	35.1	96.6	99.8	110.1	115.1	118.0	119.3	124.8	131.4	137.7	141.9	147.6
Services	32.1	20.7	25.8	60.0	64.9	70.2	75.6	81.8	86.3	92.5	100.0	106.7	113.4	122.8
Gross private domestic investment	16.2	1.4	9.3	33.0	50.0	56.3	49.9	50.3	48.9	63.8	67.4	66.6	54.9	72.0
New construction	8.7	1.4	4.8	18.8	24.2	24.8	25.5	27.6	29.7	34.9	35.5	36.1	35.8	40.3
Producers' durable equipment	5.9	1.6	4.2	17.2	18.9	21.3	21.3	22.3	20.8	23.1	27.2	28.5	22.9	25.8
Change in business inventories	1.7	−1.6	.4	−3.1	6.8	10.2	3.1	.4	−1.6	5.8	4.7	2.0	−3.8	5.9
Net exports of goods and services	.8	.2	.9	3.8	.6	2.4	1.3	−.4	1.0	1.1	2.9	4.9	1.2	−1.0
Government purchases of goods and services	8.5	8.0	13.3	40.2	39.0	60.5	76.0	82.8	75.3	75.6	79.0	86.2	92.6	97.1
Federal	1.3	2.0	5.2	22.2	19.3	38.8	52.9	58.0	47.5	45.3	45.7	49.4	52.2	53.3
State and local	7.2	6.0	8.2	17.9	19.7	21.7	23.2	24.9	27.7	30.3	33.2	36.8	40.5	43.9

[a] From *Survey of Current Business*, July, 1959, and July, 1960. The 1957 and 1958 figures were slightly revised in 1960 but have not been changed here in order to keep them consistent with the other tables in this chapter.

not. The total of capital expenditures, including those that serve merely to maintain the existing capital stock intact, represents gross private domestic investment or capital formation. If an estimate of capital consumption (depreciation and obsolescence) is now deducted, the resulting figure is net investment or capital formation. In the Department of Commerce figures, the difference between gross and net investment accounts for part of the difference between gross national product and national income, the remaining difference being explained chiefly by indirect business taxes. If we are interested in the total capital expenditures by business, the gross figure is the appropriate one. If we are interested in net additions to the nation's stock of capital, net investment is the appropriate concept. Of the two, the estimate of net investment is less reliable statistically than that for gross investment because of the difficulties that arise in attempting to measure the amount of capital consumption during any period.

The major categories into which gross private domestic investment, or business expenditures on current output, are divided are those shown in Table 4—namely, construction (which can be divided further into residential and nonresidential building), producers' durable equipment, and net change in business inventories. Of these, the first two are much the most important. The great bulk of private capital formation takes the form of new buildings or of machinery and equipment.

In prosperous years, private gross capital formation accounts for around 15 percent of the total GNP. (See Table 5.) Put differently, and assuming that a day's labor produces the same amount of GNP in the capital-goods as in the consumers'-goods industries, about 15 percent of the American labor force in periods of prosperity depends directly on business capital expenditures for employment. Such expenditures, however, are made only when profit expectations seem to warrant them; they are therefore highly unstable. For example, they dropped catastrophically during the Great Depression of the 1930's, from about 16 billion dollars in 1929 to about one billion in 1932 and 1933. They have shown considerable instability even during the mild recessions we have had since World War II. Thus, in the short business contraction of 1957–1958, gross private domestic investment declined by about 25 percent. Obviously, the more heavily we rely for job opportunities

on the behavior of capital formation, the more unstable will be the economy. Yet we *must* rely on capital formation to perform two basic functions. First, it fills the gap in total income-generating expenditures created by the fact that individuals choose to save

TABLE 5. Percentage Distribution of the Gross National Product by Types of Expenditure, 1929–1959 (In percent of total GNP) [a]

Year	Consumption Expenditures	Domestic Investment	Net Exports	Government Expenditures
1929	75.6	15.5	0.7	8.1
1933	82.9	2.5	0.3	14.3
1939	74.2	10.2	1.0	14.6
1940	71.4	13.1	1.5	14.0
1941	65.1	14.4	0.9	19.7
1942	56.4	6.2	−0.1	37.5
1943	52.2	2.9	−1.2	46.0
1944	52.0	3.4	−1.0	45.7
1945	57.0	4.9	−0.7	38.8
1946	69.8	13.4	2.3	14.5
1947	70.6	13.4	3.9	12.1
1948	68.7	16.6	1.3	13.3
1949	70.2	12.8	1.5	15.6
1950	68.5	17.6	0.2	13.7
1951	63.8	17.1	0.7	18.4
1952	63.3	14.4	0.4	21.9
1953	63.7	13.8	−0.1	22.7
1954	65.6	13.5	0.3	20.7
1955	64.6	16.1	0.3	19.0
1956	64.4	16.1	0.7	18.8
1957	64.4	15.0	1.1	19.5
1958	66.3	12.4	0.3	21.0
1959	65.1	14.9	−0.2	20.2

[a] Department of Commerce data from *Survey of Current Business*, July, 1959, and July, 1960.

part of their incomes and not spend all on consumption. Secondly, without capital expenditures, economic progress toward a higher standard of living is impossible. A steadily increasing output per capita for an increasing population is almost certain to require constant additions to the nation's stock of capital goods.

Brief comment may be made on the components of gross private domestic investment shown in Table 4. Construction includes new

homes as well as construction for business purposes. The cost of a new house represents a capital investment rather than current consumption. Investment in new houses offsets savings just as do expenditures on business plant and equipment. The importance of residential construction as a factor in total private investment should not be underrated. It accounted for more than a quarter of total gross private investment in the middle 1920's, and the failure of residential building to revive adequately had much to do with the unsatisfactory performance of the American economy after the low point of the Great Depression had been reached in 1932–1933. In the 1950's, residential construction again accounted for about a quarter or more of total private investment.

Producers' durable equipment includes all types of industrial machinery, farm machinery and equipment, office and scientific equipment, transportation equipment (including tractors, trucks, railroad cars, buses, and aircraft), and so on. Purchases of such equipment considerably exceed the amounts spent on nonresidential construction; but if we add home building, all forms of private construction add up to a larger total. Expenditures on producers' durables are very sensitive to changes in general business and have shown considerable variation, even during the mild business cycles experienced since World War II. However, building follows a cyclical pattern of its own in which the swings are of considerably greater duration than those in general business activity. Construction fell relatively much more than expenditures on producers' durables during the Great Depression, and it was much slower to recover in later years. During the last half-century it has shown a less pronounced upward trend than expenditures on producers' durable equipment.

The most volatile item of all in private investment is represented by net changes in inventories. If firms have larger stocks of raw materials or finished goods on hand at the end than at the beginning of the year, this increase represents income-generating expenditures by business which have led to a (perhaps temporary) increase in the nation's stock of business wealth. As reported by the Department of Commerce, these inventory figures measure changes in the physical volume of inventories; the dollar figures are adjusted for changes in prices within the year. Since World War II, the annual net change in inventories has ranged between

a *decrease* of 3.8 billions in 1958 and an increase of 6.4 billions in 1946. In the first quarter of 1958, the *reduction* in inventories was at an annual rate of 6.9 billions, but this rate of decline did not continue through the whole year.

If we subtract the net changes in inventories from total GNP, we get that part of the total output of the economy which is purchased by "final buyers"—i.e., by those who do not intend ultimately to resell. The Department of Commerce publishes separate figures on "final sales," with a breakdown into durable and nondurable goods.[13] The series on "final sales" is more stable than that for total GNP, because it excludes the highly volatile element of changes in inventories.

NET EXPORTS

The third major category of expenditures on the gross national product listed in Tables 2 and 4 is "net exports of goods and services." This represents the net export surplus of the United States after deducting imports. Services as well as tangible commodities are included in both exports and imports. The value of exports represents, in effect, expenditures by foreigners on American output and clearly should be added into the GNP. Imports must be deducted in order to avoid overstatement, because expenditures by Americans on these imports are included elsewhere in the GNP statement, as consumers' expenditures, private domestic investment, or government expenditures. Imported foreign cars bought by American families are, for example, added into consumers' expenditures even though they are not part of the flow of goods and services produced in the United States. Deducting imports gives us a chance to correct for this overstatement.[14]

As can be seen from Table 4, the net export surplus is the small-

[13] The Department of Commerce provides a breakdown of total GNP by type of product. The categories are "goods output" (subdivided into durables and nondurables), services, and construction. By the nature of the case, construction activity and the production of services do not give rise to changes in inventories.

[14] Until a few years ago, the Department of Commerce deducted from the net export surplus the net flow of gifts and grants (public and private) from the United States to other countries. Thus, that part of the net export surplus financed by, for example, U.S. government grants was treated as a purchase by the American government, not as an expenditure by foreigners. The present procedure results in showing our entire net export surplus as a single item regardless of the way in which it is financed. Cf. *U.S. Income and Output,* pp. 57–58.

est of the four categories of expenditure on the gross national product. A peak figure of nine billion dollars was reached in 1947, when the foreign demand for American goods after World War II was at its maximum. Thereafter, our export surplus declined and even became negative in 1953. Under the Marshall Plan and subsequent aid programs, a substantial part of the American export surplus was financed by government grants to foreign countries. The export surplus was unusually large in 1957, but fell off again thereafter. Exports and imports tend to vary with business conditions at home and abroad, a matter that we shall look into further in a later chapter.

GOVERNMENT EXPENDITURES

We now come to the last of the four categories of spending on the gross national product—namely, expenditures by government, at the federal, state, and local levels. Government income-generating expenditures, whether made at the local or at the national level, are essentially of two types. In the first place, governmental bodies buy newly produced goods from private business. This much of current output is not covered by consumers' expenditures or private capital formation but is included in government spending. In addition, government pays incomes directly to the factors of production (primarily, salaries and wages to government employees) in exchange for services which are then passed on to consumers and business in the form of fire and police protection, national defense, education, soil conservation, and the multitude of other things that modern governments provide for their citizens. These services are not sold and therefore cannot be recorded at market prices as purchases by private groups. In effect, these services are included as output by government, valued at cost to government.

Thus, the government expenditures recorded in Table 4 show the total of government purchases from private enterprise and the sum of direct government payments to the factors of production. Virtually all of the latter represents government salaries and wages. In addition to these expenditures, the government makes substantial payments to individuals which are treated as "transfer payments" and are not included in government expenditures on new goods and services. Thus, social security benefits, direct relief, veterans' pensions, payments under the G.I. Bill of Rights, and similar types of

government disbursements are excluded from the GNP because they represent merely transfers of income (similar to gifts) rather than compensation for productive services performed. Interest payments on the government debt are also excluded from government expenditures on new goods and services and therefore from the GNP.

TABLE 6. Expenditures on and Disposition of Proceeds from the Gross National Product, 1939 and 1958
(In billions of dollars) [a]

Symbol	Item	1939	1958
C	Consumption expenditures	67.6	293.0
I_D	Gross private domestic investment	9.3	54.9
I_F	Net export surplus	0.9	1.2
G	Government expenditures	13.3	92.6
GNP	Total equals GNP	91.1	441.7
D	Less: Capital consumption allowances	7.8	37.9
T_B	Indirect business taxes	9.4	39.0
A	Other adjustments	1.1	−1.4
Y	Equals: National income	72.8	366.2
S_B	Less: Undistributed corporation profits [b]	0.5	6.1
T_C	Corporation profits taxes	1.4	18.2
S_S	Social insurance contributions	2.1	15.1
P	Plus: Government interest payments	1.2	6.2
	Government transfer payments	2.5	24.4
	Business transfer payments	0.5	1.7
Y_P	Equals: Personal income	72.9	359.0
T_P	Less: Personal taxes	2.4	42.6
Y_D	Equals: Disposable personal income	70.4	316.5
	Which was divided:		
C	Consumption expenditures	67.6	293.0
S_P	Personal saving	2.9	23.5

[a] From *Survey of Current Business*, July, 1959.
[b] Includes inventory valuation adjustment; that is, inventory profits arising from price changes are excluded from corporate profits.

Though shown separately, they are in effect treated as transfer payments on the grounds that, since the bulk of government debt was created to finance wars or current expenditures of past years, the interest on such debt does not represent a payment for a current productive service. The importance of these non-income-generating expenditures is suggested by the figures in Table 6. In 1958, government transfer payments plus interest on the government debt

amounted to 30.6 billion dollars, compared to government expenditures on goods and services of 92.6 billions.

Government expenditures are usually shown separately for the federal and for state and local governments. The enormous increase in government spending during World War II is indicated in Figure 5. During most of the war period, government expenditures absorbed nearly half of the entire gross national product. (See Table 5.) By 1947, the share of the GNP absorbed by government had fallen to about one eighth, although this fraction was considerably larger if government transfer payments are included. From 1948 on, the government's share of the GNP increased with the rise in military expenditures and foreign economic aid, and this rise was accelerated after the Korean crisis led to a new spurt in military expenditures. The government's share of the GNP was about one twelfth in 1929 and about one seventh in 1939. In the last few years, it has been about one fifth. It is to be remembered that state and local governments are included in these figures.

DISPOSITION OF THE PROCEEDS FROM SALE OF THE GNP

We have seen that national income may be defined in terms either of the value of goods produced or of incomes earned. The same is true of the GNP. Expenditures on the GNP by consumers, business, foreigners, and government are, in effect, the sales proceeds of the nation's output; and these proceeds are currently distributed to those who have a claim on current output. And, to complete the circular flow, these proceeds are sufficient, if spent, to purchase the current flow of goods and services represented by the GNP.

Table 6 provides the information necessary to trace the disposition of the proceeds from the sale of the gross national product. This table deserves careful study.

The first several lines of the table present the four components of total expenditure on the GNP. Using the symbols presented at the left of the table we may then write:

$$C + I_D + I_F + G = GNP$$

For simplicity, let us combine gross private domestic investment (I_D) and the net export surplus (I_F) and refer to the total as gross investment, to which we can give the symbol I.[15] We can then write

[15] We write I_F for the export surplus because that part not financed by gifts and grants represents foreign investment.

the preceding equation representing expenditures on the gross national product in the following somewhat simpler form, as we did in Chapter 2:

$$C + I + G = GNP$$

Let us now see how the receipts from the sale of the GNP were distributed. The first deduction is for depreciation and other capital consumption allowances charged by business firms (D in Table 6), which in 1958 amounted to 37.9 billion dollars. This amount was included in the price of products sold, and hence in the GNP, but is not paid out as incomes or counted as retained profits. Depreciation charges are a part of the *gross* saving of business and are available to finance the portion of private capital formation intended for replacement of worn-out or obsolete equipment and buildings.

We have already seen (page 34) why we must next deduct the two items that now follow in Table 6—namely, indirect business taxes (T_B) and other adjustments (A). Business taxes such as sales, excise, and property taxes are paid indirectly by the consumer in the prices charged for goods and are therefore included in the GNP; yet these amounts are turned over to government and are not available as income. We have already referred to the other adjustments (A), which are deducted at this point to secure the national income (Y). The most important of these are business transfer payments, which are added back in at a later point, and the statistical discrepancy.

The national income (Y) is the sum of all incomes earned by the factors of production in producing current output. However, not all of the national income is paid out to individuals and thus made available for consumption expenditures. The withholdings or diversions from consumption are listed in Table 6. Corporations withhold part of their net income as undistributed profits. These are savings by (incorporated) business; hence the symbol S_B. A substantial amount is paid to government as corporate income taxes (T_C). Since the national income was computed before income taxes, this deduction is obviously necessary if we are to arrive finally at the total of disposable personal incomes available for consumption expenditures. Finally, we must deduct the social security contributions (S_S) which business firms turn over to the government. These also are not paid out to individuals. What is left of the national income after these various "diversions" is paid out to individuals in the

form of wages and other labor income, rent, interest, dividends, and profits of unincorporated enterprises.

However, some individuals receive "incomes" that are not a payment for any productive service and therefore are not included in the national income. Table 6 lists three types of such transfer incomes, which must be added to the "earned" incomes received by the factors of production. These are government interest payments, government transfer payments (social security payments, pensions and other benefits paid to veterans, etc.), and business transfer payments. Of these, government transfer payments have been much the most important in recent years. Indeed, they have more than doubled in the last decade, and they can be expected to expand further in the years ahead. In Table 6 we have designated all of these forms of transfer income by the symbol P.

The total we now arrive at is personal income, or the sum of all income payments to individuals (Y_P). This represents the current flow of purchasing power put into the hands of individuals through the workings of the productive system and government activity. If we now deduct personal taxes (T_P), we secure one of the most important totals shown in the GNP statement. The remainder is the disposable income of individuals (Y_D). This is the amount that individuals as consumers have to spend or save. If personal saving is deducted (S_P), the balance represents the total of consumers' expenditures already shown at the top of the table.[16]

Thus, in 1958 a GNP of 442 billion dollars, after the series of "leakages" or diversions from consumption enumerated, resulted in a disposable income of 316.5 billions. A further diversion of 23.5 billions in the form of personal saving resulted in consumer expenditures of 293 billions. The total of the leakages, after the necessary adjustments, was exactly equal to the sum of private capital formation and government spending. The larger the sum of the leakages, with a given GNP, the smaller will be consumers' expenditures and the larger must be the other components of total expenditures.

The relation between disposable income and GNP and that between consumption and disposable income are obviously of great importance in studying the causes of changes in business activity. The ratio of disposable income to GNP is determined primarily by three

[16] In practice, disposable income and consumers' expenditures are directly estimated, and personal saving is obtained as a residual.

important variables: gross business savings, government tax receipts, and the volume of transfer payments. The first varies widely over the cycle; indeed, net business savings have ranged from substantial negative to large positive values. Government taxes vary with changes in the tax laws and the variation in the size and distribution of the national income. Transfer payments vary with changes in legislation and in the numbers of people entitled to various types of government payments. Corporate savings and total taxes show a positive correlation with the business cycle; transfer payments, chiefly because of relief payments and unemployment compensation, show some inverse correlation.

Between 1929 and 1940, consumption tended to be about 95 percent of disposable income in prosperous years and more than this in years of declining business activity and depression. (See Table 7.) In the worst years of the Great Depression, consumption was actually greater than disposable income. On net balance, consumers "dissaved"; that is, personal savings were negative. During the war years, personal savings were abnormally high, for a variety of reasons that will be discussed in Chapter 15. The percentage of disposable income saved dropped sharply after the war as consumers returned to prewar spending habits and also sought to satisfy pent-up demands carried over from the war. This was particularly evident in 1947. Since 1950, personal saving has ranged between 6 and 8 percent of disposable income, or somewhat more than in the prosperous years before World War II.

The long-run tendency toward some constancy in the ratio of consumption to disposable income is discussed further in Chapter 5. One implication of this relation, and of the preceding analysis of the GNP statement, immediately suggests itself. As the national income increases in absolute amount, so, too, will personal saving; and the larger this diversion from consumption, the greater must be the volume of private investment or government spending if aggregate demand is to be sustained. Business saving will also rise with the national income, creating further diversions to be offset by private investment or government spending. Conversely, if a rise in government spending threatens to create an inflationary situation, consumption can be curtailed by raising taxes, which will reduce disposable incomes, and by means which will induce consumers to increase personal saving.

TABLE 7. Relation between Consumption Expenditures and Disposable Income, 1929–1959
(In billions of dollars) [a]

Year	Disposable Personal Income	Consumption Expenditures	Personal Saving	Saving as Percentage of Disposable Income (percent)
1929	83.1	79.0	4.2	5.1
1932	48.7	49.3	−0.6	−1.2
1933	45.7	46.4	−0.6	−1.3
1937	71.0	67.3	3.7	5.2
1938	65.7	64.6	1.1	1.7
1939	70.4	67.6	2.9	4.1
1940	76.1	71.9	4.2	5.5
1941	93.0	81.9	11.1	11.9
1942	117.5	89.7	27.8	23.7
1943	133.5	100.5	33.0	24.7
1944	146.8	109.8	36.9	25.1
1945	150.4	121.7	28.7	19.1
1946	160.6	147.1	13.5	8.4
1947	170.1	165.4	4.7	2.8
1948	189.3	178.3	11.0	5.8
1949	189.7	181.2	8.5	4.5
1950	207.7	195.0	12.6	6.1
1951	227.5	209.8	17.7	7.8
1952	238.7	219.8	18.9	7.9
1953	252.5	232.6	19.8	7.8
1954	256.9	238.0	18.9	7.4
1955	274.4	256.9	17.5	6.4
1956	292.9	269.9	23.0	7.9
1957	307.9	284.8	23.1	7.5
1958	316.5	293.0	23.5	7.4
1959	334.6	311.4	23.3	7.0

[a] From *The Economic Report of the President*, January, 1960, p. 170.

RELATING THE TWO SIDES OF THE GNP STATEMENT

We have now looked at the two sides of the GNP statement—at the components of total expenditure and at the distribution of the proceeds among the various groups of recipients. The relationships described can be summarized in the form of a simple algebraic equation. Using the symbols given in Table 6, and for simplicity omitting

the adjustment item (A), we can express the two sides of the GNP statement as follows:

$$(C + I + G) + P = GNP + P = (D + S_B + S_P) + (T_B + T_C + T_P + S_S) + C$$

All that we have done here is to rearrange the items in Table 6. Transfer payments (P) are not a part of the GNP but are included in personal incomes; hence we have added them to the sum of expenditures on the GNP on the left-hand side of the equation. What is available for distribution, therefore, is shown on the right-hand side of the equation. Part of the proceeds are saved, either as gross business saving $(D + S_B)$ or as personal saving (S_P). Part goes to the government as indirect business and corporate income taxes $(T_B$ and $T_C)$, as personal taxes (T_P), and as social security contributions (S_S). What is left comprises consumption expenditures.

All of this can be summarized in the following simple statement: Out of the total proceeds resulting from the sale of the GNP plus the receipt of transfer payments, part is saved by business and individuals, part goes to the government in taxes and social security contributions, and only the remainder is spent on consumption.

These relationships are illustrated diagrammatically for the year 1958 in Figure 6. This is an elaboration of the circular flow diagram in Chapter 2, constructed to account for all of the items in Table 6. Thus, the GNP was 441.7 billions. Total spending of this amount was divided among consumers $(C = 293$ billions), domestic investment and net exports. $(I = 56.1$ billion), and government $(G = 92.6$ billion). This total plus 32.3 of transfer payments (P) was distributed by private enterprise and government in the upper part of the diagram.[17] Note that the transfer payments (P) circle around the box representing productive activity, since these payments are transfers and not expenditures on new goods and services. As we now follow the circular flow counterclockwise, we begin to encounter the diversions from consumption. A total of 114.9 billions is siphoned off by government in the form of taxes and social security contributions $(T_B + T_C + T_P + S_S)$. A further 66.1 billion goes into gross

[17] Business transfer payments create a slight problem here. In Table 6 they are included in GNP and then subtracted out as one of the "other adjustments" so that they are not included in national income. They are then added back in as one of the types of transfer payments.

business and personal saving ($D + S_B + S_P$ and, for simplicity, including the adjustment item, A).

Government expenditures on the GNP ($G = 92.6$) and transfer payments ($P = 32.3$) totaled 124.9 billion, or 10.0 more than the

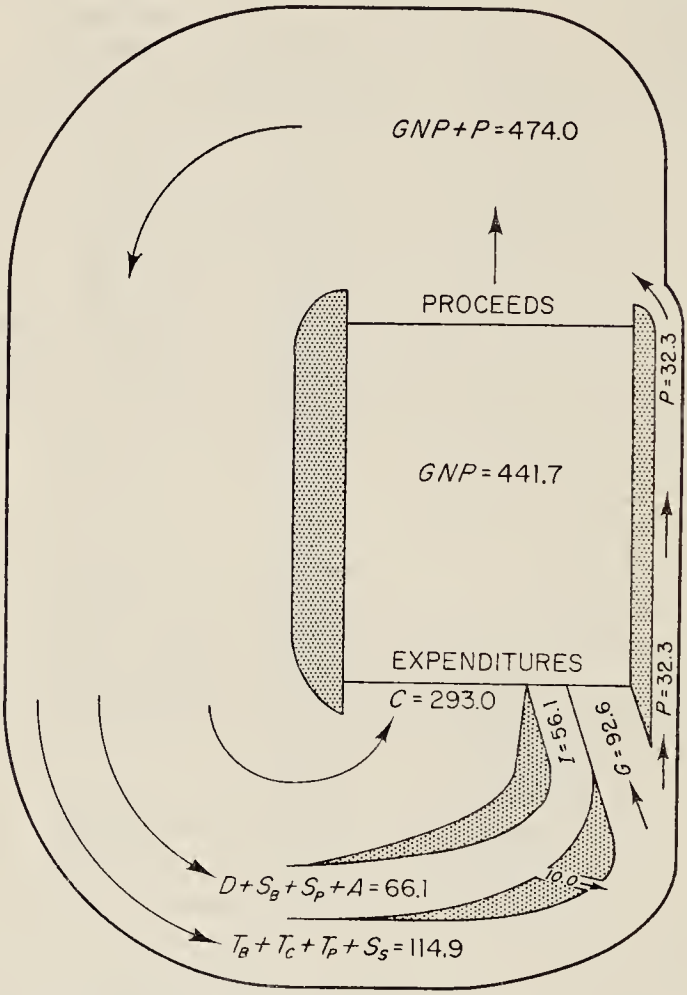


FIGURE 6. Circular Flow Representation of GNP and Its Components, 1958.

government collected.[18] Government then spent more than the total of all the tax diversions from the income stream that came into its possession. On the other hand, private capital formation (I)

[18] Of the total of 32.3 billion in transfer payments, 1.7 billion were paid by business rather than government. For simplicity, we have lumped business and government transfer payments together in the diagram, since we have no easy way of showing them separately.

amounted in all to a total of 56.1 billions, or 10.0 billions less than gross savings. This difference was exactly that by which government expenditures and transfer payments exceeded the tax diversions, and this is shown in Figure 6 as a diversion from the saving-investment pipe to government.

Or, to put it in more familiar language, government spending in 1958 exceeded tax receipts, and therefore there was a deficit, chiefly by the federal government. (1958 was a recession year, with the result that tax receipts fell off while government expenditure rose.) Government might have borrowed more or less than the 10 billions shown in Figure 6, and the money borrowed might have come from savers or might have represented new money created by the banking system. All that Figure 6 implies is what the equation on page 55 tells us. A government deficit (properly defined) is automatically matched by an excess of saving over private investment, and a government surplus is automatically offset by an excess of investment over saving. This follows necessarily from the basic identity expressed by the equation on page 55.

Returning to Figure 6, we see that total taxes and gross savings left available for consumers' expenditures a sum of 293 billion in 1958. This figure added to total investment and government spending yielded a total GNP of 441.7 billions, and thus we complete the circular flow. Had any particular diversion been less, consumer spending would have been more than 293 billion. With the same spending by business and government, this would have yielded a larger aggregate demand, a larger GNP, and higher money incomes. Out of these higher incomes some other diversion would have increased, if only temporarily, to preserve the equality between the total of all diversions and the sum of $I + G + P$.

In peacetime, individual savings are absorbed chiefly in private capital formation. During World War II, however, government expenditures were far in excess of tax receipts. Private investment was restricted to a very low level while individual savings rose, under the stimulus of expanding incomes and restraints on consumption, to unprecedented heights. Government borrowing absorbed virtually all savings, by individuals and by business. This does not mean that government bonds were sold only to individuals and business firms. A substantial fraction of government borrowing was from commercial banks, which created new bank deposits to purchase the

bonds. To the extent that the government borrowed from the commercial banks, and thus did not directly absorb savings, cash balances accumulated somewhere in the economy and were not spent. As a result, demand deposits and currency outstanding increased greatly during the war.[19]

Let us return for a moment to the equation on page 55. If we subtract consumption from both sides of the equation, we get the following fundamental identity: The sum of private investment, government spending on new goods and services, and transfer payments must be equal to the sum of all saving and all taxes, including social security contributions. Algebraically,

$$I + G + P = (D + S_B + S_P) + (T_B + T_C + T_P + S_S)$$

Now $(G + P)$ represents the sum of all government expenditures, not merely those that enter into the GNP, and this can be compared with total tax receipts $(T_B + T_C + T_P + S_S)$ to obtain the government's net deficit or surplus.[20] The preceding identity can therefore be rewritten as follows: Gross private investment plus the government's net deficit equals the sum of all private (business and personal) saving. In short, the net savings of the economy can be used in only one of two ways, either to finance private (domestic and foreign) investment or to offset a government deficit.

THE EQUALITY OF EXPENDITURE AND INCOME

It should be clear by now that the expenditures and receipts sides of the GNP statement or equation must be equal. This must be so because the accounts are constructed to make it so, in the same way that debits equal credits in any system of double-entry bookkeeping. Thus, the difference between the GNP and consumption is at the same time a measure of the sum of investment and government spending on new goods and services and also a measure of the sum of all diversions or leakages resulting from gross savings and taxes (minus transfer payments).

If we confine ourselves to gross saving and gross capital formation (investment), then in the GNP figures these two totals must be equal if taxes are equal to government expenditures. We have seen that total leakages (savings plus taxes) must equal $I + G + P$. If the tax

[19] This process is discussed in greater detail in Chapter 15.

[20] Subject to various bookkeeping adjustments, into the nature of which we need not enter here.

leakages (including social security contributions) just equal government expenditures including transfer payments, then investment (I) must equal gross savings of business firms and individuals.

We may adopt the economists' distinction between *ex post* (realized) and *ex ante* (expected or planned) magnitudes. Thus, in a given period, consumers may anticipate a certain income and may set out to save a certain amount based on this anticipation. If their actually realized income differs from that anticipated, realized saving will probably be different from planned saving. Businessmen, with certain anticipations at the beginning of the period, may plan 10 billions of real investment. If, however, an unexpected decline in demand causes an undesired accumulation of inventories, realized investment may turn out to be more than 10 billions.

Realized or *ex post* leakages must by definition be equal to the sum of actually realized $I + G + P$. If taxes exactly match G + P, realized saving and realized private investment must be equal. Let us illustrate this by turning again to the figures for 1958 in Table 6 and Figure 6. Suppose that business had succeeded in spending five billions more than it did for capital equipment, and that the savings and tax *plans* of consumers, business, and government were unchanged. The greater business spending would have meant higher incomes and a larger aggregate demand. In particular, business profits would have been higher. Business savings would have been larger; personal savings might have been somewhat more; taxes, on larger incomes, would have been moderately larger. The sum of these additional diversions, as we look back at the end of the year, would have equaled the additional investment.

Similarly, let us assume that during World War II the federal government had borrowed more than it did from the banks and to that extent had relied less on taxes or borrowing from current saving. Government spending would have remained the same but the tax diversion would have been less. Either the tax reduction would have been saved by business and individuals, or, if people had found ways of spending it despite price controls and rationing, somewhere in the economy prices and incomes would have risen and additional unplanned savings would have resulted.

Thus, so far as the statistics are concerned, saving does equal investment. This does not mean, however, that an equality of realized saving and investment implies any sort of equilibrium or stability in

the economy. Individuals and business may not want to retain the unplanned saving that results from an unexpected increase in incomes in the current month or year. If they attempt to spend more in the next period, aggregate demand and incomes then rise. The discrepancy between expectation and realization leads to further changes in aggregate demand and incomes. In 1958, for example, the equality of the realized magnitudes was achieved in the manner already indicated. But this was not a stable situation. During most of the year, incomes and spending intentions were such as to lead consumers and business to increase their spending as the economy recovered from the recession which reached a low point in April, 1958. Given their realized savings and incomes in each quarter, consumers and business both wanted to spend more than they were currently spending.

The GNP data available do not tell us anything directly about plans and expectations; they give us only the magnitudes actually realized. In analyzing the causes of economic change, the equality of realized leakages and government and business spending is less important than the plans for future saving and spending which the presently realized magnitudes generate. Thus, there is real purpose in studying, analytically if not statistically, the *ex ante* as well as the *ex post* versions of such concepts as saving, investment, and income. This is something that we shall look into further in the next chapter.

THE GNP IN CONSTANT PRICES

For many purposes, it is desirable to have a record of the changes in the GNP after correction for price changes. Correcting dollar figures for changes in prices, when we are dealing with complex aggregates whose components are constantly changing in importance, is a tricky matter. Fortunately, the Department of Commerce has done the necessary work and has published official estimates of the GNP expressed in dollars of constant purchasing power. Estimates of constant-dollar GNP, but only on an annual basis, were first published in 1951. In 1958, constant-dollar estimates on a quarterly basis were made available by the Department of Commerce.[21]

When prices change substantially, as they have done since the 1930's, it becomes particularly important that we have a measure of

[21] See *U.S. Income and Output;* also "Real National Output by Quarters—A New Major Economic Indicator," *Survey of Current Business,* December, 1958, pp. 10–15.

the change in "real" GNP and its components, i.e., after correction has been made for price changes. Thus, between 1939 and 1958, the undeflated GNP (not corrected for price changes) rose from 91 to 442 billions, an increase of nearly 400 percent. But prices more than doubled during this period, with the result that, deflated for this increase in prices, the GNP in 1958 was only a little more than double

TABLE 8. Gross National Product in Current and Constant Dollars, Selected Years, 1929–1959 (In billions of dollars) [a]

Year	Total GNP		
	In Current Dollars	In 1954 Dollars	Implicit Price Deflator [b]
1929	104.4	181.8	57.4
1939	91.1	189.3	48.1
1949	258.1	292.7	88.2
1950	284.6	318.1	89.5
1951	329.0	341.8	96.3
1952	347.0	353.5	98.2
1953	365.4	369.0	99.0
1954	363.1	363.1	100.0
1955	397.5	392.7	101.2
1956	419.2	400.9	104.6
1957	442.5	408.3	108.4
1958	441.7	399.0	110.7
1959	482.1	428.0	112.6

[a] From *Survey of Current Business*, July, 1959, and July, 1960, and *U.S. Income and Output*, Supplement to *Survey of Current Business*.

[b] This is the price index for the GNP implied by the ratio of the GNP in current prices to the GNP in constant prices.

that in 1939. (See Table 8.) A series for the GNP in constant dollars back to the 1890's is shown in Figure 1 (page 4).

Deflated or constant-dollar measures are now available on both an annual and quarterly basis for both the total GNP and each of its major components. A sample of the annual estimates is given in Table 8, along with the values of the deflating price index which is implicit in the constant-dollar estimates of GNP. The figures are expressed in terms of 1954 dollars; that is, they indicate, for any given year, what the GNP would have been in that year had all the included goods and services been valued at their 1954 prices instead of at the prices that actually prevailed in the year in question.

CHAPTER 4

INCOME AND THE PRICE LEVEL: THE INTERACTION OF AGGREGATE DEMAND AND SUPPLY

THE LAST 30 years have witnessed a significant increase in our understanding of the forces that determine the levels of spending, income, and employment at any particular time. This progress is part of the "Keynesian Revolution" in economic theory, which was touched off by the late Lord Keynes when he published his famous *The General Theory of Employment, Interest, and Money* in 1936. Although he had been anticipated in various respects by earlier writers, it is fair to say that the modern theory of income determination dates largely from the publication of Keynes' *General Theory*. As a result, economists now have at their disposal a set of new analytical tools that help us better to understand the causes of business fluctuations. In this chapter we shall see what these tools are and how they can be used to help explain why the national income and employment, as well as prices, are at a high or low level.

THE INTERACTION OF AGGREGATE DEMAND AND SUPPLY

In Chapter 2 we introduced the concepts of *aggregate demand* and *aggregate supply*. Just as demand and supply in the case of an individual commodity determine how much will be spent on that commodity and what its price and output will be, so the interaction of aggregate demand and aggregate supply determine the levels of spending, production, and prices for total output. And the level of production will determine, along with the productivity of labor, how many jobs will be available—that is, what the level of employment will be.

Starting with the general framework provided by Keynes, the mod-

ern theory of income determination does not pay much attention to aggregate supply but instead concentrates on the factors that presumably determine the level of aggregate demand—that is, the level of spending on newly produced goods and services.[1] Emphasis is put on finding an answer to the following question: Given any level of output and given the amount of income that will be earned by all factors of production in producing this output, what level of spending (aggregate demand) will result? The income (including normal profit) earned in producing a given output is the aggregate supply price of that output. This level of output will not be maintained unless aggregate demand (total spending) equals that supply price. Thus aggregate demand and supply are in equilibrium when total income (including normal profit) earned in producing a given output generates exactly the same level of spending, so that that output can be sold at prices that cover all income payments and leave a normal profit for businessmen. If, for any given level of output, aggregate demand turns out to be more or less than this equilibrium amount, business firms will try to reduce or increase output until an equilibrium between aggregate demand and supply is achieved.

Stripped to the barest essentials, this is a statement of the problem with which the Keynesian theory of income determination is concerned. The basic question is: At what level of output and employment will aggregate demand (i.e., spending) just equal aggregate supply price? Or, to put it somewhat differently, at what level of output will total spending be just enough to leave businessmen satisfied to continue producing that level of output and without strong motives either to expand or to contract production? Clearly, the answer to this question lies in the forces that determine the level of aggregate demand, and it is on a study of how aggregate demand can be expected to vary with different levels of output and employment that the Keynesian type of analysis concentrates.

THE DETERMINANTS OF AGGREGATE DEMAND

Let us restate our problem in the following way. Suppose that this month the gross national product is at some particular level—say, at

[1] Keynes' own exposition is to be found in *The General Theory of Employment, Interest, and Money,* 1936. For helpful introductory guides, see A. H. Hansen, *A Guide to Keynes,* 1953; Dudley Dillard, *The Economics of John Maynard Keynes,* 1948; and L. R. Klein, *The Keynesian Revolution,* 1947. A number of useful texts on the determination of income and employment are also available.

the annual rate of 500 billion dollars a year. How much spending (aggregate demand) will be generated by this level of income? Will just that amount of spending result next month, so that income and employment remain unchanged? Or will there by a different amount of spending, so that income and employment will be pushed up or down?

The first step toward an answer is an obvious one, and we have already taken it in Chapter 3. We can divide aggregate demand according to the main types of buyers who purchase the national output: consumers, business, and government. Thus, total spending on current output is made up of consumers' expenditures, gross private investment, and government purchases of new goods and services. What determines each of these three types of spending? In particular, what determines the amount of consumption and private investment? These two types of spending represent the contribution of the private sector of the economy to aggregate demand.

Let us review briefly the relations among consumption, investment, and total income that were brought out in Chapter 3. To simplify matters, we shall assume for the moment that there are no taxes or government expenditures. Then that part of total income which is not spend on consumption is saved. Saving represents a diversion of income away from the flow of spending on total output. If there are positive savings, then consumers alone do not spend enough to maintain the level of income.

This is why the relation between saving and investment is so important and why the theory of income determination is so concerned with the nature of this relationship. Let us assume that businessmen plan to spend on plant, equipment, and additions to their inventories an amount exactly equal to the total that individuals are currently saving. In this case, business spends on investment precisely the amount by which consumption falls short of total income. As a result, total spending (i.e., aggregate demand) is just large enough to maintain the existing level of income.

Suppose now that planned investment exceeds planned saving. Then business spending is larger than the amount by which consumption falls short of total income. This means that total spending is larger than it was before, with the result that the level of income is higher also. Investment has put back into the income stream more than the saving has taken out. Conversely, if planned investment

were less than saving, aggregate demand would decline and so would total income. In this case, saving takes out of the income stream more than is put back by investment.

It is a simple matter to bring the government into this picture. Not only saving but also taxes keep consumers' expenditures below the level of total income. Government expenditures as well as pri-

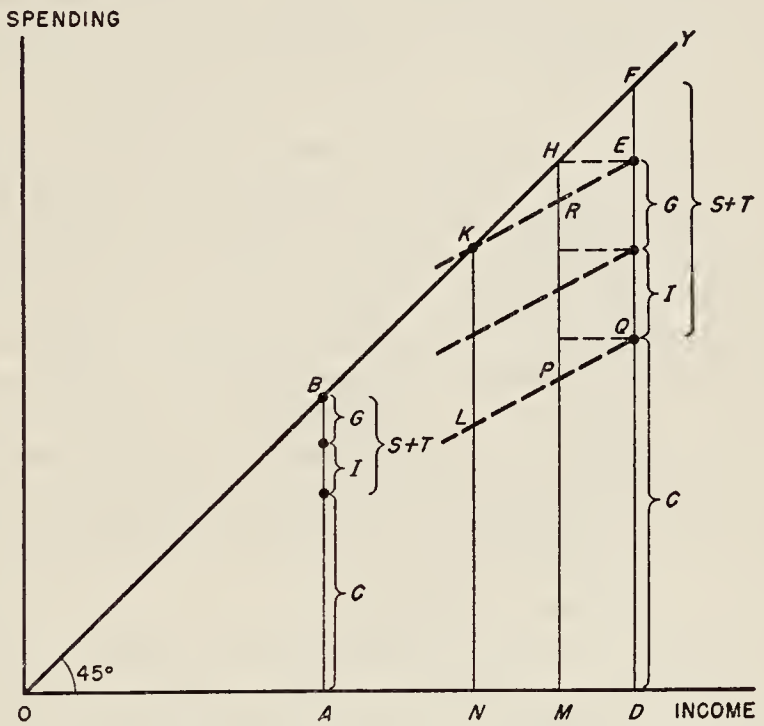


FIGURE 7. Spending and the Equilibrium Level of Income.

vate investment must be added to obtain total spending. Hence, for a given level of total income to be maintained, planned private investment plus government expenditures on new goods and services must be equal to the sum of all saving and all taxes.[2]

These relationships may be illustrated by a simple diagram. The horizontal scale of Figure 7 measures income, and total spending is measured along the vertical axis. Let us now draw the line OY at an

[2] To use the terms introduced in Chapter 3, we are obviously dealing here with *ex ante* concepts. *Ex post,* realized $I+G$ is *always* equal to realized $S+T$, for the reasons indicated in Chapter 3. In *ex ante* terms, however—that is, in terms of what people plan or try to do—$I+G$ may not be equal to $S+T$, and it is this discrepancy that causes the level of income to change.

angle of 45 degrees, so that any point on this line is the same distance from both the horizontal and vertical axes. Therefore, any point on *OY* represents the same amount of income and spending. It tells us how high spending must be to maintain any particular level of income marked off on the horizontal axis.

Let us assume that income is at the level indicated by the distance *OA*. Is this a stable or equilibrium situation, or will income immediately move to some new level? This depends on how much spending results at this level of income. Let us erect a vertical line at *A* to show the amount of spending that will occur at this level of income. The segment marked *C* represents consumers' expenditures, which, because of saving and taxes, fall considerably short of total income. There is a "gap" between total income (measured by *AB* as well as by *OA*) and consumption, and this gap is the total of all savings and taxes ($S + T$ in the diagram). To consumption we now add vertically two additional types of spending—investment, represented by the segment *I*, and government expenditures on new goods and services, shown by the segment *G*. We have assumed in this case that $I + G$ exactly equals $S + T$. Hence the total spending line *AB* is exactly equal to *OA*, the level of income. In this case, sufficient spending results to maintain income at the existing level—at least until something happens to change the amount of consumption, investment, or government expenditures.

Consider now a case in which spending is not sufficient to maintain a given level of income. Assume that income is at the level marked off by the distance *OD*. This level of income cannot be maintained unless aggregate demand is equal to *DF*—that is, unless an equal amount of spending is generated. Let us assume, measuring along the line *DF*, that at this level of income saving and taxes absorb the amount indicated by $S + T$ and that the remainder (*C*) is spent on consumption. To the consumption segment we now add vertically an assumed amount of investment (*I*) and government expenditures (*G*). The total of $C + I + G$ gives us the vertical distance *DE*, which is less than *DF*. In this case, planned (or *ex ante*) investment and government expenditures do not fully offset the diversions created by saving and taxes. The level of income *OD* has failed to generate an equal amount of spending, and therefore income (and output and employment) will decline.

How far will income fall? Let us consider two of the numerous

possibilities. First, we shall make the improbable assumption that all three types of spending remain unchanged as total income falls. In that case, income will decline only until it is equal to the unchanged level of spending. Since total spending remains unchanged, the unabsorbed saving and taxes (EF) are quickly wiped out as income falls. Total spending remains unchanged at DE. A vertical line equal in height to DE would intersect the line OY at H, and the total spending line MH ($=DE$) cuts the horizontal axis at M. OM, therefore, is the amount of income that can be supported by the level of spending $MH = DE$.

Consider now a more realistic assumption. As income falls, consumers' expenditures fall also, though not as rapidly as income. This assumed behavior of consumption is represented by the line LPQ. When income is OD, consumption is DQ; consumption falls to PM when income is OM and to LN when income is ON. We shall continue to assume that I and G remain unchanged as income declines. Under these conditions, what will be the equilibrium level of income? How far will income fall?

All we need to do is to add the assumed constant amount of $I + G$ to the consumption line LPQ, thus obtaining the line KRE (which is parallel to LPQ). The intersection of this line and the 45° line, at the point K, determines the new equilibrium level of income ON ($=NK$). Now let us see why this is so.

The line KRE tells us how total spending will change as income varies. If income falls from OD to OM, spending will fall from DE to MR. But MR is less than MH ($=OM$); there is still a "deflationary gap," measured by RH; and income will continue to fall until it is equal to ON. At an income of ON, an equal amount of spending is generated. This is the significance of the intersection of KRE and the 45° line, OY. Income will not fall further than this, because to the left of K the spending line is above OY. Any lower level of income would generate a level of spending greater than itself, and income would rise. The equilibrium level of income is ON ($=NK$).

We have here some highly useful tools that help us to explain why at any particular time the national income is at one level rather than another. And these tools can also be of use in getting at the forces that cause income and employment to fluctuate between relatively high and relatively low levels. However, we need to carry the analysis much further.

Two different kinds of questions need still to be answered. First, what determines the level of consumption and investment at any particular time? Second, what makes consumption and investment constantly change in amount, so that the level of income is continuously changing? We can restate these questions in terms of Figure 8, which is constructed in the same manner as Figure 7. The first type of question is concerned with the height and slope of the spending

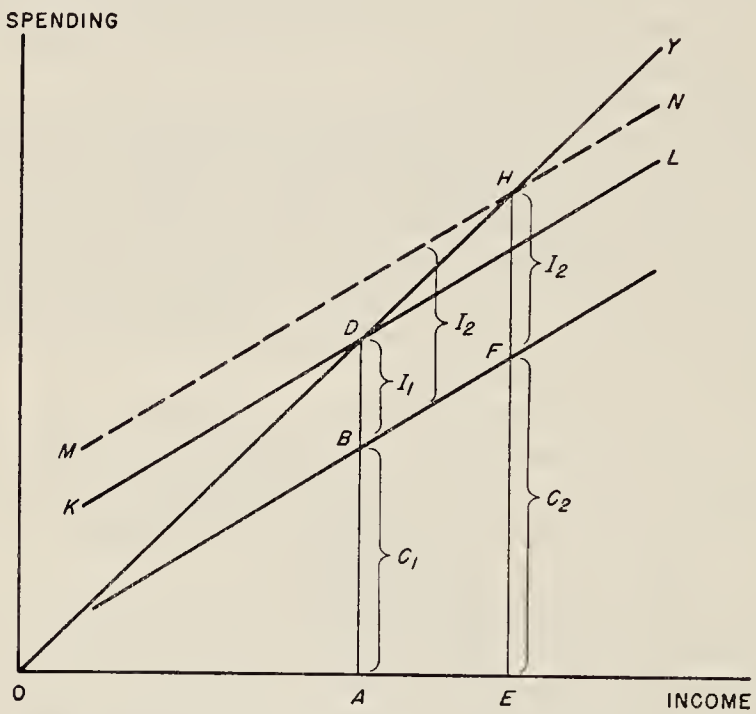


FIGURE 8. Effect of Change in Spending on Income.

line *KL*. At different levels of income, how high will *C* and *I* be, and therefore how large will be total spending? If we know this, we can then determine the equilibrium level of income. The second sort of question is concerned with a different problem—namely, why does the total spending line constantly shift its level, thereby bringing about continuous changes in the level of income and employment? In Figure 8, an increase in the amount of investment from I_1 to I_2 shifts the total spending line upward from *KL* to *MN* and pushes the level of income upward from *OA* (=*AD*) to *OE* (=*EH*). The first type of question asks: Why is aggregate demand what it is today? The second type asks: Why does the economic situation usually con-

tain within itself conditions that will make aggregate demand tomorrow different from what it is today? The first is concerned with the problem of static equilibrium; the second, with the problem of economic dynamics. When we study business fluctuations or problems of economic growth, we are dealing with a dynamic, not a static, situation. The modern theory of income determination, especially as developed by Keynes, began by being concerned almost entirely with the static problem; and this is still true of much of the literature in the field. But, since World War II, there has also been a growing interest in the dynamic aspects of the theory of income determination—in why and how changes occur in the level of aggregate demand (and aggregate supply). This interest has led to the development of various dynamic theories, or "models," that we shall look at later on, particularly in Chapters 6 and 13.

AGGREGATE SUPPLY AND THE PRICE LEVEL

It may have occurred to the reader that the analysis of aggregate demand in the preceding section told us nothing about how the price level is likely to behave as the level of income and spending rises or falls. Nor is there anything in the illustrative diagrams that we used in Figures 7 and 8 to give us any clue as to what level of prices is likely to be associated with different levels of income and spending.

This lack is characteristic of the Keynesian theory of income determination, which in its usual form excludes the price level from consideration. Keynes and his followers were concerned primarily with the way in which the level of aggregate demand affected the level of real output and employment. Their concern was with employment and with *real* income and output, i.e., with variables that had been corrected for price changes.

This emphasis is not surprising because the Keynesian analysis developed during the Great Depression when there was much unemployment and excess capacity. It could be assumed that aggregate demand (i.e., spending) could vary over a wide range without bringing about a significant change in prices. Within this range, aggregate supply was assumed to be elastic (i.e., responsive) to changes in spending, so that changes in spending brought about more or less proportionate changes in output and employment, with no significant changes in either prices or wage rates. Keynes

did not ignore the possibility of changes in prices or wages in response to shifts in aggregate demand; and indeed he emphasized the point that, if demand rose in the short run, prices would have to rise in response to increasing marginal costs.[3] But in general, neither he nor his followers chose to put much of their emphasis on the supply aspects of the problem.[4]

The United States and other countries have experienced a strong dose of inflation since World War II, and the resulting rise in prices has provided a stimulus to economists to try to work the price level into their general theory of income determination, so that we can explain at one and the same time both the level of output and the level of prices. This has not yet been done in a completely satisfactory way, in part because we are not sure how, in a world of large firms and powerful trade unions, aggregate demand and aggregate supply react on each other. This uncertainty is illustrated by the arguments that have been going on for some time as to whether, in recent years, inflation has been caused by "demand pull" or "cost push" or by some combination of the two.

In Chapter 2, we offered a brief discussion of aggregate supply and introduced a diagram (Figure 4) to show how short-run aggregate supply might shift with rising productivity and increasing supplies of labor and capital. We now propose to combine this analysis with the preceding discussion of aggregate demand to show how both the level of output and the level of prices are influenced by the same set of interacting forces—that is, by the interaction of aggregate demand and aggregate supply.

We can begin by turning to Figure 9. Figure 9a is simply the same sort of income-spending diagram we examined in Figures 7 and 8. OY is the 45° line, which is the locus of all points at which income equals spending. The lines D_1D_1', D_2D_2', and D_3D_3' are different total spending lines, only one of which would apply at any particular time. Each spending line, for example, D_2D_2', shows how the total of consumers' expenditures, investment, and government

[3] See, for example, *The General Theory, op. cit.*, p. 296; also Hansen, *op. cit.*, chap. 11.

[4] This was more true of his followers than of Keynes. The latter did introduce the concept of the aggregate supply function, and he devoted three chapters in *The General Theory* (chaps. 19–21) to the subject of wages and prices.

spending (which are not shown separately) would vary as income changes.

If aggregate demand varies in accordance with D_1D_1', the equilibrium level of income is OA $(=AB)$. If aggregate demand shifts upward to D_2D_2', the equilibrium level of income is OH $(=HE)$; and similarly, if aggregate demand shifts upward to D_3D_3'.

Now we need to be explicit regarding a matter about which we

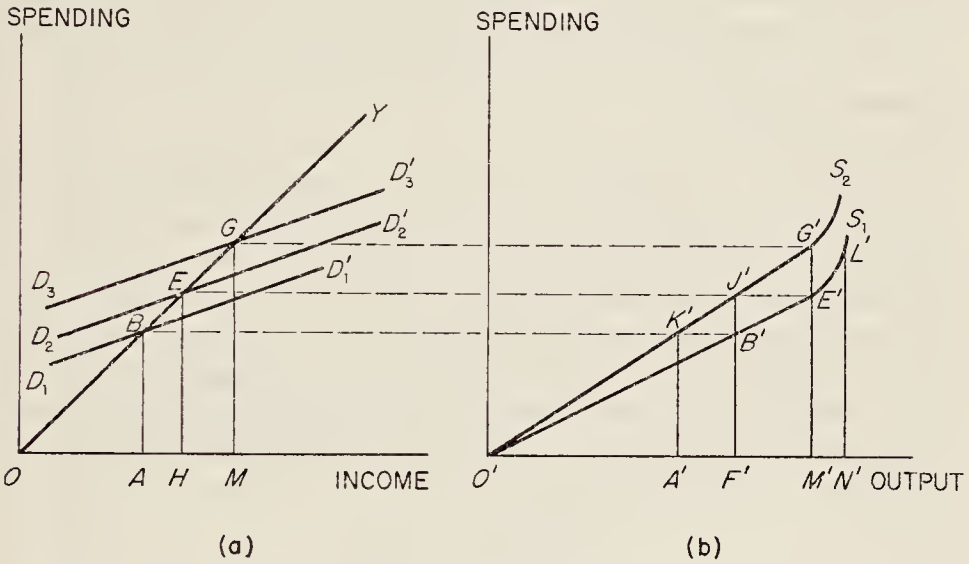


FIGURE 9. Determination of Output and the Price Level: Interaction of Aggregate Demand and Aggregate Supply.

have been purposely vague thus far. If spending and income are defined in money terms and without reference to any particular price level, then the equilibrium level of income does not uniquely determine the level of output and employment.[5] Thus on page 24 we wrote the simple identity:

$$C + I + G = (PT)_Y$$

$(PT)_Y$ is the equilibrium level of income as represented by, for example, OH $(=HE)$ in Figure 9a. But if $(PT)_Y$ is some figure, say 100 billion dollars, this might represent a large output sold at a low price level or a smaller output at a higher price level. In short, any

[5] Actually, Keynes carried on his analysis in terms of "wage units"; that is, he divided money income and expenditure by the wage rate paid to a standard unit of labor.

given money income can be obtained by an infinite number of combinations of quantities and prices.

Thus, the equilibrium level of (money) income cannot tell us anything about the equilibrium level of output until we know something about the price of the output on which that income will be spent. To complete the picture, we must turn to Figure 9b, which is an adaptation of Figure 4 on page 26. This diagram contains two possible aggregate supply curves, $O'S_1$ and $O'S_2$. Each curve indicates the amount of total spending needed (measured on the vertical axis) to induce any given amount of output (measured on the horizontal axis).

Let us consider the supply curve $O'S_1$. We have drawn it so that it is a straight line up to about the point E'. Over the range $O'E'$ each increase in spending brings forth a proportional increase in output. This is the same thing as saying that the price level is constant for all possible levels of output between zero and $O'M'$. But for larger levels of output, spending has to rise faster than output, which is to say that prices must rise—because of shortages of capacity, need to work overtime, etc. For any point on the aggregate supply curve, such as B', E', or L', the price level is given by the ratio of spending to output, that is, by such ratios as

$$\frac{B'F'}{O'F'}, \frac{E'M'}{O'M'}, \text{ or } \frac{L'N'}{O'N'}.$$

In Chapter 2, we made use of the equation of exchange

$$MV_Y = (PT)_Y.$$

For simplicity, let us use the symbol E (total expenditures for new goods and services) for MV_Y, and let us represent the price level by P and output by O, so that

$$E = PO, \text{ and}$$
$$P = \frac{E}{O}.$$

Thus, we can read the ratio $\frac{E}{O}$, and hence the price level, off the supply curve $O'S_1$ in Figure 9b. This ratio, and therefore the price level, is constant for levels of output between zero and $O'M'$ and rises for output levels beyond this.

The curve $O'S_1$ assumes some given level of wages and also a

given level of labor productivity.[6] Suppose that wages rise, the productivity of labor remaining unchanged, and that employers raise prices to compensate for some part or all of the increase in labor cost. To induce business firms to supply any particular level of output, spending will now have to be higher than before; the aggregate supply curve will shift upward and to the left, say, to $O'S_2$. It will now require, for example, total spending of $J'F'$ to bring forth the level of output $O'F'$; whereas before, spending equal to only $B'F'$ would have been required.

Thus, each of the $O'S$ curves in Figure 9b corresponds to a different level of prices (and labor costs). The higher any particular $O'S$ curve, the higher the level of prices associated with any particular output level.

Now let us combine the two parts of Figure 9. Let unit labor costs be at some level that determines the aggregate supply curve $O'S_1$. What will be the level of ouput? Let us assume that corresponding to this level of prices and wages there is a particular aggregate demand curve, D_2D_2', showing how total spending (consumption plus investment plus government expenditures) would vary with different levels of money income. The equilibrium level of income, where D_2D_2' cuts the 45° line, will be OH $(=EH)$. If we now extend a horizontal line across to Figure 9b, we see that it intersects the aggregate supply curve $O'S_1$ at E'. Thus, the equilibrium level of income OH $(=EH)$ will buy a quantity of output equal to $O'M'$, and the price level of this output will be the ratio $\frac{E'M'}{O'M'}$.

Suppose now there is an increase in wages which drives the aggregate supply curve up to $O'S_2$. It now requires spending of $G'M'$ to buy

[6] It also assumes that other cost elements and profit margins are related to output in some particular way. The curve bends upward to the right of E' because of short-run "diminishing returns" which are reflected in rising marginal costs. We assume that wage rates do not rise. Profit margins would rise if prices are equated to marginal costs, since average costs rise more slowly. The OS curves can also be drawn on the assumption that at some point wages begin to rise as a function of output. In this case, the OS curve would bend upward to reflect such wage (and price) increases. For another and somewhat different treatment of aggregate supply functions, see J. P. Lewis, *Business Conditions Analysis*, 1959, chap. 6. See also Sidney Weintraub, "The Keynesian Theory of Inflation: The Two Faces of Janus?" *International Economic Review*, vol. 1, May, 1960, pp. 143–155, and Paul Wells, "Keynes' Aggregate Supply Function: A Suggested Interpretation," *Economic Journal*, vol. 70, September, 1960, pp. 536–542.

the same output, $O'M'$. Will this amount of spending be forthcoming? This depends on what happens to aggregate demand.[7]

One possibility is that aggregate demand will not rise at all but will remain at D_2D_2' in Figure 9a. In this case, the equilibrium level of *money* income and spending remains at $OH \ (= EH = J'F')$, and this unchanged level of money spending will now buy only the smaller quantity of output equal to $O'F'$. Output falls from $O'M'$ to $O'F'$ because prices have risen while spending has remained the same. This result might arise if the monetary authorities saw to it, by a tight credit policy, that the supply of money did not expand and if, also, velocity did not increase. Thus, going back to the equation of exchange, we can recall that

$$MV_Y = E = PO.$$

If M and V_Y remain constant, E will be unchanged. But if P rises while E is constant, O will fall.

But another and more likely possibility is that aggregate money demand will rise to offset some or all of the shift in the aggregate supply curve from $O'S_1$ to $O'S_2$.[8] Suppose that, when the wage increase occurs, businessmen increase their payrolls in proportion so that there is no decline in employment. Suppose further that consumers, businessmen, and government increase their expenditures in proportion to the rise in prices. (If they are to do this, either there must be an increase in the money supply, or else firms and households must spend money that was previously idle—that is,

[7] The effect of changes in wages and prices on the level of aggregate demand is still a somewhat unsettled question among economists. It is a question to which Keynes gave some attention, and his treatment aroused considerable controversy. Involved are such issues as how prices change in relation to wages, the elasticity of the demand for labor, the effect of price changes on the behavior of interest rates, the behavior of the money supply, and how consumers and businessmen react to changes in interest rates and also to changes in the real value (i.e., purchasing power) of their holdings of money and other financial assets which result from a rise or fall in prices. For a brief introduction to this range of problems, see, for example, Hansen, *op. cit.*, chaps. 10–11; J. P. McKenna, *Aggregate Economic Analysis*, 1955, chaps. 11–13; Don Patinkin, "Price Flexibility and Full Employment," reprinted in American Economic Association, *Readings in Monetary Theory*, 1951, pp. 252–283; and Thomas Mayer, "The Empirical Significance of the Real Balance Effect," *Quarterly Journal of Economics*, vol. 73, May, 1959, pp. 275–291.

[8] As we pointed out in the preceding footnote, the effect of a change in wages and prices on aggregate money demand depends on a complex of factors that cannot be fully explored here.

velocity must increase.) This is equivalent to saying that the aggregate demand curve shifts upward in Figure 9a. If it shifts from D_2D_2' to D_3D_3', the new equilibrium level of income and spending will be OM $(=GM)$, and (looking across to Figure 9b) we see that this level of spending will continue to buy a level of output equal to $O'M'$, even at the higher price level represented by the curve $O'S_2$.

Actually, the aggregate demand curve in Figure 9a might shift upward more or less than this. If there is a general belief that the initial wage and price increases are an augury of still further increases to come, everyone may rush to buy before prices rise further; this may shift the spending line higher than D_3D_3', and output would then rise to some level higher than $O'M'$, which would bring us to the steeply sloping section of the curve $O'S_2'$, where capacity shortages will cause prices to rise still further. This may lead labor to demand further wage increases, causing another upward shift in the aggregate supply curve above $O'S_2'$ (not shown on the diagram), which may cause aggregate demand to shift upward again, and so on in an inflationary spiral.

What we have sketched out here is the interaction of the so-called "demand-pull" and "cost-push" forces that may operate to bring about a rise in prices. If the initial stimulus is a rise in aggregate demand that leads output to press on capacity, so that we reach the steeply rising part of the aggregate supply curve where prices must rise, this is a case of "demand-pull." The tight labor market may drive up wages and lead to a shift in the supply curve, but this results from the initial increase in demand. In contrast, we may have the "cost-push" case. Here the initial stimulus is either an increase in wages or an increase in prices by sellers in markets that are not perfectly competitive, even though there is not an antecedent expansion in aggregate demand. But, as we saw a moment ago, an increase in the price level arising in this way—that is, from the supply side—will lead to a decline in output unless monetary conditions permit a corresponding increase in total spending, i.e., in aggregate demand.

Figure 9 helps us to identify some of the conditions necessary if the economy is to maintain full employment without constantly rising prices. Aggregate demand must be high enough so that, given the conditions of supply, output is as large as possible without

running into the steeply rising section of the supply curve where prices would begin to rise. At the same time, labor and business must be induced not to use their economic power to push the supply curve upward, for if they do, prices will rise, and then aggregate demand must rise just to keep output and employment from falling.

FITTING ECONOMIC GROWTH INTO THE PICTURE

Actually, we should like to achieve not two but three goals: not only full employment and something close to a constant price level but also a satisfactory rate of growth. Some of the problems that arise in the attempt to achieve simultaneously all three of these goals are illustrated by Figure 10, which carries further the kind of analysis suggested by Figure 9 and by Figure 4 in Chapter 2.

Each of the three sections of Figure 10 is to be interpreted in exactly the same way as Figure 9. In each section, part (a) refers to aggregate demand; part (b) refers to aggregate supply.

Let us look at Figure 10A. Suppose that the economy is growing and that the aggregate supply curve in Figure 10Ab shifts from $O'S_1$ to $O'S_2$. Here we assume that productive capacity expands under conditions that do not lead to an increase in costs or prices (at least as long as aggregate demand does not exceed capacity under conditions of reasonably full employment). This occurs when the supply curve is extended out to the right without shifting upward; i.e., the segment $B'H'$ on $O'S_2$ is no steeper than the segment $O'B'$ on $O'S_1$.

Under what conditions are we likely to get this sort of growth in capacity without a change in the price level? In general, we can distinguish two main cases. (1) Wage rates, labor productivity, and profit margins remain unchanged while capacity and the labor force increase. (2) Labor productivity improves as the labor force and capacity expand, but wage rates rise just enough to offset the rise in productivity, profit margins remaining unchanged.

Figure 10Ab, then, refers to conditions of growth with constant labor costs (as long as full capacity is not exceeded). Now we have to ask: How fast do aggregate demand and aggregate supply expand? Some of the possibilities are illustrated in Figure 10A. Suppose we start in a given year with DD_1' and $O'S_1$, so that money income is OA and output is $O'A'$. Suppose that in the next year the rise in the labor force together with investment in new capacity

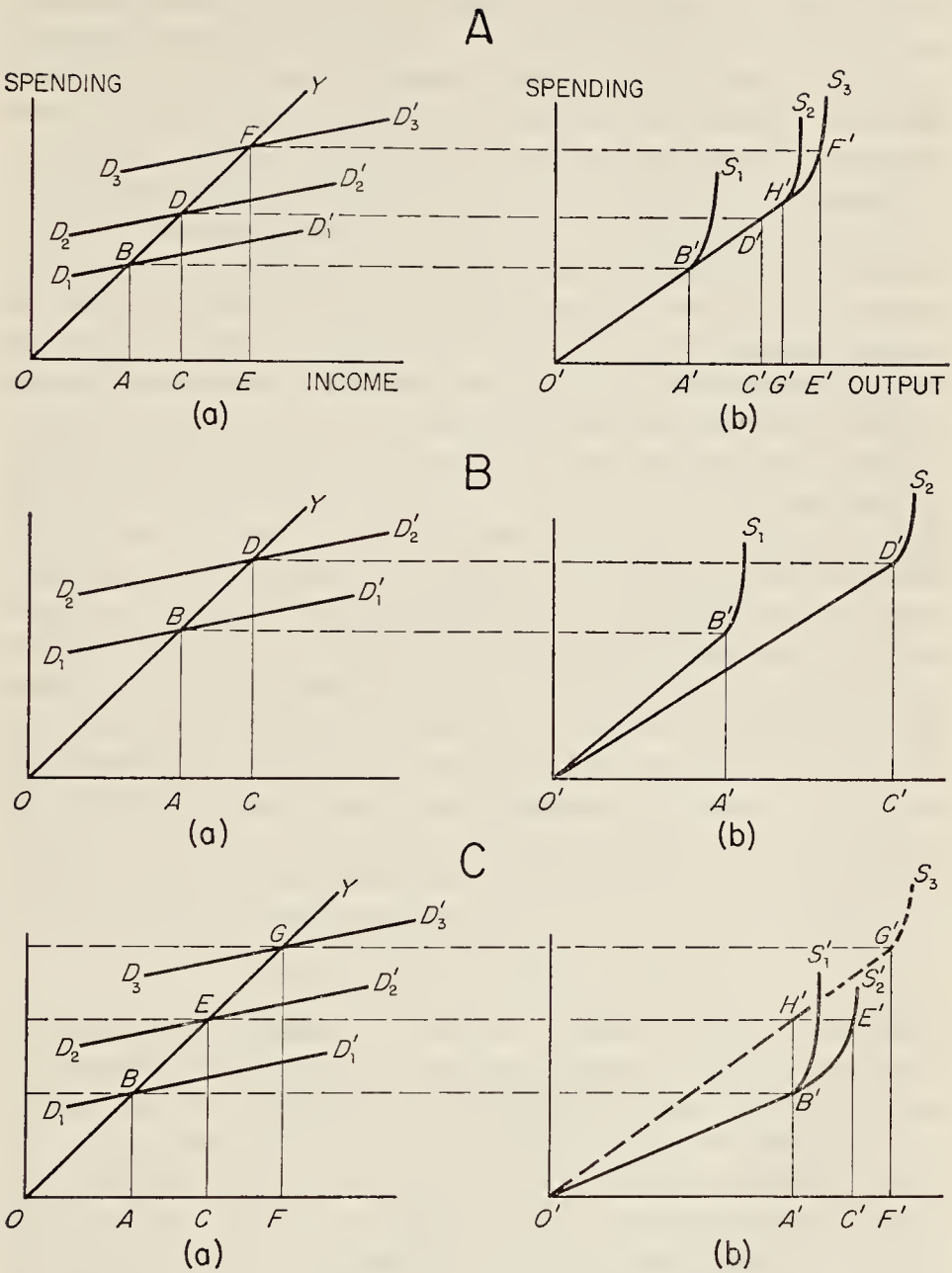


FIGURE 10. Growth and the Price Level Under Different Conditions of Aggregate Demand and Supply.

shift the supply curve to $O'S_2$ while the demand curve shifts upward only to D_2D_2'. Aggregate demand is now OC $(=CD)$, but this is sufficient to purchase only the output OC', whereas the full-employment output is something larger, about $O'G'$. In this case, aggregate demand has not risen fast enough to maintain full employment, although there has been some increase in output.

Suppose now that in the next year aggregate demand shifts to D_3D_3' while aggregate supply expands as indicated by $O'S_3$. Now income is OE $(=EF)$, but this level of demand is more than enough to purchase the new full-employment level of output at constant prices. FF' intersects the aggregate supply curve in its steeply rising section. In this case, prices rise because of "excess demand." Growth has been accompanied by rising prices because, in effect, aggregate demand has risen faster than aggregate supply.

There are other ways in which growth in output can be associated with changes in aggregate demand and supply. One of these is illustrated in Figure 10B. In this case, the aggregate supply curve shifts downward as it shifts to the right from $O'S_1$ to $O'S_2$. A larger output becomes available at a lower level of prices, since, for any level of output, $O'S_2$ is below $O'S_1$. This situation can develop if wages rise less rapidly than labor productivity increases (so that unit labor costs fall). Suppose that in this case aggregate demand rises from D_1D_1' to D_2D_2', while aggregate supply shifts from $O'S_1$ to $O'S_2$. Here we get growth and falling prices. We also get continued full employment, since aggregate demand, while it has not expanded as much as output, has still risen enough to purchase at the lower level of prices the full-capacity level of output.

Figure 10C illustrates one of the ways in which growth may be associated with price inflation. Suppose that aggregate demand rises from D_1D_1' to D_2D_2' while aggregate supply shifts only from $O'S_1$ to $O'S_2$. The new level of aggregate demand (CE) intersects the supply curve in its steeply rising section (at E'). This means that there is excess demand and that prices are higher. Suppose, as is likely, that this encourages trade unions to ask for wage increases larger than the improvement in productivity. (Business firms may also increase their profit margins.) The aggregate supply curve will therefore shift upward. If growth in capacity continues, we may have the new supply curve $O'S_3$; the full-employment level of output is larger than before but prices are also higher. Aggregate demand must rise to D_3D_3', with total spending equal to FG if full em-

ployment is to be maintained at this higher level of prices. That is, spending must be as high as FG $(=F'G')$ in order to buy the full-capacity level of output equal to $O'F'$. It is important to note also that this new and higher level of costs and prices may be maintained even if aggregate demand should later decrease—because, for example, the economy runs into a business recession. The aggregate demand curve might shift downward to D_2D_2' while the aggregate supply curve might remain unchanged at $O'S_3$. In this case, output would decline to $O'A'$. This is a case in which prices are "sticky" in the face of falling demand.

The hypothetical case just discussed suggests what has been happening in the United States since World War II and particularly since the Korean War. Inflation has accompanied economic growth. In boom years, aggregate demand has risen faster than supply. This has helped to bring about an increase in wages in excess of the average increase in labor productivity, and this has pushed up the whole level of costs and prices. During the years of mild recession that we have experienced, prices have not declined significantly. Thus, in terms of Figure 10, the aggregate supply curve has continued to move not only to the right but also upward, the upward movement being particularly rapid in boom years. Aggregate demand has also continued to rise in all but recession years. During the 1950's, aggregate demand rose rapidly enough to absorb at rising prices the output of a growing and steadily more productive labor force; yet in the late 1950's there was some excess capacity and the economy operated at somewhat less than a full-employment level.[9]

A SIMPLE KEYNESIAN MODEL

In the discussion thus far, we have talked about the interaction of aggregate demand and aggregate supply without having very much to say about what determines either side of this basic relationship. As we have already seen, the modern theory of income and employment, which stems from Keynes' *The General Theory of Employment, Interest, and Money,* puts its primary emphasis on the demand side. Since *The General Theory,* many economists have formulated theoretical "models" which summarize the factors determining the level of aggregate demand. We shall present one such model, a modification of the one Keynes originally developed.

[9] These developments are discussed in greater detail in Chapter 16. See also U.S. Congress, Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959).

We begin with the basic identity that total expenditures on new goods and services (aggregate demand) are made up of consumers' spending (C), private investment (I), and government expenditures (G), so that we can write:

$$C + I + G = Y. \tag{1}$$

We shall have to take the amount of G as being given. We treat it as an "exogenous" variable—that is, as something we do not try to explain. It is determined for us by what the government (at the federal, state, and local level) plans to do. This leaves us with the two other components of aggregate demand, which we do have to try to explain. We treat C and I as "endogenous" variables which, in ways we now have to examine, depend on yet other economic variables. Let us begin with consumers' expenditures.

The simplest assumption we can make here is that C depends merely on Y. Consumers' expenditures will rise and fall as income rises and falls. At this point, we can choose to be very general or very specific. If we want to be very general, we can write:

$$C = f_1(Y). \tag{2}$$

This says that consumers' expenditures are "a function of" income (this is the meaning of f_1). C depends on Y, but we do not say what the nature of this relationship is. As income rises, consumption might, for example, rise by a constant amount or by a decreasing amount, or some more complicated relationship might hold. The simplest specific relationship that we could assume would be that consumers' expenditures were always some constant fraction, a, of income. In that case, we could rewrite equation (2) as:

$$C = aY. \tag{2a}$$

In the next chapter, we shall examine the nature of the relationship between consumption and income in considerable detail.

Let us now turn to the other component of aggregate demand that we have to try to explain, i.e., private investment. Keynes assumed that investment could be expressed as a function of the rate of interest, and he called this relationship the marginal efficiency schedule of capital. This gives another equation:

$$I = f_2(i). \tag{3}$$

Here i stands for the rate of interest—i.e., the rate at which loan-

able funds can be borrowed for the purpose of carrying through investment expenditures. The symbol f_2 simply says that investment varies with the interest rate in some unspecified way. Presumably, the relationship is inverse; the higher the rate of interest, the lower the volume of investment, and vice versa. (This sort of relationship is illustrated in Figure 17, page 123.)

As it turns out, this type of investment relationship does not prove to be very helpful. At any particular time, investment can be considered to be related in a particular way to the interest rate only if a number of other factors are given. These include (1) the prices of the capital goods on which the investment expenditures are to be made and (2) what businessmen expect the return to be from making various amounts of new investment. Prospective returns from new investment depend, in turn, on such factors as the rate and kind of technological change, how rapidly demand is growing, the stock of capital already in existence, how confident businessmen are about the future, and a variety of other considerations.

Thus, this particular relationship between investment and the interest rate can be expected to shift, upward or downward, as conditions change. In Chapter 6 we shall see how these other conditions influence the level of investment and thus cause the marginal efficiency schedule of capital to shift. At any particular time for which we can take these other conditions as given, however, there will be a particular marginal efficiency schedule of capital, or relationship between investment and the rate of interest, such that the rate of interest determines the level of investment.

Therefore, we must explain what determines the rate of interest. Keynes did this by relating the interest rate to the supply of money. He called this relationship the "liquidity preference" schedule or function. For various reasons—to finance their ordinary expenditures, to have a reserve against contingencies, or because money is not subject to price fluctuations in the way that other financial assets are[10]—business firms and households desire to hold a certain amount of cash rather than other assets. How much they wish to hold depends, at least in part, on the interest rate. The higher the rate of interest that can be earned by putting money into income-yielding assets, the less cash (other things being equal) will people

[10] Keynes referred to these three motives for holding cash as, respectively, the transactions, precautionary, and speculative motives. For a useful introductory discussion of liquidity preference, see Hansen, *op. cit.*, chap. 6.

choose to hold. This is the idea behind the liquidity preference schedule.

Actually, as Keynes also pointed out, the amount of cash that firms and individuals want to hold will also depend, among other things, on the level of income and spending. The higher the level of spending, the higher must be the amount of cash (M) that people hold unless there is a change in velocity (V). For a given supply of money, the higher the level of income and spending, the higher will be the interest rate. Thus, we can say that the interest rate depends on both the supply of money and the level of income, and we can express the liquidity preference schedule as:

$$i = f_3(M, Y). \tag{4}$$

That is, the interest rate depends on (is a function of) both the stock of money (M) and the level of income (Y).[11]

Let us now bring together the four equations in our model. We have:

$$C + I + G = Y \tag{1}$$
$$C = f_1(Y) \tag{2}$$
$$I = f_2(i) \tag{3}$$
$$i = f_3(M, Y). \tag{4}$$

We have six variables—C, I, G, Y, i, and M. But two of them are taken as being determined by outside forces; these are government spending (G) and the supply of money (M). This leaves us with four unknowns (C, I, Y, and i) to be explained by four equations. Our model is therefore logically complete.

Put in its barest terms, this is the Keynesian general theory of aggregate demand or income.[12] There will be some level of income that will satisfy our four equations. The liquidity preference schedule (equation 4) determines the interest rate, which, given the

[11] The relation of the interest rate to liquidity preference and the supply of money is sometimes stated in the following, more precise way. Let the *demand* for money, M_D, which is the liquidity preference schedule, be expressed as:

$$M_D = f(i, Y).$$

But in equilibrium the demand for money must be equal to the supply, which is M, so that

$$M = f(i, Y).$$

This is equivalent to our expression, except that we have rearranged this equation to make the interest rate the "dependent" variable.

[12] It is also a theory of employment if we specify the level of wages.

marginal efficiency schedule, determines the level of investment.[13] The level of investment (I) and the given amount of government spending (G) can then be added to the consumption function to obtain the equilibrium level of income, as we did in Figures 7 and 8. This is illustrated further by Figure 11.

The line CC' is our consumption function, as determined by equation (2). The combination of the liquidity preference schedule and the marginal efficiency schedule—equations (4) and (3)—determine the level of investment to be added to the consumption function, giving us the line KL in Figure 11. Adding the given

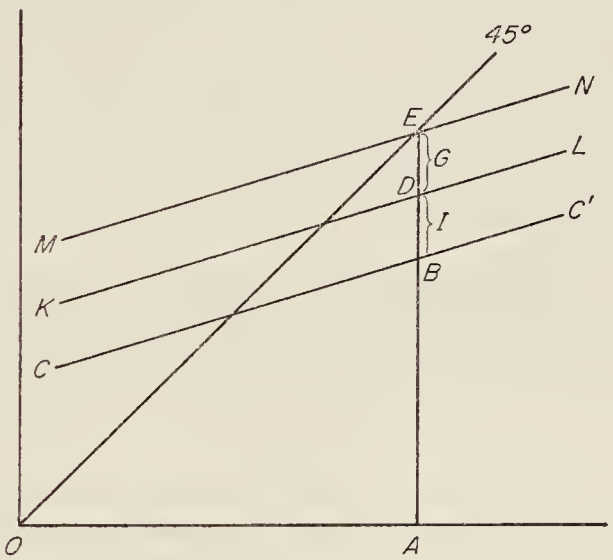


FIGURE 11. Equilibrium Level of Income in the Simple Keynesian Model.

amount of government spending (G), we get the total spending line MN. The intersection of MN and the 45° line is equivalent to that level of income which simultaneously satisfies all four of our equations. If we have the level of income OA ($=AE$), there will be no reason for it to change unless there is an upward or downward shift in one or more of the key equations or functions, or unless there is a change in one or both of the exogenous variables (G or M).

As we noted before, if we define our variables in money terms without respect to the price level, the equilibrium level of income

[13] There is an element of inexactness here that we have glossed over in order to keep the discussion as simple as possible. Since the interest rate depends not only on M but also on Y, the interest rate and the level of income are determined simultaneously. It is not really correct to speak of the interest rate being determined independently of the level of income.

is merely a dollar figure that does not tell us anything about the level of output or employment. We can take care of this by adding the aggregate supply function described earlier in this chapter. The aggregate supply function shows how output varies with income or spending, given the level of wages.[14] We can therefore write:

$$O = f_4(Y,W) \tag{5}$$

where W stands for the level of wages.

We do not try to explain the level of wages; we take it as being exogenous. We now have five equations and five endogenous variables, having added output (O) to the four we had previously.

Adding the fifth equation for aggregate supply is equivalent to doing algebraically what we did geometrically in the right-hand panel of Figure 9. We have added an aggregate supply function so that, with the determinants of aggregate demand given by the first four equations, we can solve for total output as well as total money income or spending.

We can go a step further and write another equation for the price level. This is nothing more than a definition: that the price level is determined by the ratio of spending to output—that is,

$$P = \frac{Y}{O}. \tag{6}$$

Here P stands for the price level. The other two variables we have already determined. Thus we wind up with six equations and six endogenous variables, including the price level.

It is important to note both the advantages and disadvantages of this way of looking at the determinants of the level of income, spending, and output. In an important sense, Keynes' great contribution to economic analysis—what is sometimes called the Keynesian revolution—consisted of centering economists' attention on the factors determining the level of aggregate demand and emphasizing that aggregate demand might settle at some point well below (or above) that necessary for a satisfactory level of employment. Before Keynes, it was generally assumed that aggregate demand would always naturally return to a full employment level, and that there was thus no need to make any detailed study of the determinants of the level of aggregate demand and employment.

[14] Other things must also be assumed given which we ignore for simplicity—particularly labor productivity.

Keynes expressed his determinants of aggregate demand in the terms summarized above because he began by separating aggregate demand into the two broad categories of consumption and investment. This is a highly significant way of dividing the demand for the national output. Clearly the forces determining consumption are different from those determining private capital formation. This division has led to the current emphasis in economic analysis on the relative behavior of consumption, investment, and saving.

All of this is important and useful. Indeed, we began the analysis of this book, and discussed the available data on the components of the gross national product, in terms of the extent to which diversions from consumption were likely to be offset by nonconsumption expenditures by business and government. This analysis has provided us with a valuable toolbox. But it has not explained the causes of cyclical fluctuations for us, nor does it by itself tell us much about the way an economy grows or why and how inflation occurs.

The chief weakness of the Keynesian system is its *static* character. It suggests some of the important factors that determine the level of aggregate demand at a particular time, but it does not tell us how and why the significant relationships change to create both economic fluctuations and growth. Dynamic analysis must be concerned with change, and this implies continuous shifting in some or all of the basic Keynesian functions.

Its other main weakness is that it tells almost nothing about the determinants of the most volatile part of aggregate demand, i.e., private investment. Economists have had little success in showing that investment expenditures are very sensitive to changes in the interest rate. Even if there is some slight sensitivity, the resulting marginal efficiency schedule shifts widely, both up and down, in response to a variety of other factors that cause businessmen to alter their views as to the presumed profitability of additional investment.

In Chapter 6, we shall take a more detailed look at the factors that seem to be important in determining the level of investment expenditures and why investment tends to fluctuate widely over the business cycle. But first we need to examine more closely that relationship in the Keynesian system which has stood up fairly well under the test of time, although some serious problems arise when we come to put it to practical use. This relationship, with which we shall be concerned in the next chapter, is that between consumers' expenditures and income.

CHAPTER 5

CONSUMPTION, INCOME, AND THE MULTIPLIER

THE LARGEST single component of aggregate demand consists of consumers' expenditures; and, as we saw in Chapter 4, the relationship between consumption and income plays a crucial role in the determination of the level of income and employment. This relationship has come to be called the *consumption function,* and the ratio of consumers' expenditures to income is frequently referred to as the *propensity to consume.* The consumption function describes the relation between consumers' expenditures and income as income changes. As we shall see in the discussion that follows, consumers' expenditures can be compared with various measures of income—for example, the GNP, the national income (Y), or disposable income (Y_D).

NATURE OF THE RELATIONSHIP

Figure 12 serves to bring out some of the possible characteristics of the relation between consumption and income. As in our earlier diagrams concerned with aggregate demand, the vertical scale measures spending (in this case, on consumption only) and the horizontal scale measures the level of income.[1] The various lines illustrate some of the ways in which consumers' expenditures and income might be related.

[1] Any of the variants of income (GNP, national income, or disposable income) may be used. While we shall keep the discussion as general as possible, most of the discussion in this and the next section implies that the horizontal scale measures disposable income. Later on, we shall look at the empirical evidence bearing on the relation between consumers' expenditures and both disposable income and GNP.

Suppose that the economy always consumed its entire income, however large or small it might be. In this case, the consumption-income relationship would be described by the 45° line OY. At every level of income—for example, the horizontal distance OA—consumption (measured by AD) would be equal to income. The *average propensity to consume*—defined as the ratio of total consumption to total income (i.e., $\frac{C}{Y}$)—would always be 1, or 100 percent. If we compare *changes* in income with the corresponding

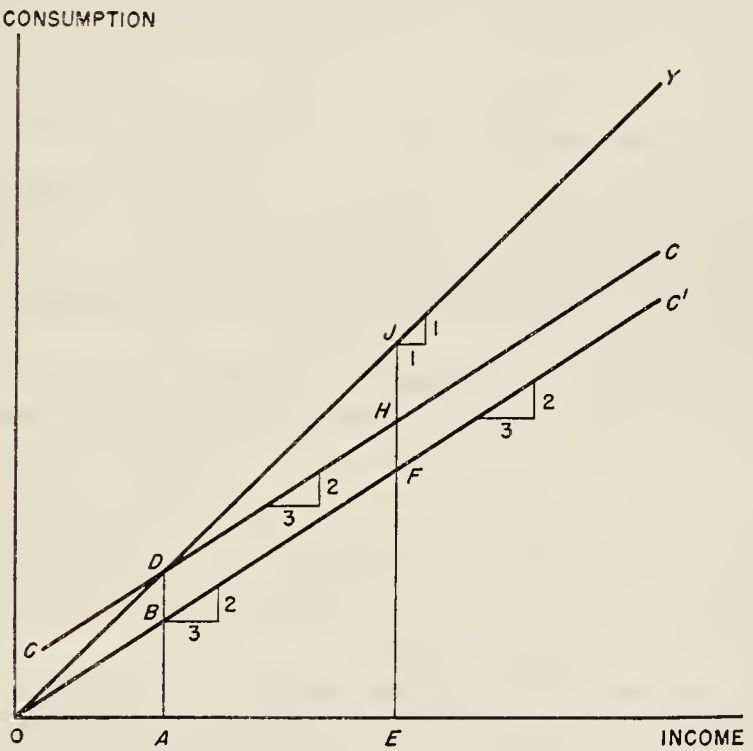


FIGURE 12. Relation Between Consumption and Income.

changes in consumption—which comparison gives us the *marginal propensity to consume*—we find that for every increase of a dollar in income there is an increase of precisely a dollar in consumption. The marginal propensity to consume, which can be expressed algebraically as $\frac{\Delta C}{\Delta Y}$, has the value of 1 at every point. This is illustrated by the small triangle erected at the point J. If ΔY is 1, then ΔC is also 1.

Actually, of course, consumption is seldom equal to total income, and ordinarily the total of saving and taxes keeps consumers' expenditures well below total income.[2] One possibility is illustrated by the consumption function OC' in Figure 12. As income rises to the right, consumption rises by a constant amount, but not as rapidly as income. The marginal propensity to consume is less than 1. In the case of OC', the marginal propensity to consume is $\frac{2}{3}$ at all levels of income. For every increase of three dollars in income, consumption rises by two dollars. Since this is true at all levels of income, we are dealing with a constant marginal propensity to consume $(=\frac{2}{3})$. The *average* propensity to consume is also constant and equal to $\frac{2}{3}$. Thus, the ratio of total C to total Y is measured by $\frac{AB}{OA}=\frac{2}{3}$, and this is the same as the ratio of C to Y at any other level of income—say, OE. The average propensity to consume will be constant and equal to the marginal propensity to consume if the consumption function is a straight line which passes through the origin. The latter condition means that consumption is zero when income is zero.

Actually, this is not a realistic assumption. At very low levels of income, consumption will be greater than income, and positive savings will not arise until some minimum level of income is reached. Thus, in Figure 12, the consumption function CC is in closer accord with the facts than is the line OC'. The latter yields positive savings even at very low levels of income, whereas the former yields negative savings up to an income of OA and positive savings only at levels of income greater than OA.

Let us now look at the relations between the marginal and average propensities to consume for a consumption function such as CC. The marginal propensity to consume is always measured by the slope of the line; and, since CC is a straight line, the marginal propensity is the same for all levels of income. We have drawn it parallel to OC', so that in this case the marginal propensity is also $\frac{2}{3}$. But how does the average propensity behave as income varies?

For all incomes from zero up to OA, the ratio of consumption to

[2] If the horizontal axis represents disposable income of individuals (which is computed after taxes), the only diversion is personal saving. If we use GNP as our measure of income, the relevant diversions include all forms of business and personal saving and all taxes.

income is greater than 1, since the consumption function lies above the line OY. At an income of OA, consumption is exactly equal to income, and the average propensity is therefore 1. For all higher levels of income, consumption is less than income, and the absolute amount of savings increases rapidly. It is also clear that the ratio of savings to income steadily increases, and the ratio of total consumption to income steadily decreases. Thus, the average propensity to consume falls as income rises, even though the marginal propensity to consume is constant. This is always true when the consumption function intersects the 45° line to the right of the origin.[3]

In each case thus far discussed, the marginal propensity to consume was taken to be constant for all levels of income, and the various consumption functions were drawn as straight lines. Another possibility is that the marginal propensity to consume declines as income rises. If this were so, the consumption function would take the shape of a curved line, concave to the horizontal axis, which rose by constantly decreasing amounts. As will be indicated later, however, the available evidence does not lend much support to the assumption that $\frac{\Delta C}{\Delta Y}$ declines as the national income or disposable income rises.

The importance of the consumption function for dynamic economic analysis should be immediately apparent. If, as income rises, consumption does not expand by the full amount of each increase in income, diversions are created which must be offset by private investment or government expenditure. In short, the consumption function states how diversions ("saving") will change as income varies and therefore how large I and G must be to maintain the level of aggregate demand consistent with each level of income. If we are comparing C and Y_D, the relevant consumption function

[3] The line CC in Figure 12 is described by the equation

$$C = a + bY$$

where C is consumption, Y is income, a is the amount of consumption at zero income, and b is the marginal propensity to consume. To obtain the average propensity, we divide C by Y. Thus

$$\frac{C}{Y} = \frac{a}{Y} + b$$

and this clearly declines as Y increases.

tells us how personal savings will change with variations in income and therefore how private investment and government expenditures must vary if these savings are to be offset and aggregate demand maintained. If we compare C and GNP, the relevant consumption function tells us how the total of all diversions from the flow of spending will tend to change as GNP varies. In this case, all forms of gross saving and all taxes are included; and the total of these diversions, at each level of income, measures the offsets $(I + G)$ required to maintain aggregate demand and income at that level.

THE THEORETICAL CONSUMPTION FUNCTION

It is necessary to distinguish among three kinds of consumption-income relationships: the *theoretical* and the *historical* consumption functions, and the *family* or *household* consumption function.

The theoretical consumption function expresses the relationship between alternative levels of income and the resulting amounts of consumption on the assumption that other conditions remain unchanged—that there are no changes in population, in the level and structure of prices, or in any other variables which might affect the consumption-income relation but which do not themselves necessarily change as a function of incomes. Like the demand and supply curves of economic theory, the theoretical consumption function expresses the relationship between two sets of alternative magnitudes on the assumption that nothing else changes. The theoretical consumption function, like the demand curve of price theory, is not directly observable. In any given period, income will be some particular amount, and with it will be associated some amount of consumption. But we cannot say for sure what consumption would have been had income been more or less than it was in that period. Suppose now that income rises 10 billions in the next period and that consumption increases 9 billions. This may or may not tell us something about the theoretical consumption function. If other conditions have changed—a rise in prices, a change in the availability of consumers' credit, a new war scare, or what not—the actual behavior of consumption will reflect not merely the rise of income along one consumption function but also a shift to a new relationship.

Such a situation is illustrated in Figure 13. In period one, in-

come is Y_1, the consumption function is the line C_1, and consumption is Y_1A. Let income rise in the next period to Y_2; but, because of some change in the situation, we assume that the theoretical or "true" consumption function shifts upward to C_2. Then consumption will be Y_2B. The empirical evidence will show only the two points A and B, but a line through them will not indicate how con-

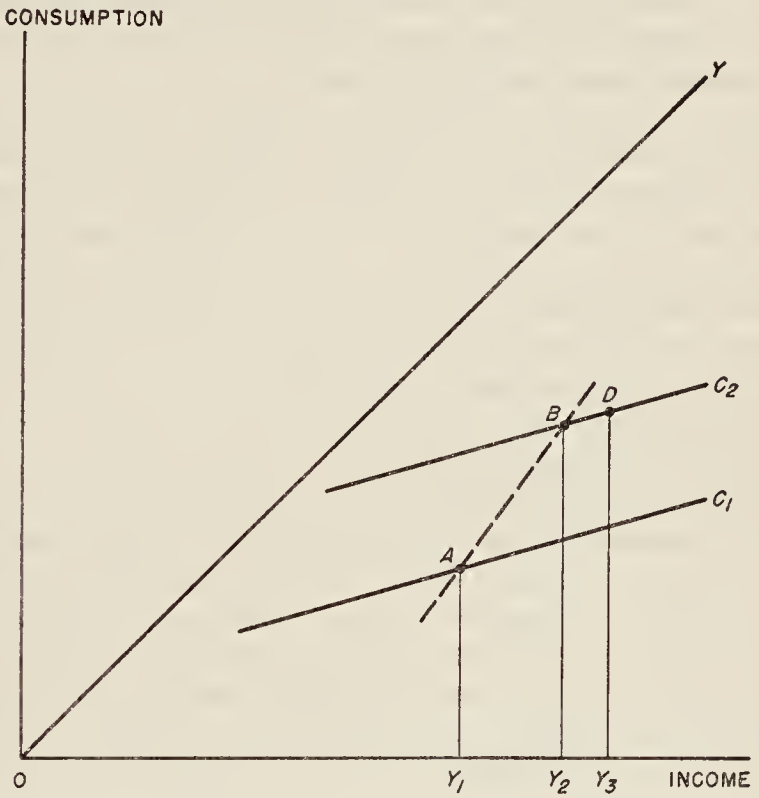


FIGURE 13. Effect of a Shift in the Consumption Function.

sumption reacts to changes in income if all other conditions remain the same. If income now increases to Y_3, but there is no further shift in the consumption-income relationship, then consumers' expenditures will be Y_3D and not some larger amount suggested by the dashed line AB. Since the short-run consumption function is constantly shifting, the actually observed data, as illustrated by the points A, B, and D in Figure 13, are imperfect guides to what the true consumption-income relationship is at any particular time. A line fitted as closely as possible to the points A, B, and D is what we referred to as the historical consumption function.

A variety of conditions determine the nature of the relationship between consumption and income. In this connection, it is important to distinguish between short-run and long-run relationships. It is also possible that there may be different short-run consumption functions, depending on the particular circumstances. The distinction between short and long run is concerned especially with the related facts that it takes time for people to adjust their spending habits to a change in income and that a standard of living once achieved is not readily given up.

We know that, over the decades, consumers have adjusted upward their standards of living as their incomes have risen, and the evidence suggests that, as a long-run tendency, the percentage of the national income saved has remained fairly constant since the closing decades of the nineteenth century.[4] Over long periods, as the national income gradually rises, standards of living rise more or less in proportion, and the economy continues to save about the same percentage of the national income. There is some evidence, then, that there is a long-run consumption function looking something like the line *OC* in Figure 14.[5]

The long-run relationship, however, is likely to be a poor guide to how consumption will react to short-period changes in income. It does not reveal the shape of the short-period function, and it does not tell us how this function is likely to shift about in response to all the influences that operate during the ups and downs of the business cycle. How the economy tends to respond to a gradual rise in income over a long period is different from the way it responds to rapid changes in income in the various phases of the business cycle.

What are some of the factors that influence the response of consumption to short-period changes in income? It has been argued that one important consideration is where the current level of income stands compared to the highest level of income achieved in

[4] Cf. Raymond Goldsmith, *A Study of Saving in the United States,* vol. I, 1955, chap. 1; Simon Kuznets, *National Income: A Summary of Findings,* 1946, pp. 52–54; and J. S. Duesenberry, *Income, Saving, and the Theory of Consumer Behavior,* 1949, chaps. 3–4.

[5] However, even if we can speak of a long-run consumption function, we must recognize that it may change its shape and position as important structural changes occur in the economy—for example, changed attitudes toward thrift, changes in the distribution of income and in the age distribution of the population, and so on.

the past.[6] According to this argument, people tend to adjust their standards of living upward as the general level of income rises secularly. But when incomes fall, consumers struggle to maintain the highest standard of living previously achieved, so that consumption falls relatively slowly as income declines. It has also been pointed out that, without reference to the previous peak in income, the amount of consumption resulting at a given level of income will tend to differ depending on whether income is rising or falling. One writer has listed a number of factors that determine the slope and level of the short-run consumption function. Among them are: consumers' expectations regarding such matters as their future incomes and employment prospects, recent price changes, the size of liquid asset holdings, and recent changes in the level and distribution of income.[7]

In the last few years, a new hypothesis regarding the relation between consumption and income has attracted considerable attention. We may refer to it as the "permanent income" (or "normal income") hypothesis.[8] According to this theory, a family's expenditures on consumption are determined not by its *current* income but by its *permanent* or *normal* income. Thus, current changes in income do not affect consumption unless these changes also affect the household's belief as to what its normal income is likely to be, averaged over a number of years. This approach offers an alternative explanation for the different behavior of the long-run and short-run consumption functions. The apparent frequent shifting

[6] Cf. Duesenberry, *op. cit.*, chap. 5; and Franco Modigliani, "Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting," in Conference on Research in Income and Wealth, *Studies in Income and Wealth*, vol. 11, 1949, pp 371–443.

[7] Cf. R. V. Rosa, "Use of the Consumption Function in Short Run Forecasting," *Review of Economics and Statistics,* vol. 30, May, 1948, pp. 91–105. For further discussion of the factors affecting the relation between consumers' expenditures and income, see, for example, Ruth Mack, "Economics of Consumption," in B. F. Haley, ed., *A Survey of Contemporary Economics,* vol. II, 1952; W. W. Heller *et. al.,* eds., *Savings in the Modern Economy,* 1953; and the references cited in the other footnotes in this section.

[8] See Milton Friedman, *A Theory of the Consumption Function,* 1957; and F. Modigliani and R. E. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data," in K. K. Kurihara, ed., *Post-Keynesian Economics,* 1954. For an excellent critical review of this type of explanation and additional references, see M. J. Farrell, "The New Theories of the Consumption Function," *Economic Journal,* vol. 69, December, 1959, pp. 678–696.

of the short-run consumption function is to be explained in good part by the fact that "normal" income moves more sluggishly than current income, and consumption is related chiefly to the former.[9] Over a long period, however, the trend in income can be taken as the actual course of "normal" income, and consumption can be expected to rise more or less in proportion to such normal income.[10]

It is fair to say that we are not yet able to evaluate precisely the influence of all of the factors that we have mentioned, and of others that we have not mentioned, on the way that consumption reacts to short-period changes in income. Some of the possibilities, however, are illustrated in Figure 14.[11] We assume that OC (which is not intended to be a 45° line) is the long-run consumption function as consumers gradually adjust their standards of living to a rising trend in the national income. Let us begin with an income of OY_1, and a consumption level indicated by Y_1C_1. If income now rises gradually to a new level at OY_2, consumption may rise along the long-run function to Y_2C_2. Suppose that a depression now forces income back to $OY_3 (=OY_1)$. Consumption will fall not along the long-run function to Y_1C_1 but back along a flatter short-run function to Y_3C_3. The upward adjustment of living standards when income rose to OY_2 means that consumption declines relatively slowly as income falls.

Suppose that income now starts to rise rapidly. Consumption will perhaps return along the consumption function A as far as C_2, or it may rise along some different function not shown on the diagram.[12]

[9] Thus, when current income falls, consumption will decline relatively slowly, because the decline in income is not expected to be permanent. Similarly for an unusually rapid rise in income. It should be added that this sort of explanation usually does not treat expenditures on consumers' durables as consumption but rather as investment in an asset which is then gradually consumed over the life of the asset.

[10] It is not an essential feature of this explanation that the long-run average propensity to consume (and ratio of saving to income) be constant.

[11] For somewhat similar diagrams illustrating some of the possible short- and long-run relations between consumption and income, see Duesenberry, *op. cit.*, p. 114; A. H. Hansen, *Business Cycles and National Income,* 1951, p. 165; and P. A. Samuelson, "Full Employment After the War," in S. E. Harris, ed., *Postwar Economic Problems,* 1943, p. 35.

[12] The more rapid the rise in income, other things being equal, the lower the level of consumption associated with any given level of income. Thus, if income recovers rapidly from the depression level OY_3, consumption may rise along a line which is at a lower level than the function marked A.

If income continues to rise rapidly beyond the previous peak level to OY_4, consumption does not necessarily rise along the long-run function OC. If the rise in income is rapid, spending habits may be slow to adjust themselves, and consumption may move along a short-run function such as the dashed lines D or E—say, to the level Y_4C_4.[13] Suppose now that income remains stable for a while. Spend-

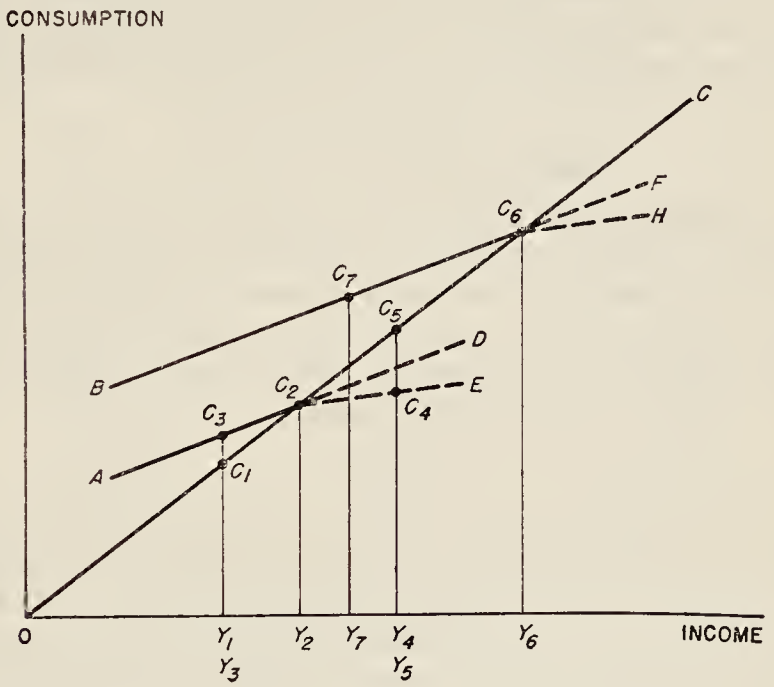


FIGURE 14. Possible Reactions of Consumption to Income.

ing habits will become adjusted to the new income level, and consumption will rise to Y_5C_5, even without a further rise in income. If there is a further *gradual* increase in income—say, to OY_6—consumption may rise along OC to Y_6C_6. If income should now decline because of a new business downswing, consumption will move back along the *short-period* function B to Y_7C_7.

[13] The shape and position of these lines as income rises above its previous peak levels depends on a number of circumstances—how rapidly income rises, the behavior of prices, changes in the distribution of incomes, and so on. If money income rises beyond the full-employment level, prices will rise, there will be little further increase in *real* income, and there may be significant changes in the relation between wages and profits. All of these factors affect the ability and willingness of consumers to increase their expenditures as total money income rises.

These illustrations are enough to drive home the point that we should expect considerable variation in the way consumption responds to short-period changes in income. A corollary may also be drawn that is highly important if we try to predict the future behavior of consumption from past experience. Although the historical relationship between consumption and income may tell us something about the underlying long-run relationship, granted that we have a long enough period and the relevant conditions have not changed too radically, it does not tell us very much about the character of the short-period consumption function.

THE EMPIRICAL EVIDENCE

The statistical material available regarding the relation between consumption and income is of two sorts. We have the historical record of the relationship between the national totals of income and consumption in different years, and there are also family budget studies which show how much, on the average, is consumed and saved at different levels of family income.

THE FAMILY CONSUMPTION FUNCTION

Household budget studies reveal a consumption-income pattern which we may call the family consumption function. These budget studies each deal with a sample of families with different incomes in a particular period. They thus eliminate the difficulty of shifting relationships over time. They tell us how, with a given distribution of incomes, consumption and savings vary with size of family income at a given time and place.[14]

An example of the sort of consumption-income pattern that emerges from family budget studies is given in Table 9. These figures, which are based on data of the United States Bureau of Labor Statistics, show what percentage of income is saved at different levels of family income.[15] According to these figures, households with an income of less than $1000 in 1950 spent some 80 percent more than their income, and expenditures were slightly in excess of in-

[14] For references to such studies and what they tell us about how consumption and saving vary with family income, see Dorothy S. Brady, "Family Saving, 1888 to 1950," in Goldsmith, *op. cit.*, vol. III, pp. 139–273; Irwin Friend and Stanley Schor, "Who Saves?" *Review of Economics and Statistics*, vol. 41, May, 1959, part 2; and M. R. Fisher, "Exploration in Savings Behavior," *Bulletin of the Oxford University Institute of Statistics*, vol. 18, August, 1956, pp. 201–277.

[15] Only urban families are included.

come for those with an income from $1000 to $3000. Positive savings emerge only for incomes above $3000. Above this level, the percentage of income saved rises rapidly as we move to higher levels of family income. Other studies, for earlier and later years, also emphasize this same tendency: as the size of family income increases, the percentage saved increases rapidly. Or, to put it differently, the average propensity to consume declines as the size of family income rises.

Unfortunately, family consumption functions are of limited use-

TABLE 9. Ratio of Saving to Income by Income Class, 1950 [a]

Income (after taxes)	Saving-Income Ratio (percent)
Under $1,000	−81.7
$ 1,000–1,999	−6.2
2,000–2,999	−1.7
3,000–3,999	2.4
4,000–4,999	4.5
5,000–5,999	6.5
6,000–7,499	10.0
7,500–9,999	16.3
10,000 and over	30.7

[a] From Irwin Friend and Stanley Schor, "Who Saves?" *Review of Economics and Statistics*, vol. 41, May, 1959, part 2, p. 232.

fulness in determining the relation between aggregate saving and aggregate income for the country as a whole. There is fairly conclusive evidence that the family consumption-income relation tends to shift upward secularly as the national income rises. It is true that, as the total national income rises, most families move to a higher income bracket. But this does not mean that each family, in its improved position, will save as much as a family with that size income did earlier when all incomes were lower. How much a family saves depends in part on the incomes of other families and on the standards of decency and comfort that are created by a given level and distribution of incomes. Or, to put it differently, the percentage of income consumed seems to depend to some degree on a family's *relative* position in the income scale rather than on the absolute size of its income.[16] If the total national income gradually rises but

[16] Cf. Duesenberry, *op. cit.*, and Dorothy S. Brady and Rose D. Friedman, "Savings and the Income Distribution," in *Studies in Income and Wealth*, vol. 10,

the shape of the income distribution does not significantly change, the percentage of total income saved is likely to remain approximately constant, even though every family has a higher income than before. Thus, although the family consumption function displays a declining average propensity to consume for families at different income levels at any one time, this does not mean that, for the economy as a whole, total consumers' expenditures will behave in this way as total income rises.

THE HISTORICAL CONSUMPTION-INCOME RELATIONSHIP

Let us now turn to a study of the historical behavior of consumption and income. Over the past years, how have total consumers' expenditures reacted to changes in income, and what if anything does this historical relationship tell us about the way consumption may be expected to respond to future changes in income, both in the short run and in the long run?

As already noted, aggregate consumption can be compared with various measures of aggregate income—for example, GNP, net national income, income payments, or disposable income. Which we use depends on which of the diversions from consumption we wish to study. In the remainder of this section we shall study the behavior of consumption in relation to disposable income (Y_D) and to GNP. The latter comparison tells us something about the behavior of total diversions; the former tells us something about a single, homogeneous type of diversion—namely, the savings of individual income receivers. If we are interested in a "psychological propensity to consume," which is related to personal spending habits, it is the comparison between Y_D and C that is particularly relevant. In practice, consumers' expenditures are usually related to disposable income.

Figure 15 depicts graphically the historical relationship between

---

1947, pp. 247–265. James Tobin has produced evidence that suggests that *relative* family income may be less important, and the *absolute size* of income more important, in determining the percentage of income saved than the studies by Duesenberry and Brady and Friedman seemed to indicate. See his paper, "Relative Income, Absolute Income, and Saving," in *Money, Trade, and Economic Growth: In Honor of John Henry Williams*, 1951, pp. 135–156. For further discussion of these and other factors that may affect the relation between family income and saving, see Brady, *op. cit.;* the references in footnote 8, above; also Heller, *op. cit.*, chap. 13.

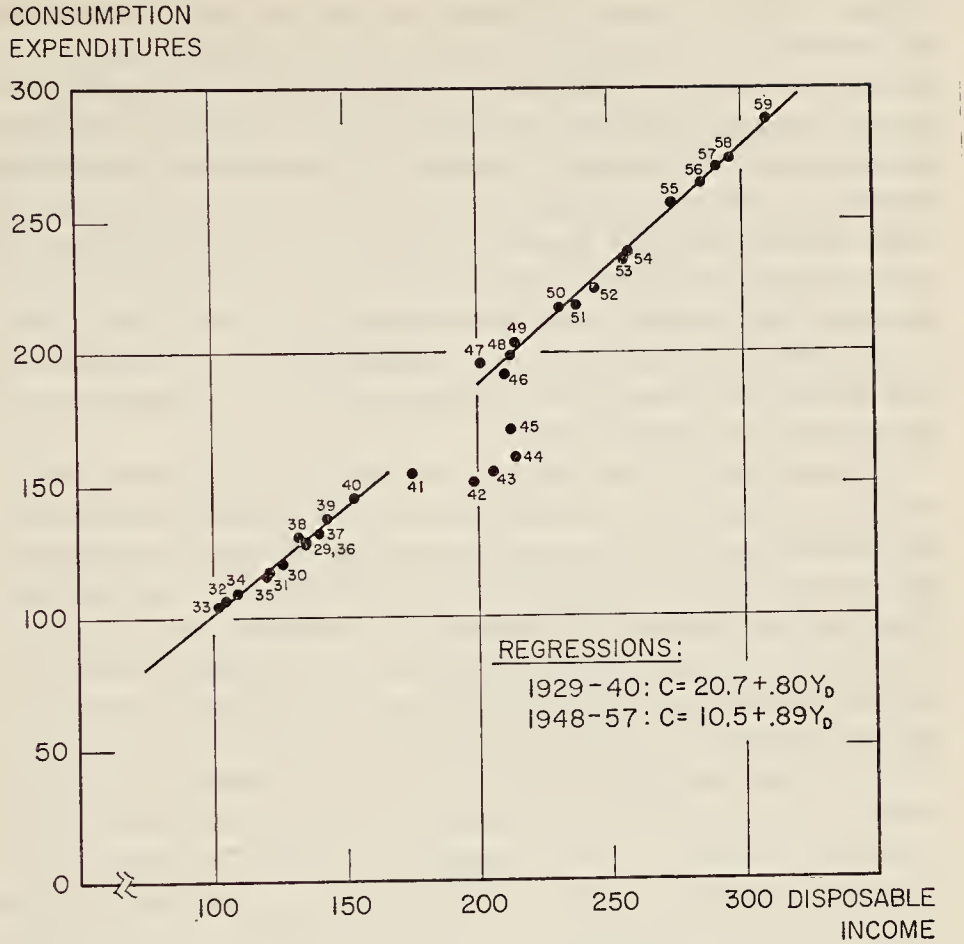


FIGURE 15. Relation Between Consumers' Expenditures and Disposable Income, Annually, 1929–1959.

Based on Department of Commerce data.

consumers' expenditures and disposable income for the period 1929–1959. The vertical scale measures consumers' expenditures; the horizontal scale, disposable income. Both series are expressed in dollars of constant (1954) purchasing power. Each point on the chart is identified by the year to which it refers. Thus, to take one case, the point for 1959 indicates that in that year consumers' expenditures were about 288 billion dollars (in 1954 prices), and disposable income was about 310 billion dollars.

It is clear from Figure 15 that at least three distinct periods must be demarcated in describing the historical relationship between consumers' expenditures and disposable income since 1929. From

1929 to 1940, the points fall closely around a straight line, which we can refer to as the prewar historical consumption function. During the war years, consumption increased much less rapidly than income and hence fell considerably below the prewar relationship. Beginning in 1945, consumption began to move upward toward the prewar relationship. For the years after 1946, the relationship between consumption and income can again be described fairly well by a straight line, but one which seems to be at a slightly higher level and with a slightly steeper slope than that describing the prewar relationship. At least, this is suggested by the two regression lines on the chart, one of which has been fitted to the prewar and the other to the postwar data.

The slope of the line fitted to the data for 1929–1940 yields a marginal propensity to consume of about 0.8. That is, an increase of 10 billion dollars in disposable income was associated, on the average, with an increase in consumers' expenditures of about 8 billions. Other studies utilizing Department of Commerce data agree that the historical consumption function based on disposable income for the period between the early 1920's and World War II can be accurately described by a straight line with a slope of not far from 0.8.[17] A linear relationship with a not much different slope still holds, as a matter of fact, if the data are put on a *per capita* basis or if somewhat different prewar periods are used. The close fit to a linear regression also holds if the data are not corrected for price changes. The precise equations obtained by different investigators have varied somewhat depending on the particular years included, whether the figures have been corrected for price and population changes, and whether any independent variables other than income are introduced into the equation.

Figure 15 suggests that the postwar consumption-income relationship can also be described by a straight line, the slope of which

[17] See, for example, Robert Ferber, *A Study of Aggregate Consumption Functions*, National Bureau of Economic Research Technical Paper 8, 1953. Compare the comparable savings functions (for personal savings with consumers' durables excluded from saving) in Goldsmith, *op. cit.*, vol. III, part 4; also Irwin Friend, *Individuals' Savings: Volume and Composition*, 1954, chap. 8. It should be noted that a regression line fitted to the 1920's alone, excluding the depressed years after 1929, would have a somewhat steeper slope. See the reference to Hickman, footnote 25, below.

is perhaps slightly higher than for the period 1929–1940. That is, the postwar historical marginal propensity to consume seems to be a bit higher than for the prewar period including the 1930's; the postwar regression line fitted in Figure 15 has a slope of .89, compared to .80 for the period 1929–1940. In general, any given level of disposable income in the postwar period seems to have generated a slightly higher level of consumers' expenditures than we might have predicted from a mere extrapolation of the fitted prewar regression line. But the difference does not seem to be very great, and a line fitted to the entire period 1929–1959 (excluding the years of World War II) would fit the data fairly well.[18] It would, however, have a steeper slope than that for the prewar period taken alone.

The lines of regression fitted in Figure 15 describe linear consumption functions that fit the actual data for the prewar and postwar years very closely. This is not surprising, since in effect we are correlating disposable income with another series (consumption) which is equal to more than 90 percent of disposable income. This does not mean however, that $\frac{\Delta C}{\Delta Y_D}$ did not vary significantly from year to year or quarter to quarter. Actually, $\frac{\Delta C}{\Delta Y_D}$ fluctuated quite widely, even for the prewar period. In short, the slope of the historical consumption function, even though the function fits the

[18] For a more detailed recent discussion of consumption-income relationships before and after World War II, see L. J. Paradiso and Mabel A. Smith, "Consumer Purchasing and Income Patterns," *Survey of Current Business,* March, 1959, pp. 18–28. This article measures the income sensitivity or elasticity of consumers' expenditures; that is, it relates *percentage* changes in consumers' expenditures to *percentage* changes in income, whereas the usual type of consumption function deals with absolute changes. The income elasticity of consumers' expenditures, defined as the percentage change in spending divided by the percentage change in income, is equal to the marginal propensity to consume divided by the average propensity. If ΔC and ΔY represent absolute changes in consumption and income, then

$$\frac{\Delta C}{\Delta Y} \div \frac{C}{Y} = \frac{\Delta C}{C} \div \frac{\Delta Y}{Y}.$$

The right-hand expression is the income elasticity of consumers expenditures. Concern with percentage relationships leads to putting the consumption-income relationship in logarithmic form, for example, $\log C = \log a + b \log Y$, so that $C = aY^b$. In these equations, b is the income elasticity of consumers' expenditures.

original data very closely, is only a rough guide in estimating the actual changes in consumption that are associated with short-period changes in income. This conclusion is reinforced if we examine the quarter-to-quarter changes in the postwar period as shown in Figure 16. Note, for example, the quite different slopes of the lines connecting successive quarters in the period covered by the chart (1953–1959). Although there was clearly an *average* relationship between C and Y_D for the postwar period, this average relation frequently provided a poor guide for predicting how consumption would change from quarter to quarter as income changed. The short-period consumption function is unstable; and, as we saw in Figure 13, the historical relation does not tell us how the short-period relationship shifts from period to period.

The average propensity to consume—or its counterpart, the percentage of income saved—gives a greater impression of stability than do the figures for $\frac{\Delta C}{\Delta Y_D}$. If we exclude the war years and the years of deep depression in the 1930's, the ratio of $\frac{C}{Y_D}$ seems to remain moderately stable within a range of 92 to 95 percent, which is to say that, in normally prosperous years, saving seems to average about 5 to 8 percent of disposable income. (Compare Table 7 on page 54.) In wartime, when there were special incentives to save as well as shortages of goods, price control, and rationing, saving was considerably higher than this. In years of extremely depressed incomes, as in 1932–1933, consumers in the aggregate spent more than their total disposable income. On the whole, the following seems to be a safe generalization which is supported by both the facts and the general arguments advanced in an earlier section: There is a long-run tendency for the percentage of disposable income saved to remain approximately constant, but in the short run the ratio of saving to income tends to vary with the business cycle, being relatively low in depression years and relatively high in years of prosperity.[19]

The short-run variations in $\frac{C}{Y_D}$ and the high degree of variability

[19] During the 1950's, saving averaged somewhat higher as a percentage of disposable income than in prewar prosperous years—that is, between 6 and 8 percent rather than about 5 percent.

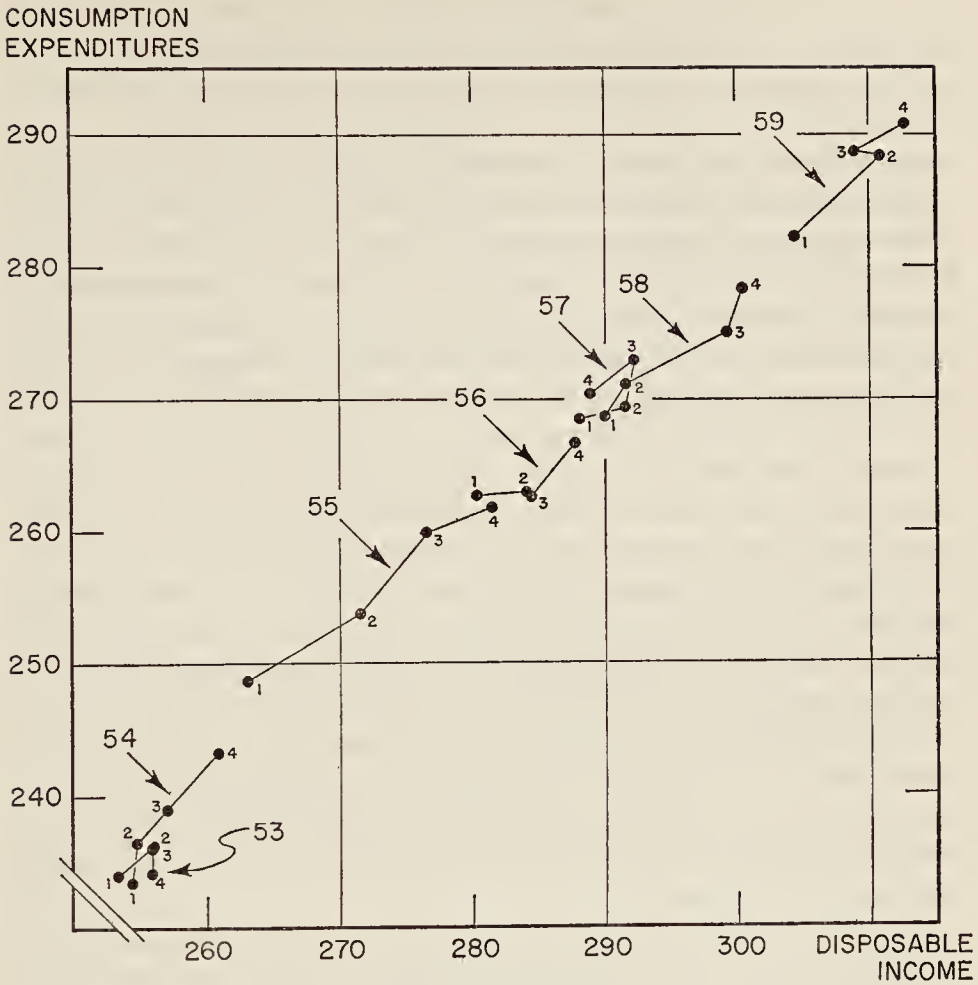


FIGURE 16. Relation Between Consumers' Expenditures and Disposable Income, Quarterly, 1953–1959.

Based on Department of Commerce data.

in $\frac{\Delta C}{\Delta Y_D}$ strongly suggest what we have already been led to expect—namely, that historical consumption functions such as that in Figure 15 conceal cyclical shifts in the "true" short-period consumption function. Complicated statistical techniques have been used to isolate the short-period functions that are assumed to hold in different phases of the business cycle, but the results thus far are inconclusive. Though the evidence clearly indicates that the short-

period relation between consumption and income does change over the course of the business cycle, we have not been able to isolate the short-period relation and explain satisfactorily how and why it shifts.

A number of writers have tackled this problem, and they have not all come up with the same answers, although a good many of the answers point in the same direction. Thus, an early study by L. H. Bean suggests that, for a given level of income, consumption is larger at a low level of business activity than at a high level.[20] This is consistent with the conclusion reached by Duesenberry and Modigliani that the percentage of income saved depends on the relation between current income and the highest level of income reached in the past. The rate of saving falls as current income declines below the previous peak and rises as current income returns toward the previous peak. Other studies have suggested that the direction and rate of change in income affect the relation between consumption and income, and we listed on page 93 additional factors that might also affect the position and slope of the short-period consumption function.[21] One of these, which is particularly important in business-cycle analysis, is the intangible element of expectations. An example is the sharp rise in consumers' expenditures relative to incomes after the outbreak of the Korean crisis, when consumers rushed in to anticipate the expected rise in prices and scarcity of goods.[22]

One final example may be given of the numerous attempts to compute a short-run consumption function. Using quarterly data corrected for price changes, Zellner obtained fairly good results with the following two equations, among others:[23]

[20] In *Review of Economic Statistics,* vol. 28, November, 1946, pp. 199–207. A level of income that would have yielded full employment twenty years ago obviously would not do so today.

[21] For a recent study, with references to the earlier literature, see Arnold Zellner, "The Short-Run Consumption Function," *Econometrica,* vol. 25, October, 1957, pp. 552–567. See also the references cited in earlier footnotes in this chapter.

[22] For extended discussion of the role of consumers' expectations, see George Katona and Eva Mueller, *Consumer Attitudes and Demand, 1950–52,* 1953, and *Consumer Expectations, 1953–56,* n.d. See also "Consumer Buying Plans," in *The Conference Board Business Record,* vol. 16, June, 1959.

[23] Zellner, *op. cit.,* p. 560. The variables are expressed in billions of constant dollars (1947–1949 = 100). The equations are fitted for the period 1947–1955. The results were somewhat better for the first of the two equations given here.

$$C = -18.96 + 0.375Y_D + 0.489C_{t-1} + 0.219L_{t-1}$$
$$C = -23.02 + 0.458Y_D + 0.369C_0 + 0.272L_{t-1}$$

Here C_{t-1} stands for consumers' expenditures of the preceding quarter; C_0 is past peak consumption; and L_{t-1} represents individual holdings of liquid assets at the beginning of the quarter. Note that the introduction of past (or past peak) consumption and of liquid assets considerably reduces the dependence of current consumption on current disposable income. Most of this effect is due to the use of the lagged (or past peak) consumption variable. The more current consumption depends on past consumption or income, the less will be the effect of *current* income on current consumers' expenditures.[24]

THE PROPENSITY TO CONSUME GNP

Although we shall not examine the data in detail, we should take a brief look at the relationship between consumers' expenditures and GNP. Here again, we get for the prewar years a fairly stable relationship closely approximating a straight line. If we were to prepare a scatter diagram like that in Figure 15, using GNP instead of Y_D as our measure of income, the points for the prewar years would be close to a straight line with a slope of not far from 0.65.

The difference between GNP and C can be described in two different ways: as the sum of $I + G$ or as the sum of all the diversions from consumption, including gross business savings and taxes as well as personal saving. Both descriptions suggest that $\frac{C}{GNP}$, the average propensity to consume the gross national product, should vary inversely with the business cycle. Private capital formation tends to be a larger fraction of the GNP in prosperous years than in depressions, and the reverse is true of consumption. Similarly,

[24] Thus, when consumption is made to depend only on current income and liquid assets, Zellner obtained the following equation:

$$C = -21.91 + 0.708Y_D + 0.368L_{t-1}.$$

Notice how much larger is the influence of current income here than in the two equations cited in the text. Here the change in consumption is equal to 70.8 percent of the change in *current* income, but to only 37.5 and 45.8 percent in the two equations that also include some measure of past consumption.

business savings tend to be relatively large when business is good and relatively small when profits are low. To express the matter in another way, consumption rises less fast (and private investment faster) than the GNP during the upswing, and consumption falls more slowly (and private investment faster) than the GNP during downswings. The resulting variation in the ratio of C to GNP is shown in Table 5 (page 45).

The short-term relationship between C and GNP points up the extent to which historical *ex post* data may fail to reveal the underlying *ex ante* magnitudes. The ratio $\frac{C}{GNP}$ and particularly the marginal propensity $\frac{\Delta C}{\Delta GNP}$ reflect the wide fluctuations in business savings and in inventory accumulation. Both of these variables are subject to large unexpected changes, with the result that $\frac{\Delta C}{\Delta GNP}$ for any particular year may give only a rough indication of the plans and expectations of consumers and businessmen. One would be rash, therefore, to use the average historical relationship between ΔC and ΔGNP to predict what change in consumption was likely to be associated with a given change—during, say, a year—in the GNP. Here, as in the case of disposable income, the average propensity provides a better basis for prediction than does the marginal propensity, although we must allow for the certainty of cyclical shifts and the possibility of secular changes.

Actually, it is somewhat misleading to speak of a marginal or average "propensity" to consume or save when relating consumption to the gross national product. The sum of the diversions measured by the difference between GNP and C is not a homogeneous total. It is made up of such heterogeneous elements as personal savings, depreciation, net business saving, social security contributions, and personal and business taxes, from which transfer payments must be subtracted. These do not all necessarily behave in the same way as GNP changes, cyclically or secularly. The total of these diversions represents a composite product of the spending habits of consumers, the expectations of businessmen, customary financial practices, tax laws, social security legislation, and so on. Over considerable periods, there are likely to be underlying tendencies mak-

ing for a given distribution of the GNP among consumers' spending, private capital formation, and government expenditures. But, to repeat, this distribution is subject to considerable variation, particularly in short periods.

As we shall see in later chapters, one set of factors contributing to the relative stability of the American economy in the postwar period has been the greater importance of the so-called "automatic stabilizers," which tend to cushion the effect on consumers' expenditures of a decline in GNP. These stabilizers include undistributed corporate profits, corporate and personal income taxes, and transfer payments. As the GNP declines, undistributed profits and corporate and personal income taxes tend to fall off rapidly, thereby reducing the amount by which consumers' expenditures would otherwise have to fall, and consumers' disposable incomes are further supported in a recession by a rise in transfer payments (for example, in unemployment compensation benefits). The result is that in the postwar years, $\frac{\Delta C}{\Delta GNP}$ has on the average, been lower than before World War II. This effect is particularly noticeable in short, especially recession, periods.[25]

THE THEORY OF THE MULTIPLIER

Even though the short-period consumption function is not completely stable, it is clear that the level of income is an important determinant of the amount of consumers' expenditures. Consumption varies as income changes, although we may not always be able to predict the precise nature of the relationship at any particular time.

[25] Hickman gives the following results for 1921–1929 and 1947–1958 (annual data in billions of current dollars):

$$1921\text{–}1929: C = 7.9 + .69GNP$$
$$1947\text{–}1958: C = 26.4 + .58GNP$$

Note the decline in the "marginal propensity to consume" GNP between the two periods, from .69 to .58. A good part of this decline is due to the effects of the automatic stabilizers mentioned in the text. Cf. B. G. Hickman, *Growth and Stability of the Postwar Economy,* 1960, chap. 9. For a more detailed discussion of the postwar relation between C and GNP, see Duesenberry, Eckstein, and Fromm, *Stability and Instability of the American Economy,* Harvard University, multilithed, 1958 (later published in *Econometrica,* October, 1960).

We can draw an important corollary from this fact. If income increases for any reason—say, because of an increase in investment or government spending—consumption will also expand by some amount; this creates a further increment of income, which in turn leads to more consumption and hence to still another increment of income; and so on. The initial rise in spending has a multiplied effect on income because a succession of rounds of induced consumers' spending is added to the initial stimulus. The ratio of the total increase in income to the original increase in spending is called "the multiplier."

The theory of the multiplier in its most general form states a truth of great importance for business-cycle analysis. It throws additional light on the self-generating process that goes on during cyclical expansions and contractions. It helps to explain how a dollar of spending generates additional dollars of spending through the automatic reaction of consumers to changes in income. The trouble comes when we try to get more specific than this, particularly when we try to ascertain just how large will be the total increase in income resulting from an initial rise in private investment or government spending.[26]

We have defined the multiplier as the ratio of the total change in income to the initial change in spending with which it is associated. We can write ΔY for the total change in income (we could also use ΔGNP), and, for the purposes of this discussion, we shall let ΔI represent the initial change in spending, whether it takes the form of private investment or government spending.[27] If we call the multiplier m then

$$m = \frac{\Delta Y}{\Delta I}.$$

[26] For a critical evaluation of the substantial literature on the multiplier, see Hugo Hegeland, *The Multiplier Theory*, 1954; also Ralph Turvey, "Some Notes on Multiplier Theory," *American Economic Review*, vol. 43, June, 1953, pp. 275–295.

[27] It could also take the form of a change in consumers' spending brought about by a shift in the consumption function. This possibility is usually ruled out in the theoretical literature by the assumption that the consumption function is stable. It should also be noted that we have defined the multiplier in terms of the relation between monetary magnitudes. We could, however, express the multiplier in terms of employment or real output (after correction for price changes). The employment multiplier relates an initial change in employment due to an increase in investment with the total change in employment which ultimately results.

THE SIMULTANEOUS MULTIPLIER

An immediate question that arises has to do with the dating of the numerator and denominator in the above fraction. Is ΔY to be taken for the same period as that in which ΔI occurs, or is ΔY to be interpreted as the increase in income over some period in the future following an initial increase in investment? The latter seems to be the more reasonable approach. However, Keynes made use of what we may call the simultaneous multiplier, in which ΔY and ΔI are taken as applying to the same period. Let us see why this approach does not yield very useful results.[28]

We shall examine this simultaneous multiplier relationship under the simplified assumption that aggregate demand is composed of only the two elements C and I. What happens to income if investment is increased by some amount ΔI? *Ex post,* there must be an equal increase in saving, since realized saving is always equal to realized investment. But saving is assumed to depend on income in some predetermined way which is given by the consumption function. Income must therefore rise sufficiently so that, given the marginal propensity to consume, saving will rise by an amount equal to ΔI. If the marginal propensity to consume is .8, and if I increases by one billion dollars, then Y must rise by $\frac{1}{1-.8}$ or five billions if saving is to increase by one billion to equal the increase in investment. An increase in investment thus leads to an increase in income large enough, given the marginal propensity to consume and save, to yield an amount of saving equal to the increased investment.

The multiplier thus turns out to be the reciprocal of the marginal propensity to save, and the marginal propensity to save is one minus the marginal propensity to consume. One way of showing this is as follows. If c is the marginal propensity to consume, the change in consumption (ΔC) resulting from any change in income (ΔY) will be given by

$$\Delta C = c\Delta Y.$$

But ΔY is simply the sum of the changes in C, I, and G, which we can write as:

$$\Delta Y = \Delta C + \Delta I + \Delta G.$$

[28] For a sympathetic exposition of Keynes' views on the multiplier, see A. H. Hansen, *A Guide to Keynes,* 1953, chap. 4.

Substituting the first equation into the second, we have:

$$\Delta Y = c\Delta Y + \Delta I + \Delta G$$
$$\Delta Y - c\Delta Y = \Delta I + \Delta G$$
$$\Delta Y = \frac{\Delta I + \Delta G}{1 - c}.$$

Thus, according to this theory of the multiplier, if we know the marginal propensity to consume, we can predict the level of income that will result from any change in the rate of investment or government spending. Obviously, the theory of the multiplier is of great practical importance—unless it is based on assumptions that limit its application to the real world. Let us examine these assumptions.

The main weakness of the simultaneous multiplier can be illustrated by a simple example. Let us assume that in a given period $I + G$ rises by two billion dollars and that in the same period GNP rises by six billions. Consumption therefore rises by four billions. To say in this case that the multiplier is 3 is the same thing as saying that the "marginal propensity to consume" is ⅔ (i.e., four billion divided by six billion). The "marginal propensity to consume" in this case does not represent a stable relationship or a psychological propensity. *Ex post,* the part of the GNP not consumed represents not merely planned personal saving but also the volatile element of corporate saving, including unexpected changes in profits, as well as unintended personal saving, taxes, and certain minor leakages. We have seen that this version of the "marginal propensity to consume" varies significantly and somewhat erratically from year to year. In view of all the unexpected things that can happen, we have no way of knowing how much Y or GNP will rise in the same period in which I or G increases.

Thus, the simultaneous multiplier merely states the *ex post* relationship between a realized change in national income or GNP and the realized change in investment and government spending in the same period. To repeat, this concept of the multiplier is only distantly related to the propensity of individuals to consume their disposable income, and it tells us nothing about any causal connection between the saving habits of individuals and the amount of net income or GNP that will be generated within any period by a given increase in investment or government expenditures.[29]

[29] Cf. Gottfried Haberler, "Mr. Keynes' Theory of the 'Multiplier': A Methodological Criticism," reprinted in *Readings in Business Cycle Theory,* selected by

THE PERIOD MULTIPLIER

We may approach the multiplier effect of a change in investment or government spending on income in another and on the whole a more significant way, although this approach also has serious limitations. One difficulty with the concept of the simultaneous multiplier thus far considered is that it relates a change in investment to a *simultaneous* change in income. Yet, if there is to be any stimulating effect, the passage of time is necessary so that the initial increase in spending can generate further increases in spending. It is only reasonable to assume that the full multiplying effects of a change in investment are not felt immediately, and thus there may be little significance in the ratio of a current change in income to the change in investment in the same period.

If we introduce the idea of a train of secondary spending effects occurring through a succession of income periods, we arrive at the idea of a period multiplier, the nature of which we must now investigate. The period multiplier states a relation between a given increment of investment or government spending and the increase in income that ultimately results by the end of some specified period, which may be much longer than that during which the original change in investment takes place.

Demonstration of the nature of the period multiplier involves the tracing of money flows around the income circuit along the lines described in Chapter 2. As in the case of the simultaneous multiplier, we assume a given marginal propensity to consume. We shall deal in gross terms and assume that, for any change in GNP, $\frac{\Delta C}{\Delta GNP}$ is .6. Let us assume further that currently the GNP is running at the rate of 100 billions per quarter, or 400 billions per year. Now let us suppose that private business decides to invest an additional billion dollars on new plant and equipment, and that this is the only change that takes place. We could assume equally well that the increase came through government spending. We could also assume that the change came through increased consumers' spending brought about by an upward shift in the consumption function. However, the

---

a committee of the American Economic Association, 1944, pp. 193–202. See also Fritz Machlup in the same volume, p. 218, and Hegeland, *op. cit.*, chaps. 4 and 6.

multiplier theory assumes that the consumption function is stable and that consumption does not change unless there is first a change in income.[30]

The billion dollars of additional investment would ordinarily be divided between direct payment to workers and purchase of materials and equipment. Part, therefore, immediately becomes income for the workers on the investment project. Let us say this amounts to 400 millions. The rest passes through a succession of business firms, which produce additional goods in response to the increased demand for building materials and equipment. Of the original 600 millions used in this way, let us say that various contractors, jobbers, and manufacturers retain 250 millions in the form of net business saving, depreciation, and reserve for additional taxes, and that the remaining 350 millions becomes income paid out in the form of wages, dividends, and the like. For simplicity, we shall assume that the increased production of materials and equipment takes place immediately so that the 350 millions is available as income payments at the same time that the 400 millions is paid directly to construction workers. The original billion of investment has thus far given rise to 750 millions of income payments —400 directly and 350 indirectly.

The larger part of this sum will very quickly be spent on consumers' goods. If we assume a marginal propensity to consume personal income of .8, then 600 of the 750 millions will go for consumers' expenditures and 150 million will take the form of personal saving and taxes. Thus, of the original increase in I and GNP of one billion, 600 million is eventually spent again for consumption. The marginal propensity to consume personal income is .8, but, because of business diversions, $\frac{\Delta C}{\Delta GNP}$ is only .6.

We must now give some attention to the lapse of time. Let us say the income velocity of active money—that is, money that is not idle —is four per year, and that it therefore takes, on the average, three months for a dollar of the active money supply to travel around the income circuit—from the time it is received by one income recipient until it returns to another income receiver, or alternatively, from

[30] The assumption is also implicitly made that either the money supply or income velocity increases sufficiently to support the increased spending. If MV_T remains constant, there will be no net increase in aggregate demand.

the time it is first spent on new final output until it is again spent on current output.[31]

We may therefore assume that, on the average, within about three months after the initial investment of one billion, 600 millions more will have been added to the GNP through the additional consumers' expenditures. Business firms receive the 600 millions that are thus spent on consumption and retain part as gross savings and to pay taxes, and the remainder becomes income payments to individuals. The latter retain part for personal savings and

TABLE 10. The Multiplier Process for a Single Increment of Investment
(In billions of dollars)

Period (in calendar quarters)	ΔI	ΔC	ΔGNP	Total GNP[a]	Cumulative Addition to GNP
First quarter[b]	1.000	—	1.000	101.000	1.000
Second quarter	—	.600	0.600	100.600	1.600
Third quarter	—	.360	0.360	100.360	1.960
Fourth quarter	—	.216	0.216	100.216	2.176
Fifth quarter	—	.130	0.130	100.130	2.306
Sixth quarter	—	.078	0.078	100.078	2.384
Seventh quarter	—	.047	0.047	100.047	2.431

[a] Assuming that the GNP was 100 billion per quarter before the increment of investment occurred.

[b] The initial investment is assumed to take place at the end of this quarter.

taxes, and the remainder again becomes consumers' expenditures. If the marginal propensity to consume the gross national product remains at .6, 360 millions will be spent on consumption this second time. Now the process repeats itself, and by the end of another three months there will be a further expenditure on consumption amounting to .6 times 360 millions or 216 millions, and so on.

This multiplication process through successive periods is illustrated in Table 10. The effect on the GNP of the initial act of investment gradually diminishes and approaches zero as a limit; that is, if no further additional investment takes place, the GNP eventually returns to the original figure of 100 billions.

[31] The ratio of annual GNP to *M* in 1959 was in the neighborhood of 3.3. (*M* is here defined as total adjusted demand deposits and currency.) Not all of this money supply was active. Hence, an income velocity for the active money supply of something like 4 per year is probably a reasonable assumption.

What will be the total amount added to the GNP over any succession of periods? The answer can be easily read from the infinite series representing the multiplying process described in Table 10. If we let ΔGNP represent the total cumulative addition to the GNP, then:

$$\Delta GNP = 1 + .6 + .6^2 + .6^3 + \cdots .6^n.$$

More generally, letting ΔI represent the initial increase in investment and c the marginal propensity to consume,

$$\Delta GNP = \Delta I(1 + c + c^2 + c^3 + \cdots c^n).$$

The sum of the infinite series within the parentheses is $\frac{1}{1-c}$. But c is the marginal propensity to consume. Thus, we have the same relation between the multiplier and the marginal propensity to consume as we had in the case of the simultaneous multiplier. In this case, however, the multiplier gives the effect of an increment of investment in a given period on income over all periods into the future; it does not show the relation between ΔGNP and ΔI in the same period.

In the case we have been discussing, the total multiplier over all future periods would be $\frac{1}{1-.6} = 2.5$. If $\frac{\Delta C}{\Delta GNP}$ had been .5, the multiplier would have been 2; if it had been $\frac{2}{3}$, the multiplier would have been 3; and so on.

In practice, we are not likely to be interested in the *total* multiplier over the infinite future. We can be certain that the increments of GNP in the distant future which are attributable to the original investment will be small and will be lost in the effects of new changes that occur. We are interested, however, in what the *period* multiplier—the multiplier over some given finite period—will be. The answer is given by the progression already presented, provided we know the income velocity of additions to the money supply.

Thus in Table 10 we see that at the end of one income period, or three months after the initial increase in investment, the total increase in the GNP is 1.6 billions, of which one billion represents the original investment and .6 billion represents the first increase in consumption. During the next three months, consumption increases by a further .36 billion, so that at the end of six months the period multiplier is 1.96. It is 2.176 at the end of nine months and 2.306 at

the end of one year. Thus, about 92 percent of the total multiplier effect will have been felt within twelve months. The greater the income velocity of money, the greater will be the period multiplier within any given interval of time. With a given income velocity, the fraction of the total multiplier that is felt within any given period will be larger for low values of the total multiplier than for large values.[32]

Thus far we have assumed a single act of increased investment which was not repeated in subsequent periods. Now let us suppose that the increased investment is continued through a succession of quarters. What will be the effect on the GNP? We proceed as before, but in each quarter a new train of income-generating effects is added to the results of the preceding investment. Table 11 shows what would happen, assuming as before a constant $\frac{\Delta C}{\Delta GNP}$ of .6 and an income period of three months.

At first, GNP increases by the amount of the original investment. During the next income period of three months there is another increment of investment of one billion, but in addition consumption rises by .6 billions as a result of the first quarter's investment. GNP in the second quarter is therefore 101.6. In the next quarter, there is again a billion of ΔI; there is now .6 billion of additional induced consumption from the second quarter's investment; and there is another increment of consumption of .36 billion resulting from the first quarter's investment. The GNP in this period is 101.96. As the additional investment continues through succeeding quarters, the effects on consumption and GNP can be followed in Table 11. The diagonal lines trace the succession of increments of consumption resulting from each quarter's investment.[33]

[32] Cf. F. Machlup, in *Readings in Business Cycle Theory*, p. 220; also H. H. Villard, *Deficit Spending and the National Income*, 1941, pp. 252–257. The paper of Machlup's, entitled "Period Analysis and Multiplier Theory," is perhaps the best simple presentation that has appeared of the period approach to the theory of the multiplier.

[33] For the sake of simplicity, we shall ignore in this discussion the complications that arise because of the existence of international trade. The increase in consumption will lead to additional imports, and this in turn may stimulate exports. The matter is not important for a country such as the United States but is very important for a country such as Great Britain. When the increased domestic spending stimulates imports, to that extent it is foreign rather than domestic incomes which are increased; but this may be offset in part by an induced expansion of exports because of the rise in foreign incomes. See the further discussion of this range of issues in Chapter 22.

If the additional investment continues indefinitely, GNP will approach 102.5 billions as a limit. Indeed, it nearly reaches this figure in the eighteen months after the first increment of investment covered in Table 11. If the new rate of investment continues indefinitely, we eventually obtain a *simultaneous* multiplier of 2.5; the additional investment of one billion eventually makes the GNP per

TABLE 11. The Multiplier Process with a Continuing Increment of Investment
(In billions of dollars)

Period	ΔI	ΔC [a]	Total GNP [b]
1	1.00	—	101.00
2	1.00	.60	101.60
3	1.00	.60 + .36	101.96
4	1.00	.60 + .36 + .22	102.18
5	1.00	.60 + .36 + .22 + .13	102.31
6	1.00	.60 + .36 + .22 + .13 + .08	102.39
7	1.00	.60 + .36 + .22 + .13 + .08 + .05	102.44
		(And so on through succeeding quarters)	

[a] Diagonal lines trace the increments of consumption in successive periods which are attributable to the increment of investment in each period.

[b] Represents ΔI plus ΔC plus assumed initial GNP of 100.0.

period 2.5 billion larger than it was before the increased rate of investment began. But notice: a simultaneous multiplier of 2.5 is not obtained immediately but only after the lapse of a considerable period of continued investment at the higher rate.

This suggests what is wrong with the simple theory of the simultaneous multiplier expounded earlier in this section. This theory stated that, with a marginal propensity to consume of .6, an increment of investment of one billion would lead to a *simultaneous* increase of 2.5 in the GNP. This would have occurred if the *ex post*

marginal propensity to consume had actually been equal to .6, but it was not. The simultaneous multiplier deals with *ex post* magnitudes; the period multiplier with *ex ante* magnitudes. This difference can be explained by referring again to Table 11.

When the first billion of investment was made, but before any additional consumption had taken place, the GNP was a billion larger than before, but there had as yet been no increase in consumption. *Ex post,* $\frac{\Delta C}{\Delta GNP}$ was for the moment zero and not .6. The increase in GNP corresponded to unintended saving, a situation that would be corrected with the lapse of time. In the next income period, GNP was 101.60, or 1.60 higher than at the beginning. Consumption was .60 larger. This gives us a realized $\frac{\Delta C}{\Delta GNP}$ of less than .4. In this second period, the *ex ante* marginal propensity to consume of .6 has given rise to .60 of consumption and hence .40 of *voluntary* saving. But ΔI in the second period is again one billion. Hence, there must be an additional .60 of unintended saving (in the form of additional incomes received but not yet spent) to make *ex post* saving equal to the billion of new investment. In the next period voluntary saving rises to .64 (.40 out of the preceding period's billion of investment-created income and .24 out of the preceding period's consumption-created income), and unplanned savings are therefore .36. Because of these unplanned savings, the *ex post* marginal propensity to consume is still not equal to .6. Eventually, voluntary savings will rise to one billion per quarter, at which time GNP will be at 102.5. At this point, planned savings are equal to planned investment; the *ex post* marginal propensity to consume is equal to the *ex ante* propensity; and the period and simultaneous versions of the multiplier yield the same results if the same rate of investment is continued in the future.

EVALUATION OF THE THEORY OF THE MULTIPLIER

How useful is the theory of the multiplier in the period version here described? There is no question that it expresses an important truth: that an increase in investment will have secondary consequences which, if not offset, will lead to an increase in income larger than the increase in investment. But there are serious limitations in applying the multiplier concept in practice.

An obvious limitation is that the theory assumes a constant marginal propensity to consume. We have seen that $\frac{\Delta C}{\Delta GNP}$ is not stable, particularly in the short run. The comparison of C with GNP or Y means that business savings, which are highly volatile, must be included in total savings. Further, it is safe to assume that the process by which the multiplying effects work themselves out will itself create changes in $\frac{\Delta C}{\Delta GNP}$.

Perhaps more serious is the theory's neglect of induced investment. It is assumed that despite the increase in personal and business savings and despite the increased demand for consumers' goods, there will be no change in investment other than that originally assumed. Yet it is inconceivable that businessmen will not change their investment plans as they see the effects on consumers' demand of the assumed change in investment. Apart from any increase in physical facilities and inventories required by expanding consumers' demand, changing business expectations will certainly lead to changes in investment that will enhance or partially offset the multiplying effects of the original investment.

The way in which private investment reacts to the increase in aggregate demand begun by the multiplier process may vary in different phases of the cycle and with the nature of the original increase in investment. In the early stages of the upswing, businessmen may meet the increased demand out of existing inventories, so that disinvestment from this source offsets the original increase in investment. Later, inventories may be replaced and new facilities added, so that additional investment enhances the multiplying effect of the initial investment. The original investment may also have repercussions on the availability of credit, interest rates, and prices, which in turn will affect total investment.[34]

It follows from this that perhaps the greatest weakness of multi-

[34] On all this see the excellent discussion in Machlup, *op. cit.*, pp. 228–234. Cf. also J. M. Clark, "An Appraisal of the Workability of Compensatory Devices," reprinted in *Readings in Business Cycle Theory*, esp. pp. 302–305; Hegeland, *op. cit.*; and Gardner Ackley, "The Multiplier Time Period," *American Economic Review*, vol. 41, June, 1951, pp. 350–368. For a good technical discussion of the monetary implications of multiplier analysis, see S. C. Tsiang, "Liquidity Preference and Loanable Funds Theories, Multiplier and Velocity Analysis," *American Economic Review*, vol. 46, September, 1956, pp. 539–564.

plier theory is its exclusive emphasis on consumption. It would be more realistic to speak of a "marginal propensity to spend," rather than to consume, and then to consider the repercussions of an initial increase in investment not only on consumption but also on total private investment and on government spending. This would require that the multiplier approach pay much more attention than it does to changes in business attitudes toward spending.[35] If the problem is put in terms of the effects on changes in total spending, we should also recognize the small likelihood that a constant "marginal propensity to spend" will endure through a succession of circuit flows and through different stages of the business cycle. Thus, it is not likely that in a dynamic world we shall be able to predict with much accuracy the multiplying effects over the cycle of a given increase in private investment or public spending.[36]

While these limitations must be recognized, the theory of the multiplier contains an element of truth which is highly important for an understanding of business-cycle fluctuations. Increased spending on new output will expand employment and incomes as it sets up a succession of further increases in spending; and the impact on incomes and employment will be greater, the smaller the diversions or leakages that reduce the secondary spendings. Since private investment, much more than consumption, fluctuates in response to stimuli other than changes in income, it occupies a particularly strategic role in business-cycle analysis. Changes in the level of investment cannot be explained merely by prior changes in income; and the fluctuations in investment that occur, for whatever reason, have a magnified effect on total output and employment through the process described by the period version of the multiplier theory. It is no wonder, therefore, that so many

[35] Cf. J. W. Angell, *Investment and Business Cycles*, 1941, chaps. 9–11.

[36] These observations are reinforced by the fact that the successive increments of income may follow each other at varying intervals of time. Thus, if consumers' expenditures increase, retailers may be slow or fast in increasing orders from manufacturers, and the latter may increase production and payrolls before they are paid by retailers, utilizing funds already on hand or borrowing new money for the purpose. Thus, a wholesaler's or manufacturer's contribution to the increase in income resulting from new consumption is not rigidly geared to the rate at which the money spent by consumers flows through the channels of distribution from retailers to manufacturers of finished goods to suppliers of raw materials and so on. See Ackley, *op. cit.*, pp. 350–368.

theories of the business cycle center their attention on the causes of changes in the volume of investment.

The possibility of a secondary expansion that may result from an increment of new spending explains why so much emphasis has been placed on government spending as a means of bringing about recovery from business depressions. The theory of the multiplier suggests that additional government spending, not offset by increased taxes, will in all probability raise the GNP by more than the primary increase in spending. This much we can safely say. But it is dangerous to forecast what the value of the multiplier will be, because we cannot say in advance precisely how consumption and particularly private investment will react. Under favorable circumstances, induced private investment, stimulated by improving expectations, may add its stimulus to that provided by the initial government spending; in other cases, private investment may decline and offset, wholly or in part, the stimulating effect of the increased government expenditures.

CHAPTER 6

THE DETERMINANTS OF INVESTMENT

WE HAVE SEEN that the level of income depends on the level of aggregate demand, and in Chapter 5 we studied the forces determining the size of the largest component of aggregate demand—namely, consumers' expenditures. This analysis led us to a twofold conclusion. Consumption depends chiefly on income; and, since this is the case, changes in investment have a multiplied effect on income and employment because of the induced changes in consumption that result. To this we can add another fact. As we saw in Chapter 3, the cyclical fluctuations in private investment are extremely wide—much wider relatively than variations in consumers' expenditure.

All this points up the key role that investment plays in the business cycle and brings us to the question raised at the end of Chapter 4: What determines the level of investment, and why does investment tend to fluctuate widely over the business cycle? If we can find a simple relation between investment and a few other variables and if we know what influences these other variables, we can then show how the level of investment is determined and why it is so sensitive to changing economic conditions. If we can do this, we shall have taken a long step toward explaining why and how business cycles occur.

As we shall see, also—and this is a fairly obvious point—investment plays a key role in determining not only economic fluctuations but also the rate of long-term economic growth. Through the multiplier, a steadily rising level of investment helps to support a rising level of aggregate demand. At the same time, this investment creates additional capacity and permits the economy to take

advantage of technological improvements that improve the productivity of economic resources. Thus, we rely on expanding private investment to generate growth in both aggregate demand and aggregate supply.[1] In the next to last section of this chapter, we shall take this basic idea and incorporate it into a simple "model" of economic growth.

THE POSSIBLE DETERMINANTS OF INVESTMENT

What, then, determines the level of investment? A very general answer would be that it depends on the expected profitability of investment and on the cost and availability of the necessary funds. The more a businessman plans to invest, other things being equal, the lower will be the net return he expects from the last dollar invested. Presumably he will invest up to that point where the expected net return from the last piece of investment just covers the cost of obtaining the necessary funds. Obviously, we have here a problem in demand-supply analysis. At any one time, business firms stand ready to invest (spend on capital goods) varying amounts depending on the price at which they can obtain the necessary funds.

THE MARGINAL EFFICIENCY SCHEDULE

We can represent the willingness of businessmen to invest in the form of a schedule that describes how much would be invested at various possible rates of return. We call this the marginal efficiency schedule of investment.[2] An example of such a schedule is given in Figure 17. The vertical axis represents the marginal rate of return expressed in percent. The horizontal axis represents

[1] See, for example, the discussion and diagrams in Chapter 4, pp. 76–79.

[2] Keynes used the phrase "marginal efficiency of capital," which he defined as being equal to that rate of discount that would make the present value of the series of expected returns from a capital asset during its life just equal to the cost of that asset. (*The General Theory of Employment, Interest, and Money*, p. 135.) Thus, if a machine costs $2000 and is expected to yield a return (before depreciation) of $1000 for each of three years and then to have no scrap value, we can write

$$\$2000 = \frac{\$1000}{1+i} + \frac{\$1000}{(1+i)^2} + \frac{\$1000}{(1+i)^3}.$$

Here i is the marginal efficiency, or rate of return, for this piece of investment. It will be higher the larger are the prospective returns and the lower is the cost of the machine.

the volume of investment. Thus, in Figure 17, the line I_1I_1', which slopes downward to the right like the usual demand curve, tells us how much, under a given set of conditions, would be invested at different possible marginal rates of return. The schedule I_2I_2' tells us how much would be invested at various rates of return under a different and more favorable set of conditions.

Suppose that at a particular time the marginal efficiency schedule is as given by I_1I_1'. Then, if the amount OA is to be invested, the

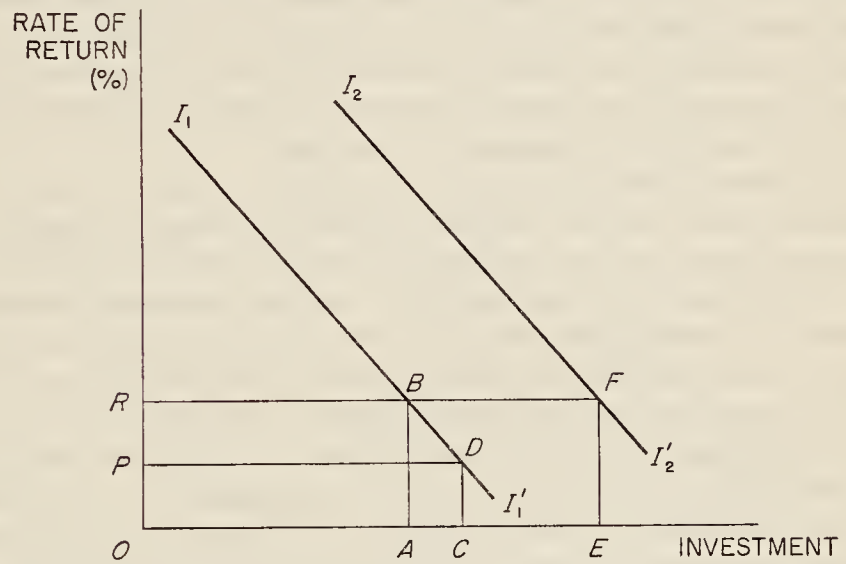


FIGURE 17. The Marginal Efficiency Schedule of Investment.

rate of return on the last bit of such investment is $AB\,(=OR)$. If businessmen maximize profits, they will invest the amount OA if the cost of obtaining the necessary funds (the interest rate) is equal to the marginal rate of return. If the interest cost is lower than OR—say, OP—then investment will be not OA but OC. Here we are making the volume of investment depend on the interest rate, which is what we did when on page 80 we wrote an equation for investment as

$$I = f(i).$$

If we use this sort of analysis, we can see that the volume of investment may change for either of two reasons: either the interest rate may change or else the marginal efficiency schedule itself may shift upward or downward. Thus, as we have seen, if the

interest rate falls from OR to OP, investment will expand from OA to OC. Or, if the schedule shifts from I_1I_1' to I_2I_2' without a change in the interest rate, investment will rise from OA to OE.

According to this way of looking at things, how sensitive investment is to changes in interest rates depends on the elasticity of the marginal efficiency schedule. Most of the evidence available suggests that investment is not very sensitive to changes in interest rates; i.e., the marginal efficiency schedule at any particular time is relatively inelastic, at least in the range within which interest rates usually fluctuate.[3] This does not mean that investment is completely insensitive to interest rate changes, but it does mean that we must look elsewhere for the main causes of the wide fluctuations in investment that we observe in real life.

It has been argued by some that the marginal efficiency schedule is actually more elastic than the preceding discussion implies. If investment does not seem sensitive to changes in the interest rate, the reason may be not so much an inelastic marginal efficiency schedule but rather inelasticity in the supply of funds available to borrowers.[4] In this connection, what is important is not the market rate of interest but what potential borrowers *believe* to be the cost of raising additional funds. Thus, if a firm has a tradition of never selling bonds (or if it is too "unseasoned" or too small to make the sale of bonds feasible) and if those in control also have serious objections to selling more stock, then the funds available will be largely limited to what the firm can reinvest out of current earnings.[5] Whatever may be prevailing interest rates, the cost of obtaining capital will *seem* to be high to this firm, and this will shut off new investment even if the marginal efficiency schedule is relatively elastic. This is illustrated in Figure 18. Here II_1 is the marginal efficiency schedule; RQS represents the cost of obtaining funds from the point of view of the borrower. To obtain any amount of funds beyond RQ entails rapidly rising costs. Hence,

[3] The literature on this subject is too large to summarize here. For a recent critical survey of some of the evidence suggesting that investment is relatively insensitive to interest-rate changes, see W. H. White, "Interest Inelasticity of Investment Demand," *American Economic Review*, vol. 46, September, 1956, pp. 565–587.

[4] See, for example, J. S. Duesenberry, *Business Cycles and Economic Growth*, 1958, chaps. 4–5.

[5] We exclude short-term borrowing from the banks, which is ordinarily not available to finance long-term investment.

investment is limited to OA, although it could have been as large as OC if unlimited funds had been available at an interest cost of $OR (= DC)$. The rapidly rising cost of obtaining funds, represented by the steeply rising section QS, may be more or less subjective to the borrower. Even so, his investment will be restricted.

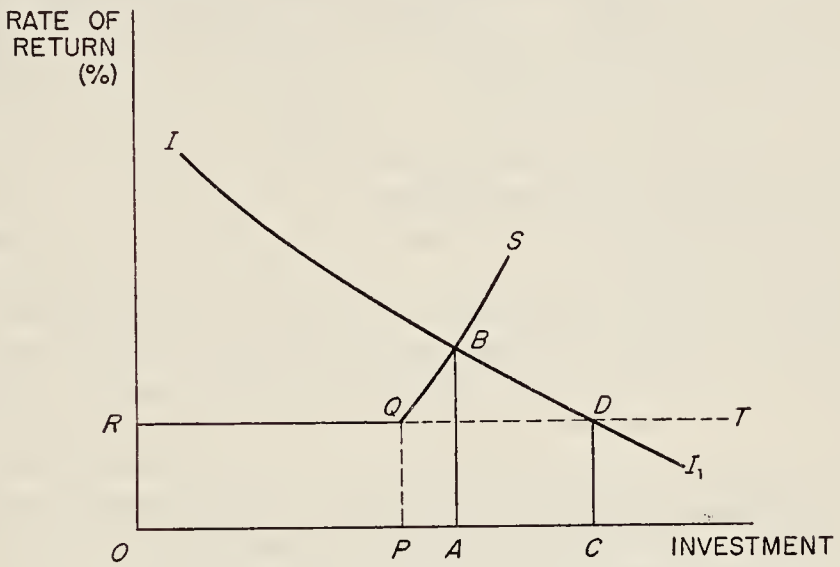


FIGURE 18. Effect of the Supply of Funds on Investment.

Further, QS will remain steep and may change little even if market rates of interest rise or fall.

SHIFTS IN THE MARGINAL EFFICIENCY SCHEDULE

It is not essential that we resolve this question as to just how elastic the marginal efficiency schedule is, because there can be little doubt that *shifts* in the schedule are much more important for our purposes than is the shape of the schedule at any particular time. A study of the determinants of investment must be concerned particularly with the forces that lead to changes in the willingness of businessmen to invest, i.e., to shifts in the marginal efficiency schedule. We must also take some account of changes in the availability of funds, although, as we have seen, such changes may be only very indirectly related to changes in interest rates. Indeed, the availability of funds is likely to be closely related to fluctuations in profits. Profits generate funds available for invest-

ment, and a good earnings record also makes it easier to obtain funds in the capital markets. But changing profits also lead to shifts in the marginal efficiency schedule. Hence, we shall be wise to concentrate on those forces that affect the marginal efficiency schedule of investment.

Shifts in the marginal efficiency schedule are associated with changes in the expected profitability of investment. The most important factors bringing about such changes can be classified under the following headings:[6]

1. The psychological element of expectations, whether or not such expectations have a factual basis.
2. Technological change resulting in new products or new ways of doing things.
3. Changes in the relative prices of labor and capital goods. If labor becomes more expensive, this is likely to stimulate the demand for laborsaving investment. Higher prices for capital goods tend to depress investment, all other things being equal.
4. Changes in the relation between current or expected output, on the one hand, and the stock of capital (or level of capacity) available to produce that output, on the other.
5. The factors affecting the volume of investment in housing (i.e., residential construction). It is obvious that population change plays a particularly important role in this type of investment.

So far we have been talking as if all investment were *net* investment that added to the existing stock of capital. But a substantial part of *gross* investment is to replace capital equipment that has worn out or become obsolete. It is more useful to work in terms of gross investment. Hence we must add another heading:

6. Replacement demand arising from wear and tear and obsolescence.

Let us now look in a little more detail at each of the above sets of investment determinants. Modern cycle and growth theory

[6] The discussion that follows implies that we are concerned with investment only in buildings and equipment. It should be remembered, however, that investment in inventories, while not usually a large fraction of total capital formation, is subject to wide fluctuations and plays a particularly important role in so-called minor cycles. Inventory investment is discussed briefly on p. 140, below, and more fully in later chapters.

concentrates on the fourth (the relation between the stock of capital and the level of output or demand), but we cannot afford to ignore the others.

EXPECTATIONS

The psychological element is clearly important, but there is not much we can say about it at this point. What makes businessmen more optimistic or pessimistic than the facts justify depends on a variety of factors arising out of the way a private-enterprise economy reacts to change and uncertainty. We shall have a good deal to say about this in later chapters, particularly Chapters 8 and 11. At this point it is only necessary to emphasize that expectations play an important role in determining the volume of investment, and that because of this the volume of investment fluctuates more widely than would otherwise be the case. The willingness of business to invest depends not on the actual but on the *expected* profitability of investment.

TECHNOLOGICAL CHANGE

As we shall see in later chapters, technological change is a major factor influencing the volume of investment. New products and new ways of doing things are continually opening up new investment opportunities for businessmen. Because of the technological factor, therefore, we can expect to get spurts in investment, which will generate still further expansion in business activity through the rough workings of the multiplier process.

Technological change has both a long-run and a short-run influence on investment. As we shall point out in Chapter 8, technological change has been one of the most important secular forces operating on the American economy in the last century and a half. Without it, investment and output would have grown much more slowly than they did. Technological change also affects investment in the short run. The introduction of new ways of doing things does not occur smoothly. One wave of innovations may be succeeded in time by another, and there is considerable evidence that this unevenness in the introduction of technological change has had a good deal to do with some of the wide cyclical swings in investment that have occurred in the past. This is a matter that we shall look into further in Chapter 11.

Although there is general agreement that technological improvements affect the profitability of investment and hence the willingness to invest, a good many economists look in another direction to explain the wide cyclical fluctuations in the volume of investment. Technological change is not a simple, clear-cut variable, and it does not have a dimension that is easily measured. Also, and more important, it is not related in a clear and simple way to some other variable whose changes we can explain. Hence, investment resulting from technological change is frequently referred to as "autonomous" investment. It is not induced by changes in some other variable that we have already accounted for. Thus, there has been a tendency to look for other relationships that might explain the behavior of investment—relationships that could be stated precisely and used as a basis for prediction. This helps to explain the modern emphasis in business-cycle theory on our fourth determinant, the relation between output or demand and the capital stock.

RELATIVE FACTOR PRICES

It is clear that, all other things being equal, the marginal efficiency schedule will be depressed if the prices of capital goods rise. This may be of some importance in the long run, but it is doubtful if the demand for capital goods is very elastic with respect to price in the short run. A more important consideration is likely to be changes in the relative cost of labor. The higher are wages in relation to the prices of the things made with labor, the greater will be the incentive to substitute capital for labor. This has been a factor of some importance in stimulating investment in the United States, but here again the long-run effects are more important than those in the short run.[7]

REPLACEMENT

If we are dealing with gross investment, we must also take account of the replacement demand for capital goods. Here we need to make a distinction between *depreciation* and *replacement*. Depreciation represents the loss in value in a durable good resulting from both wear and tear and obsolescence. The charge for depreciation included in the price of the product being sold generates

[7] Cf. J. R. Meyer and Edwin Kuh, *The Investment Decision*, 1957, p. 182.

funds which can be used for replacing fixed assets when they lose their usefulness. But depreciation does not automatically lead to replacement. An asset begins to depreciate from the moment it is acquired, but it will be replaced only some time later. In a growing economy, depreciation is always larger than replacement because, to put it simply, depreciation is figured on the total capital stock now in existence while replacement is necessary only for that part of the present capital stock which was acquired at some point in the past.

Replacement expenditures can vary widely over the business cycle. In particular, they tend to be depressed during cyclical contractions, for then business firms have excess capacity; replacement can be readily deferred; and, when pessimism is general, firms tend to hold on to the funds generated by depreciation charges. When conditions improve, there is likely to be an upsurge of replacement.

Thus, the instability of replacement expenditures provides another reason for the shifting that occurs in the marginal efficiency schedule of (gross) investment.[8]

THE RELATION BETWEEN CAPITAL AND OUTPUT

It is quite obvious that, if output is pressing on capacity and this situation is expected to continue, investment in new capacity will be needed. In general, the faster the demand for output rises, the more new investment is required. On the other hand, if capacity is growing faster than the demand for output, this will tend to depress the level of investment. Thus, the inducement to invest (the marginal efficiency of investment) will change with the changing relation between the existing stock of capital and the demand for output.

We can express this relationship in a fairly simple way as follows. Let k dollars of capital be required to produce one dollar of output. (We assume that prices do not change.) Then, if K represents the needed capital stock and Y stands for output,

$$K_{t-1} = kY_{t-1}$$

and

$$K_t = kY_t.$$

[8] For further discussion of replacement as a factor leading to fluctuations in investment, see R. C. O. Matthews, *The Business Cycle*, 1959.

That is, in time $(t\text{-}1)$ the stock of capital required to produce Y_{t-1} is equal to kY_{t-1}, and a similar situation holds for time t (i.e., the present).

Now let us assume that in period $(t\text{-}1)$ just the required amount of capital, equal to kY_{t-1}, was in existence and that there was no excess capacity. If output now rises from Y_{t-1} to Y_t, the capital stock should rise from K_{t-1} to a new level equal to kY_t. The needed increase in the capital stock, or the amount of required net investment, is

$$I_t = K_t - K_{t-1} = kY_t - kY_{t-1} = k(Y_t - Y_{t-1}).$$

This states that net investment depends directly on the amount of change in output (or income).[9] The smaller the increase in output, the smaller the amount of net investment; the larger the increase in output, the greater the amount of investment. This relation between the volume of investment and the absolute rate of change in output goes under the name of the *acceleration principle*. The coefficient k, which states how much capital is required per unit increase in output, is called the *accelerator*.[10] The acceleration principle states that the volume of investment will decline if the rate of expansion in current output falls off and will rise if the rate of expansion is accelerated. It posits a precise technologically determined relation between the volume of investment and the rate of change in production.[11]

Some economists have seized on the acceleration principle as the missing link with which they can forge a complete theory of the

[9] Practice varies as to what set of lags (i.e., time subscripts) should be used in this relationship. It is fairly common to make current investment depend on the output of the two preceding periods, i.e.,

$$I_t = k(Y_{t-1} - Y_{t-2}).$$

In this case, investment in the current period (t) is a lagged reaction to the change in output between time $(t-2)$ and time $(t-1)$.

[10] It is also referred to as the (marginal) capital coefficient or capital-output ratio.

[11] It is worth noting what sort of marginal efficiency schedule is implied by the acceleration principle. Since nothing affects investment except the amount of change in output, curves such as I_1I_1', and I_2I_2' in Figure 17 will be vertical lines which shift to the right or left depending on whether the change in output is increasing or decreasing. If the accelerator is constant and the change in output is also constant, then the marginal efficiency schedule will be a vertical line fixed in one place.

business cycle, in a way that we shall examine later in this chapter and in Chapter 13. Others take a more restrained view. They emphasize that the stock of capital and the level of output both obviously do influence the rate of investment, but they deny that there is the sort of precise relation between the level of investment and the rate of change in output that the acceleration principle calls for. Instead of writing the acceleration formula

$$I_t = k(Y_t - Y_{t-1})$$

they suggest that a looser sort of relationship holds, which in its most general form may be written

$$I_t = F(K_{t-1}, Y_t).$$

Here the symbol F means merely "a function of" or "is related to." The precise nature of the relation is not stated in the above expression. We can say from other evidence that the relation is of the following sort: I_t will vary inversely with K_{t-1} and directly with Y_t. One explicit way of writing this, as a simple linear equation, is:[12]

$$I_t = -aK_{t-1} + bY_t.$$

Those who prefer this sort of investment relation do not believe that investment is necessarily tied to changes in output in the fixed way implied by the acceleration principle. What is emphasized, rather, is that rising output increases the profitability of investment and should therefore lead to some increase in the amount of investment. The emphasis is on the relation between output and prospective profits, on the one hand, and between prospective profits and investment on the other. This is a much less rigid relationship than that suggested by the acceleration principle. Similarly, other things being equal, a growing stock of capital reduces the prospective profitability of still further investment.[13]

It will be noted that neither of these two ways of relating investment to output and the capital stock—either rigidly through the acceleration principle or more loosely through the broader formulation just described—takes into account the other sets of

[12] It can readily be shown that the acceleration principle is a special case of this somewhat more general formulation.

[13] Cf. Duesenberry, *op. cit.*, chaps. 3–4; also the discussion of business-cycle and growth models in chapter 13 below.

investment determinants previously discussed: expectations, technological change, or the price of labor and of capital goods.[14] Nor is consideration given to the cost and availability of the funds required to undertake any given amount of investment.[15]

THE CASE OF RESIDENTIAL BUILDING

It is fairly obvious that investment in housing stands on a somewhat different footing from business investment in the manufacturing, mining, trade, transportation, or public utility industries. The nature of the housing market is different from that for business plant and equipment; residential buildings are extremely durable, with the result that the existing stock of housing is always very large relative to the amount of new construction; and there are special financing problems associated with house building.

The result is that this highly important form of capital formation (amounting to around 25 percent of gross investment) has been subject to extremely wide fluctuations and has gone through long cycles of wide amplitude. (See pages 243–246.) The set of forces affecting residential building is quite complex. At the risk of some oversimplification, we can say that these influences have to do primarily with changes in the size, geographical distribution, and age composition of the population, the level of income, and the cost and availability of financing. Population change plays a more important role in residential building than it does in the other forms of investment.[16]

The factors affecting housing investment can be thought of as a somewhat special case of the capacity-demand relationship affecting nonhousing investment which was discussed on pages 129–131. Population change and growth in income call for an increase in the stock of housing and stimulate building. At the same time, such building increases the stock of housing, which tends to reduce the stimulus to further construction.

[14] Depreciation and the need for replacement are taken into account. Depreciation reduces the capital stock and therefore increases the need for (replacement) investment.

[15] Indirectly, the broad formulation does take some account of this factor. Given the stock of capital, a larger output increases profits, which can go into investment, and this also increases the firm's ability to float new securities.

[16] For a more detailed and technical discussion of the factors affecting housing investment, see Duesenberry, *op. cit.*, chap. 7; also Matthews, *op. cit.*, chap. 6.

THE LONG-RUN WORKING OF THE ACCELERATION PRINCIPLE

The acceleration principle has played a key role in recent years in economists' writings about both business cycles and economic growth. There is no doubt that the principle contains an important element of truth, and even economists who do not make it a cornerstone of their theories of the business cycle attach considerable importance to it. In view of the important role it plays in the modern analysis of both cyclical fluctuations and economic growth, we shall examine the acceleration principle in some detail in the following pages. Our conclusion from the analysis can be briefly stated here. Interpreted properly, the acceleration principle offers an important tool for business-cycle analysis, but it is not enough with which to build a realistic theory of the business cycle. As the discussion in the preceding sections suggests, the determinants of investment fluctuations are much more complex than the acceleration principle implies, and we can get at these complex forces only by the sort of detailed analysis of the functioning of the economic system that is contained in later chapters.

The acceleration principle has its greatest validity if it is interpreted as a long-run relationship and is considered with reference to a single industry rather than the whole economy. In this form, the dependence of investment on the rate of change of output can be illustrated by the growth of almost any important industry. Net investment, over and above replacement, results in additional productive capacity; additional capacity is not justified, however, unless there is a market for the increased output. Conversely, if present capacity is full employed, further expansion of output requires additions to capacity—that is, new net investment. If new output should continue to rise, but at a decreasing rate, smaller additions to capacity would become necessary; the rate of net investment in the industry would decline. If eventually the level of output should show no further increase but remain stable, the amount of additional capacity required would fall to zero. No further net investment, over and above replacement, would take place until required by new technological changes.

Some such relationship as this must hold true, at least in a

rough way, in the long-run growth of particular industries. The American railroad network, for example, was largely built between 1850 and 1914. Investment in new roadbed, track, and bridges was at a higher level in the half-century preceding World War I than it was later, when little additional railroad mileage was being built and only replacement and modernization were necessary. All industries tend to follow a typical pattern of growth, in which rapid increase in output is followed by less rapid expansion and eventually by relative stability or decline. This implies that expenditure on new plant and equipment will eventually begin to fall unless the effect of the decline in the rate of growth is offset by growing replacement needs arising from physical depreciation and from obsolescence created by technological change.

We have here, then, an important factor affecting the level of investment in individual industries. Subject to important modifications to be considered later, the volume of net investment in particular industries depends on the changing absolute rate of growth in output in these industries. Investment means primarily purchase of capital goods. We can therefore express the same principle by saying that, in addition to technological change, a basic factor determining the demand for capital goods is the rate of growth in industry generally. Stable production in the capital-goods industries requires continued expansion in the rest of the economy.

This relationship between investment and output cannot be expected to hold precisely. Continued technological change may require additional investment, even if no increase in output has taken place. Perhaps even more important in the short run, investment will reflect the current state of businessmen's anticipations. They may "build ahead of the market" if they are sufficiently confident, or they may refuse to invest despite an increase in demand because they lack confidence that the higher level of demand will endure. Further, as an industry continues to expand, replacement requirements will rise and thus to some extent—in some cases perhaps completely—offset the decline in *net* investment.

While the long-run relationship, then, is a rough one, it has important cyclical implications. The growth of new industries stimulates investment; the maturing of these industries retards investment. A succession of innovations, taken in conjunction with the

rough working of the acceleration principle, the volatility of business expectations, the availability of an elastic credit supply, and the secondary effects described by the multiplier, is sufficient to create cyclical fluctuations in business. The process by which they may be created is described in later chapters.

Thus far we have alluded only to a long-run relationship between net investment and changes in output and have emphasized that at best this relation is a rough one, subject to important qualifications. Economists, however, have attempted to generalize the principle involved in order to explain the amplitude and timing of the short-run fluctuations in the production of durable goods generally—both capital goods and durable consumers' goods. It is in this form that it has been most widely used in business-cycle analysis.[17]

THE ACCELERATION PRINCIPLE IN THE SHORT RUN

The principle of acceleration is used to explain the wide fluctuations that occur in the demand for producers' and consumers' durable goods by relating this demand to the rate of change in the demand for the commodities or services that these goods produce. The demand for the durable goods—whether machinery, buildings, or automobiles—is a derived demand; it is derived from the demand for the finished goods and services produced by the durable goods.

HOW THE ACCELERATION PRINCIPLE OPERATES

The acceleration principle states that the derived demand for durable goods will vary with the absolute amount of change in the production of the finished goods or services produced by the durable goods. Fluctuations in the output of durable goods will ordinarily be wider than in the finished goods and services produced with their aid, and the turning points in the former will precede

[17] For a good summary of the earlier literature on the acceleration principle, see G. Haberler, *Prosperity and Depression* 4th ed., 1958, pp. 85–105. For more recent critical evaluations, see A. D. Knox, "The Acceleration Principle and the Theory of Investment: A Survey," *Economica,* vol. 19, August, 1952, pp. 269–297; S. C. Tsiang, "Accelerator, Theory of the Firm, and the Business Cycle," *Quarterly Journal of Economics,* vol. 65, August, 1951, pp. 325–341; R. S. Eckaus, "The Acceleration Principle Reconsidered," *ibid.,* vol. 67, May, 1953, pp. 209–230; G. H. Fisher, "A Survey of the Theory of Induced Investment," *ibid.,* vol. 18, April, 1952, pp. 474–494.

those in the latter.[18] These relationships are illustrated in the three examples in Table 12.

In Section A of the table, we consider the case of durable capital goods. It is assumed that five dollars worth of capital equipment is required to produce one dollar of finished product, and that this relationship does not change. The capital equipment is assumed to last 10 years, so that one-tenth is replaced each year, except that none of the new equipment will be replaced until it is 10 years old. Replacement demand is therefore 50 millions per period. We also assume that all prices remain unchanged, so that the value figures reflect accurately changes in physical volumes.

When finished output rises by 5 millions in period 2, the stock of capital equipment must rise by 25. Total demand for equipment, including that for replacement, rises from 50 to 75 millions. The increase in finished-goods output is magnified five times in the increase in demand for capital goods. In the third period, a further increase of 10 in the output of the finished product creates a total demand for capital goods, including replacement requirements, of

TABLE 12. Illustrations of the Working of the Acceleration Principle
A. Durable Capital Goods
(Figures in millions of dollars)

Period	(1) Output of Finished Product	(2) Stock of Capital Goods Required	(3) Addition to Capacity Required	(4) Replacement Demand	(5) Total Demand for Capital Goods
1	100	500	—	50	50
2	105	525	25	50	75
3	115	575	50	50	100
4	120	600	25	50	75
5	122	610	10	50	60
6	120	600	−10	50	40
7	115	575	−25	50	25
8	105	525	−50	50	0
9	100	500	−25	50	25

[18] It is not inevitable that the fluctuations in durable goods be wider, but the size of the accelerator and the usual length of business cycles will ordinarily have this result. Cf. W. J. Baumol, "Acceleration Without Magnification," *American Economic Review*, vol. 46, June, 1956, pp. 409–412.

B. Consumers' Durable Goods:
Dwelling Units

Period	(1) Number of Families	(2) Annual Rental Value of Housing Units Required [a]	(3) Value of Total Units Required [b]	(4) Replacement Demand Units	(5) Replacement Demand Value [b] (thousands)	(6) Total Construction Units [c]	(7) Total Construction Value [b] (thousands)
1	1,000	$600,000	$6,000,000	40	$240	40	$ 240
2	1,100	660,000	6,600,000	40	240	140	840
3	1,300	780,000	7,800,000	40	240	240	1,440
4	1,400	840,000	8,400,000	40	240	140	840
5	1,450	870,000	8,700,000	40	240	90	540
6	1,500	900,000	9,000,000	40	240	90	540
7	1,500	900,000	9,000,000	40	240	40	240

[a] Assumes unchanging rental value of $600 per period for each unit.
[b] Assumes unchanging original cost of $6,000 per dwelling unit.
[c] Equals increase in number of families plus replacement demand.

C. Inventories
(In thousands of dollars)

Period	(1) Sales	(2) Required Inventories [d]	(3) Change in Inventories	(4) Purchases Required to Replace Goods Sold [e]	(5) Total Purchases [f]
1	500	250	—	500	500
2	600	300	50	600	650
3	800	400	100	800	900
4	1,000	500	100	1,000	1,100
5	1,100	550	50	1,100	1,150
6	1,100	550	—	1,100	1,100
7	1,000	500	−50	1,000	950
8	800	400	−100	800	700
9	600	300	−100	600	500
10	500	250	−50	500	450
11	500	250	—	500	500

[d] Assumed to be 50 percent of one period's sales.
[e] See footnote 20 on page 140.
[f] Sum of columns (3) and (4).

100. Thus a total increase in demand for the finished product of 15 percent, from 100 to 115, has increased the demand for capital goods from 50 to 100, or by 100 percent.

In the fourth period, output of finished goods rises further, but there is now a substantial decline in the demand for capital goods. This occurs because of the decline in the amount of increase in finished-goods production in the fourth period. Finished-goods output rises by only 5, which requires only 25 in additional capacity, compared to the 50 of additional capacity required in the preceding period. In period 5, there is a further small increase in output; but, because this requires a smaller addition to capacity than in the preceding period, demand for capital goods declines again.

From period 6 on, finished-goods output actually declines. In each period, therefore, less capacity is required than in the preceding period. This accounts for the minus signs in column (3). To the extent that the reduction in capacity is effected through failure to replace fully depreciated machines, total demand for capital goods can fall below the replacement rate, as it does in column (5).

Clearly, the important relationship that is involved here is the ratio of required capacity to output of the finished product. This ratio, which is called the acceleration coefficient or the accelerator, in turn depends on two factors: the durability of the equipment and the extent to which costs other than depreciation enter into the price of the final product. In our hypothetical illustration, we assumed that the equipment had an average life of 10 years and that depreciation accounted for half the price of the finished product. These assumptions established the relation of five dollars worth of installed capacity per dollar of finished-goods output, and this meant an additional demand of five millions for new equipment for each increase of one million in finished-goods output. The more durable the equipment and the less important other costs, the greater will be the accelerator. If the equipment had had a life of 20 years, and if the other costs had been half those assumed in Table 12, the ratio of capital to output would have been raised to 10 to 1, and this would have been the acceleration ratio or accelerator.

The accelerator relates the absolute, not percentage, increase in finished-goods output to the absolute amount of new capital equipment required. To make this comparison, both series must be ex-

pressed in dollar terms, through multiplication of the physical volume of output and of purchases of capital equipment by unchanging prices. We cannot express the accelerator in terms of a comparison between percentage changes. The ratio between the percentage changes in finished-goods output and in the demand for capital goods depends on how the output of finished goods varies. The relation between the percentage changes is different for each period in the table.[19]

Because of the operation of the acceleration principle, also, the swings in the demand for capital goods are likely to show a lead over the corresponding swings in finished-goods output. Thus, the peak in demand for capital goods occurs in period 3; that in finished-goods output, in period 5. There is also a lead at the low point in the table. This timing relationship arises because of the assumption that the amount of increase in finished-goods output will begin to decline before there is an actual fall in the level of output. The demand for additional capital goods depends on the change in finished-goods output, and not on its absolute level.

The same principles operate in the case of consumers' durable goods, of which the two most important are houses and automobiles. Section B of Table 12 illustrates the operation of the acceleration principle in the case of housing. The relevant relation is one between the demand for housing services, which may be measured by annual or monthly rental value, and the demand for new dwelling units as reflected in the value of new construction.

Thus, in period 2 there is an increase of 10 percent in the number of families and in the demand for shelter which, if satisfied at the average rental assumed, increases rents by $60,000. Since we assume a ratio of 10 to 1 between the cost of housing units and rent per period, the increased demand for housing services of $60,000 leads to an increase of $600,000 in the value of dwelling units required. Put simply, a hundred new families need a hundred new dwelling units which will cost $6000 each. Replacement demand, assuming that the average life of a dwelling unit is twenty-five years, is 40 per year. Total construction, therefore, rises from 40 to 140—or, in dollar terms, from 240,000 to 840,000. An increase of $60,000

[19] See, for example, Simon Kuznets, "Relation Between Capital Goods and Finished Products in the Business Cycle," in *Economic Essays in Honor of W. C. Mitchell,* 1935, esp. pp. 225–256.

in the demand for housing services has given rise to an increase of $600,000 in the value of new dwelling units built. In the next period, a further increase in demand for shelter, reflected in a further rise in the rental value of required dwelling units of $120,000, causes new construction to rise by a further $600,000. As soon as the amount of increase in the demand for housing services begins to decline, as it does in period 4, the volume and value of new construction begin to decrease—although the level of demand for housing services continues to rise until period 6. Taking the table as a whole, an increase of 50 percent in the number of families and hence in the demand for housing services resulted first in a rise in new construction of 500 percent (from 40 to 240 units) and a subsequent decline of 83 percent (from 240 to 40 units). In this case, the acceleration ratio is 10 to 1. Each dollar of additional demand for dwelling services gave rise to 10 dollars worth of demand for new construction.

Stocks of goods maintained by businessmen—their inventories of raw materials, work-in-process, and finished goods—also provide a channel through which the acceleration principle may operate. This is illustrated in Section C of Table 12, which assumes inventories are maintained at a constant percentage (in this case, 50 percent) of sales.[20] As long as sales are increasing, reorders will rise relatively faster than sales, because reorders must be large enough to cover the increase in required inventories as well as the goods actually sold. When, in period 6, sales fail to increase further, even though they do not decrease, business purchases decline because there is no further need to expand inventories.[21] On the downswing, purchases fall faster than sales because the reduction in sales permits some reduction in inventories. When sales stop declining, even though they do not rise, purchases increase because there is no further reduction in inventories.

LIMITATIONS OF THE ACCELERATION PRINCIPLE

So much for the explanation of how the acceleration principle operates, given the assumptions that are involved. Returning to our

[20] For simplicity, we have assumed that inventories and purchases are valued at sales prices. In practice, they are valued at cost, which is less than the sales price. The result of this is to reduce the acceleration ratio, which relates the change in value of sales to the change in the value of inventories.

[21] Note that the amount of net investment in additional inventories starts to fall in period 5, when the rate of increase in sales first begins to decline.

main problem, to what extent does the acceleration principle satisfactorily account for the cyclical fluctuations in investment that we find in practice? Are cyclical fluctuations in investment actually induced by current or prior changes in consumption or total output in any clearly observable way? The factual evidence available is largely in the negative.[22] Whatever value the acceleration principle may have as an explanation of certain long-run relationships, it seems to be of limited value in explaining short-period fluctuations in investment. Private capital formation depends on a good deal more than merely the current amount of change in the output of finished goods and services. This is not surprising when we look more carefully at the assumptions on which the acceleration principle is founded.

The basic assumption underlying the acceleration principle is that the ratio of capital equipment to output remains constant. Unfortunately, this assumption never holds in practice, and for several reasons. In the first place, every cyclical expansion begins from a position of excess capacity in industry generally. The upswing in finished-goods production can go on for some time on the basis of already available capacity. In the early stages of the upswing there is no need to order new capital goods, over and above replacement needs, to produce the additional output called for by rising demand. At the bottom of the depression, the capacity-output ratio is abnormally high, and businessmen seek to reduce it by not expanding capacity as output begins to rise.[23]

When will businessmen begin to expand capacity as the output of

[22] For some of this factual evidence see J. Tinbergen, "Statistical Evidence on the Acceleration Principle," *Economica,* vol. 5, n.s., May, 1938, pp. 164–176; Kuznets, *op. cit.;* Thor Hultgren, *American Transportation in Prosperity and Depression,* 1948, pp. 157–169; L. M. Koyck, *Distributed Lags and Investment Analysis,* 1954, chap. 4; and Meyer and Kuh, *op. cit.* The evidence for the acceleration principle becomes more favorable if the latter is interpreted flexibly—for example, if investment is related to changes in output over a number of past periods, if proper allowance is made for unutilized capacity, and some account is taken of expectations. See Koyck, *op. cit.,* and Robert Eisner, "A Distributed Lag Investment Function," *Econometrica,* vol. 28, January, 1960, pp. 1–29.

[23] Offsetting this to some extent may be an increase in gross capital expenditures due to deferred replacement. In depressions, equipment continues to depreciate, but businessmen can put off making replacement expenditures. When the upswing begins, returning confidence and improved financial position may lead to a spurt of replacement demand. It is worth noting here that failure to replace outworn equipment in depressions is scarcely ever sufficient to prevent excess capacity from developing. Output ordinarily falls more than capacity can be reduced by deferring replacement.

finished goods continues to rise? They may begin before or after the excess capacity, however we define that troublesome concept, is eliminated. If businessmen have no confidence that the higher level of demand will continue for very long, or if they have trouble raising the necessary funds, they may not expand capacity at all. They can try to squeeze more production out of their existing plant and equipment—through overtime and extra-shift operations, through using relatively inefficient stand-by facilities, and so on.

But suppose they finally do become willing to expand their facilities. There is no reason to expect that the expansion of capacity will be precisely that required to take care of the *current* increase in output. As a matter of fact, this is ordinarily an inefficient way of increasing capacity. Businessmen are likely to look ahead and will, at what they think is the most propitious moment, add facilities to provide for the anticipated expansion in output over some considerable period in the future, even if they will not be able to operate all of the additional facilities at their economically full capacity as soon as they are completed. Increases in capacity depend more on anticipations about future output than they do on currently observed changes in demand.

In some cases, the relevant business anticipations may cover a considerable period into the future. This is particularly likely to be true where very durable, expensive installations are involved, as in the case of the public-utility industry. Such long-term investment is not likely to bear a constant relation to current changes in the demand for the industry's product.

It is worth noting that investment in industry generally is likely to reach its peak at about the peak in the cycle of general business activity, even though the expansion in output may have begun to taper off. This is true because the state of business confidence is likely then to be most favorable for expansion. At this time, also, security prices are high, and the investment market is receptive to new issues. Replacement and modernization are likely to take place on a large scale at this stage, adding to the demand for capital goods.

The stage of buoyant optimism, leading to a willingness to anticipate future increases in demand, may well begin before excess capacity is eliminated in the larger part of industry. As improving conditions generate business confidence, firms accelerate replacements that were deferred during depression; they begin to adopt improve-

ments that have become available in order to reduce costs and improve their competitive position; and they begin to plan to expand capacity if they are confident that demand will continue to rise and if their financial position and the capital markets are favorable. Some excess capacity is a normal situation in many industries, and expansion may well ensue before output catches up with existing capacity.

When capacity is expanded, and also when it is replaced, the ratio between capital investment and output may change significantly. If there are new improvements to be adopted, the new equipment may involve more or less investment per unit of output than the old equipment.

Strictly speaking, the acceleration principle relates the demand for capital or durable goods to the change in output of those particular goods or services that the former produce. If we attempt to apply the principle to the relation between total investment, in all industry, and changes in total output, a further qualification becomes necessary. Output in one industry may be rising while in another industry it may be falling. Total output may not be changing. Yet net new investment in the expanding industry is not likely to be offset by an equivalent amount of disinvestment in the contracting industry. Hence, there may be total net investment even through there is no net increase in total output. We must also recognize that the amount of investment required by a given change in total output will vary, depending on the composition of the change in output. A million-dollar increase in the production of electric power, in which the ratio of capital to output is high, will induce more investment than a million-dollar increase in production in an industry that has a low capital-output ratio, for example, light manufacturing.[24]

We must remember also that, even when we relate investment in a single industry to changes in total output in that industry, we are comparing investment in a variety of different kinds of plant, equipment, and inventory with changes in the output of the variety of products produced in that industry. The use of such broad aggre-

[24] For an excellent discussion of the relation between differential rates of change in the output of different industries and the behavior of aggregate investment, see B. G. Hickman, "Diffusion, Acceleration, and Business Cycles," *American Economic Review,* vol. 49, September, 1959, pp. 535–565. This is another study, in addition to those previously cited, that casts doubt on the explanatory value of the acceleration principle.

gates may well conceal whatever effect the acceleration principle might have. The situation is worse when we relate total investment in the economy as a whole—in everything from power plants to changes in the retailers' inventories—to changes in total output or in the output of consumers' goods. We said in Chapter 2 that the study of broad aggregates is useful in the analysis of business-cycle fluctuations, but that a study of broad aggregates alone is not enough. This is the case here.

INTERACTION OF THE MULTIPLIER AND THE ACCELERATOR

With this warning, let us look briefly at the attempts that have been made to show how the multiplier and accelerator may interact to create cyclical fluctuations. The simplest formulation of this relation is that developed by Hansen and Samuelson, which can briefly be summarized as follows.[25]

Suppose that an initial increase in I or G takes place for some reason and that this increased expenditure is continued for some time. Let us assume that $\frac{\Delta C}{\Delta GNP}$ is .5 and that each dollar of additional output requires an additional eight tenths of a dollar of investment in new capacity. The initial expenditure leads (via the multiplier) to increased consumption in the next period, but the increased consumption requires additional private investment (acceleration principle); this, along with the continued investment expenditure, generates a further increase in consumption, which in turn requires more investment, and so on. This is illustrated in Table 13.

Let us assume that, in period 2, I or G expands by 100 million above its previous level (column [1]). With $\frac{\Delta C}{\Delta GNP}$ of .5, this will increase consumption by 50 million in period 3. We assume that each dollar of increase in output requires 80 cents of investment in new capacity in the period following that in which the increase in

[25] Cf. Paul Samuelson, "Interactions Between the Multiplier Analysis and the Principle of Acceleration," reprinted in *Readings in Business Cycle Theory*, pp. 261–269; A. H. Hansen, *Fiscal Policy and Business Cycles*, 1941, pp. 276–283, and *Business Cycles and National Income*, 1951, chap. 11. Some of the more recent theories based on this sort of interaction are summarized in Chapter 13, below. In the original Samuelson formulation, investment was made to depend on the increase in consumption only. We shall follow the current and better practice of making investment depend on the change in total output.

income takes place. Thus, the increase in GNP of 100 million in period 2 requires 80 million of investment in period 3. This represents the working of the acceleration principle. Thus GNP in period 3 is 230 million—100 of autonomous investment in column 1, 50 of induced consumption in column 2, and the 80 of new investment resulting from the accelerator shown in column 3. In period 4, consumption is 115 (one half of 230); new investment required by the

TABLE 13. Interaction of the Multiplier and Accelerator
(In millions of dollars) [a]

Period	(1) Autonomous Increase in I over Base Period	(2) Induced Consumption [b]	(3) New Investment Required [c]	(4) Total Change in GNP from Base Period
1	0	0	0	0
2	100	0	0	100.0
3	100	50.0	80.0	230.0
4	100	115.0	104.0	319.0
5	100	159.5	71.2	330.7
6	100	165.4	9.4	274.8
7	100	137.4	−44.7	192.7
8	100	96.3	−65.7	130.7
9	100	65.3	−49.6	115.7
10	100	57.9	−12.0	145.9
11	100	73.0	24.2	197.2

[a] Rounded to the nearest tenth.
[b] Equal to one half of previous period's GNP as given in column 4.
[c] Equal to eight tenths of increase in GNP in column 4 during preceding period; i.e., an accelerator of .8 is applied to the increase in GNP between period $t - 2$ and period $t - 1$.

accelerator is 104 (eight tenths of the increase in GNP in the preceding period); and, adding in the continuing autonomous investment in column 1, total GNP in period 4 is 319 higher than in the base period from which we began.

If this chain of induced effects is followed through succeeding periods, a cycle in GNP results. If we took other values for $\frac{\Delta C}{\Delta GNP}$ and for the accelerator, we would get other types of cyclical fluctuations which might show smaller or greater amplitudes with the passage of time.[26]

[26] Cf. Samuelson, *op. cit.* See also J. R. Hicks, *A Contribution to the Theory of the Trade Cycle,* 1950; and Matthews, *op. cit.,* chap. 2.

Table 13 illustrates a simple model of the business cycle that is based entirely on the interaction of the multiplier and the acceleration principle. We can describe this model in simple algebraic terms as follows. We assumed that consumption was equal to one half of GNP (which we shall call Y) during the preceding period. Using the subscripts t and $t-1$ to identify our periods, we can write the consumption function as

$$C_t = .5Y_{t-1}.$$

We also assumed that induced investment was equal to eight tenths of the *change* in GNP during the preceding period—that is, between period $t-2$ and period $t-1$. To this we have to add the 100 of autonomous investment which continues unchanged from period 2 on. We can therefore say that investment in any period following period 2 will be:

$$I_t = .8(Y_{t-1} - Y_{t-2}) + 100.$$

We also have our basic identity

$$Y_t = C_t + I_t.$$

We can now substitute into this last equation the corresponding expressions for C_t and I_t derived from the two preceding equations. This gives us

$$\begin{aligned} Y_t &= .5Y_{t-1} + .8(Y_{t-1} - Y_{t-2}) + 100 \\ &= 1.3Y_{t-1} - .8Y_{t-2} + 100. \end{aligned}$$

This is called a "difference equation," which can be used to obtain directly the figures for Y_t in column 4 of Table 13.[27] It states that current output or income depends on the output of the two preceding periods. Thus, if we know what output is for two successive initial periods, this model will then proceed to generate all succeeding levels of output in the manner illustrated in Table 13.[28]

[27] For a good elementary discussion and economic interpretation of difference equations, see W. J. Baumol, *Economic Dynamics,* 2nd ed., 1959. The equation in the text is a "second-order" difference equation since it contains a term that is lagged by two periods.

[28] The reader is invited to compare the set of equations leading to this *dynamic* model with those comprising the *static* Keynesian model discussed in Chapter 4 (pp. 79–85). Note the following differences: (1) We have eliminated the dependence of I on the interest rate and therefore need no equation (liquidity preference schedule) to explain the interest rate, and (2) we have introduced *lagged* relationships, in which C and I are made to depend on past values of Y. It is this second difference that makes this model *dynamic,* so that the model generates a changing Y as time passes—instead of the unchanging static equilibrium that results from the Keynesian system.

We can now give a more general formulation of the multiplier-accelerator model. Let a be the marginal propensity to consume; let the accelerator be equal to k; and we shall assume some particular amount of autonomous investment equal to A. Following the procedure used before, we have

$$\begin{aligned} Y_t &= C_t + I_t \\ C_t &= aY_{t-1} \\ I_t &= k(Y_{t-1} - Y_{t-2}) + A \end{aligned}$$

and therefore

$$\begin{aligned} Y_t &= aY_{t-1} + k(Y_{t-1} - Y_{t-2}) + A \\ &= (a + k)Y_{t-1} - kY_{t-2} + A. \end{aligned}$$

What kind of fluctuations this model will generate depends on the values of a and particularly of k. Among the possibilities are: (1) Y will move steadily upward (or downward) at an increasing rate; (2) Y moves upward (or downward) at a decreasing rate, converging toward a new equilibrium level; (3) Y fluctuates through a series of cycles of wider and wider amplitude; (4) Y fluctuates through a series of cycles of smaller and smaller amplitude until the cycles virtually disappear; and (5) as a very special case, Y fluctuates through a series of cycles that are of constant amplitude. Of these possibilities, cases (1) and (3) give us an *explosive* movement, while cases (2) and (4) are examples of a *damped* movement toward a new equilibrium position.

Of these five cases, only the fifth gives us a regularly recurring cycle of constant amplitude and duration. In the other cases, we obviously have to bring in some additional explanation to get the kind of recurring cycles that we have in the real world. If we have an explosive model (which results from relatively large values for a and k), we have to assume that some sort of restraint is imposed on the economy to keep it from running away. One such restraint would be the ceiling on output imposed by full employment. With very low values for the multiplier and accelerator, we would get a damped system in which cycles would tend to disappear unless we assumed that new disturbances—say, in the form of new autonomous changes in government spending or investment—were constantly occurring to touch off a new cyclical movement.[29]

[29] For more detailed discussion of the dynamic characteristics of multiplier-accelerator models, see the references cited in footnote 26; also Baumol, *op. cit.*

We have taken the time to explain in some detail the possible dynamic interaction between the multiplier and accelerator because, as we shall see in Chapter 13, this interaction has played a very important role in the recent theoretical literature on cyclical fluctuations and growth.[30] Our own opinion is that the relationship, while having some illustrative value, does not carry us very far in explaining what occurs in the business cycle. We explained in Chapter 5 that the marginal propensity to consume GNP or national income does not remain constant over the cycle, and in the preceding sections of this chapter we emphasized that the volume of investment is more a product of business expectations—as to both the short- and the long-run future—than it is of current changes in output.

Although the acceleration principle does not provide a satisfactory explanation of short-run fluctuations in total investment, it is of value if we do not try to do too much with it. It helps to point up at least two facts of importance in business-cycle analysis. First, it brings out the long-run relationship between the rate of growth in an industry and the volume of net investment that will result in the absence of continued technological change. A tendency toward slackened growth in industries that have been expanding rapidly exerts a downward pressure on the volume of investment. Second, the acceleration principle offers a strong reminder that the behavior of current output does in some way influence the volume of investment, and that the fluctuations in the latter will be relatively wide. What we have emphasized, however, is that there is in fact no rigid relation whereby the volume of investment is automatically determined by the current rate of change in output. A more realistic causal sequence is one that runs from present and recent changes in output to business expectations regarding the future to the volume of investment. And this is never more than part of the story, since there are also other factors that influence the expected profitability of investment. Simple models of the business cycle that depend on the mechanical interaction of a constant multiplier and accelerator emphasize relationships that we cannot neglect, but they are much too simple and artificial to provide a satisfactory picture of the complex interrelationships that generate business fluctuations in the real world.

[30] See also the discussion of the Harrod-Domar growth model, pp. 150–153, below.

SAVING, INVESTMENT, AND THE BUSINESS CYCLE

A brief summary may be desirable at this point. We saw in Chapter 4 that the level of income and of prices results from the interaction of aggregate demand and aggregate supply. In the determination of aggregate demand, the key variables are consumption and investment. In Chapter 5 we looked at the determinants of consumers' expenditures, and in this chapter we have examined the possible determinants of investment. So far as the business cycle is concerned, our broad conclusions can be summarized as follows.

1. There is, historically, a reasonably close, systematic relation between consumers' expenditures and disposable income and, to a somewhat less extent, between consumers' expenditures and GNP. The relationship, however, does not hold closely for short-period changes in consumption and income, particularly for consumption and gross national product. We are not certain as to the precise nature of the changing relationship between consumers' expenditures and income over the course of the cycle, but it is fairly clear that the percentage of the GNP saved tends to rise during upswings and fall during downswings. In any event, it is clearly established that consumption does not rise, in either the short or the long run, by the full amount of any increase in income, and hence additional investment or government spending is necessary to offset the increases in saving generated by a rise in incomes.

2. We saw, further, that in the mutual interactions between consumption, investment, and income, investment rather than consumption is the dynamic, initiating element. Changes in any kind of spending—by consumers, government, or business—have secondary repercussions in creating still further changes in income through the sort of process described by the theory of the multiplier. We saw that the conventional multiplier analysis is much too artificial in its simplifying assumptions—particularly in its assumption of an unchanging propensity to consume and its neglect of induced investment. Nonetheless, changes in investment, which is the most dynamic element in aggregate demand, do have important multiplying effects on income and employment because they induce still further changes in total spending.

3. This suggests that private investment is of crucial importance in business-cycle fluctuations, and it is therefore imperative to locate

the causes of the wide swings in capital formation that we observe in practice. It is here that dynamic analysis in terms of a few broad aggregates tends to break down. Although, in an important sense, capital formation can be said to depend on continued expansion of current output, our investigation of the acceleration principle showed that the wide cyclical fluctuations that occur in private investment cannot be explained by current changes in total output or in consumption. To explain what happens to private investment we must turn to the details of technological change and to the other factors that operate on businessmen's expectations and their willingness to invest.

We can set up a simple relationship between consumption and consumers' incomes. Somewhat less successfully, we can set up a simple relationship between investment and total spending. But we cannot, in our present state of knowledge, set up a simple relationship that will explain fluctuations in investment in terms of one or two other elementary variables.

A NOTE ON ECONOMIC GROWTH

Increasingly since World War II, economists, public officials, and business and labor leaders have come to emphasize the need for a satisfactory rate of growth. Rivalry between Russia and the United States is thought of partly in terms of relative rates of growth; immense efforts are being made to stimulate growth in the poorer countries of Asia, Africa, and Latin America; and there has been a tremendous upsurge of interest in the details of the process by which economic growth takes place.

As a result, a large body of literature on "growth theory" has developed since World War II, and a good many economists have experimented with "growth models"—i.e., with abstract theories that attempt to specify the main determinants of the rate of growth of total output. Two of the pioneers in this field were Roy Harrod and Evsey Domar, who independently put forward very similar growth models shortly after the end of World War II.[31] Both models depend, essentially, on the interaction of the accelerator and multiplier. To

[31] See R. F. Harrod, *Towards a Dynamic Economics*, 1948, and E. D. Domar, *Essays in the Theory of Growth*, 1957. The latter includes articles first published in 1946–1948. Harrod's first formulation of his model dates back to an article in 1939, which received very little attention until the later appearance of the book just cited.

that extent, they are subject to the criticisms that we have advanced in the preceding pages. But, as we saw also, the multiplier and accelerator are likely to be more useful as explanations of rough tendencies in the long run than they are if used to explain short-run fluctuations in consumption and investment. Partly for this reason, the Harrod-Domar type of growth model provides a useful way of looking at some of the conditions necessary for rapid and sustained growth and offers another example of the way in which economists have used the multiplier and accelerator.[32]

This type of growth theory assumes that, as a rough approximation, the marginal and average propensities to consume and save are constant and that, therefore, the multiplier is constant. It also assumes that the accelerator is constant. Then the question to be asked can be phrased in the following way: What percentage rate of growth will insure that aggregate demand and aggregate supply will grow in such a way that continued growth at that same rate remains possible?

We can look at the problem in this way. An increase in investment affects both aggregate demand and aggregate supply. Through the multiplier, there will be a rise in spending and income, i.e., in aggregate demand. But the same increase in investment causes a growth in productive capacity (aggregate supply) through the accelerator relationship. For growth to be maintained, the increase in aggregate demand must equal the increase in aggregate supply.

We can restate these relationships as follows.[33] An increase in investment (ΔI) leads to an increase in income (ΔY) in accordance with the multiplier formula:[34]

$$\Delta Y = \frac{\Delta I}{1 - c} = \frac{\Delta I}{s}. \tag{1}$$

Here s is the marginal propensity to save, which is 1 minus the marginal propensity to consume.

The same increase in investment increases the stock of capital and

[32] A good general reference on the issues dealt with in recent growth theory is D. Hamberg, *Economic Growth and Instability*, 1956. See also chapter 13, pp. 371–380, below.

[33] The following formulation follows Domar (with some slight modifications) rather than Harrod. See the former's *Essays in the Theory of Economic Growth*, pp. 70–108.

[34] See p. 110, above.

the economy's ability to produce additional output. How much additional output can be produced depends on the accelerator or capital-output ratio. If it takes k dollars of investment to create the capacity for one dollar of additional output, it will take $k\Delta Y$ dollars of investment to produce an increase in output equal to ΔY. Thus,

$$I = k\Delta Y$$

and

$$\Delta Y = \frac{I}{k}. \tag{2}$$

We now have two equations for ΔY. Equation (1) gives the increase in aggregate demand and equation (2) gives the increase in potential aggregate supply. But sustained growth requires that they be equal. Setting them equal, we have:

$$\frac{\Delta I}{s} = \frac{I}{k} \tag{3}$$

and

$$\frac{\Delta I}{I} = \frac{s}{k}. \tag{4}$$

Equation (4) says that, to maintain continued equality of demand and supply, investment must grow at a percentage rate equal to $\frac{s}{k}$. But since saving (and therefore investment) are assumed to be a constant percentage of Y, this means that Y must also grow at this rate, so that we can write

$$\frac{\Delta Y}{Y} = \frac{s}{k}.$$

This is the Harrod-Domar formula for the equilibrium rate of growth. It states that if we are to have sustained growth, with demand and capacity keeping in equilibrium with each other, output must grow at a rate equal to the propensity to save divided by the accelerator. If the saving rate is 10 percent and the accelerator is 2, the equilibrium rate of growth is 5 percent.

This is a simple yet striking formula. It is valuable in pointing up two basic determinants of growth—namely the rate of saving and the required ratio of capital to output (the accelerator). If we wish to

increase the rate of growth, we must save a higher fraction of the national income or else we need to have a smaller k. A smaller k means using methods of production that do not require so much capital per unit of output. The relevance of these considerations for the so-called underdeveloped countries, which are trying to accelerate growth and industrialize while their rate of saving is relatively low, is obvious.

The formula also contains some suggestions for the wealthier, more advanced countries. For example, a wealthy country with a high saving rate needs to grow relatively rapidly if *ex ante* saving is not to run ahead of *ex ante* investment and thus precipitate a decline. Also, the less capitalistic are methods of production (the smaller is k), the higher is the rate of growth needed to generate the necessary amount of investment.

While the Harrod-Domar type of growth model provides some useful insights, we should not forget that it is subject to all the limitations inherent in attempts to explain economic change solely through multiplier-accelerator analysis. In particular, this kind of analysis fails to consider the variety of forces operating on investment other than those implied by the acceleration principle (see the discussion of these other factors earlier in this chapter); it fails to take account of government spending and revenues; it deals with *net* investment and ignores the relation between replacement and depreciation; it takes no specific account of the role of population growth, technological change, or trends in the money supply; and so on.[35] Despite these limitations, this type of growth model has stimulated a good deal of useful theoretical and empirical work on the conditions needed for sustained growth.[36]

AGGREGATIVE ANALYSIS AND THE STUDY OF ECONOMIC CHANGE

In this and the preceding chapters we have sought to develop a kit of analytical tools with which we could study the causes of economic change—the fluctuations we call business cycles and the longer-run

[35] See the discussion of the forces making for secular change in Chapter 8.

[36] For further discussion of growth models, see Chapter 13. We might mention here that other types of growth theory—including those of the English classical economists, as well as Marx and Schumpeter—also exist. For a brief survey, see Benjamin Higgins, *Economic Development: Problems, Principles, and Policies*, 1959, chaps. 3–8.

trends in output, prices, and other economic variables. We must emphasize that what we have so far are only the tools, not the answers that we are looking for.

This is the case because thus far we have been dealing only with large aggregates—total income, spending, investment, consumption, and so on—and with a few very broad, aggregative relationships. But this is not enough to get at the complex set of causes that lead to economic instability and growth. At this point we must desert the simple aggregative type of analysis and proceed to examine the complex of forces that operate on production, prices, costs, profits, and business expectations in various types of industries. The nature of these forces changes, both in the short run and over longer periods, with the result that every new business cycle differs in important respects from its predecessors—and with the result also that the long-run trends in output and prices also change from time to time.

There is probably still too much of a tendency among economists to state the problem of economic fluctuations and growth in terms of a few broad aggregative relationships that tend to remain stable over relatively long periods of time. While theory must abstract from some of the complexities of real life if it is to have something useful to say, we should not be content with the level of abstraction—with its consequent neglect of important dynamic forces—that is involved in the use only of the analytical tools developed in this and the preceding two chapters. In the rest of this book, we shall be concerned with the more detailed statistical, historical, and theoretical treatment that we think is required to achieve an understanding of the dynamic forces that are now operating, or have operated in the past, on the American economy.

PART II

THE NATURE AND CAUSES OF BUSINESS FLUCTUATIONS

CHAPTER 7

MEASURING CHANGES IN BUSINESS ACTIVITY

BEFORE going on to examine in detail the nature of business fluctuations, we need to pause and ask a deceptively simple question: What is business (or economic) activity, and how do we measure changes in it? Any sort of study of business fluctuations requires an answer to this question, but the answer is not as simple as might at first appear. In fact, there is more than one answer.

THE MEANING OF BUSINESS ACTIVITY

Let us first dispose of the distinction between business activity and economic activity. Strictly speaking, a distinction should be made, though it is permissible to use the two terms as synonyms in any country where economic activity is predominantly geared to the profit motive. This is, of course, the case in the United States. Nonbusiness economic activity includes the work of housewives, which is not priced at all in the market and in exchange for which no money passes hands; and it includes also the activities of government, which are not carried on in the hope of profit but which do involve money transfers in the form of taxes and payment by government for the goods and services it buys. There are, of course, many private nonprofit organizations, also.

We have seen that the level of business (i.e., profit-making) activity depends on the level of aggregate demand. This provides a clue as to what types of nonbusiness activities should be included in our concept of "economic activity" for purposes of business-cycle analysis. "Economic activity" should include all activities that affect aggregate demand—essentially, that is, all activities that generate money

incomes. By this criterion, "business" or "economic" activity includes the activities of business firms and of government, but not of housewives.

This is just the beginning, not the end of our problem. We have not yet defined business (or economic) activity in a way that will permit us to measure changes in it. What do we mean when we say that business is good or bad, or that business activity is rising or falling?

RATE OF USE OF RESOURCES

There are two approaches to the problem of finding a definition that will permit us to set up measures of changes in business activity. The simpler is merely to define economic activity as the rate or intensity of use of economic resources. This, however, creates a new problem. Economic resources are of many types—all kinds of labor, the infinite variety of capital goods, and so on. How do we measure the rate at which each kind of resource is used, and how do we combine these incommensurables into a single, homogeneous total? The answer, very briefly, is that we cannot—but two fairly good approximations to a solution can be found.

From any point of view, labor is far and away the most important economic resource. One way to measure changes in business activity, then, is by the volume of employment—particularly if the figures on the number of men working are adjusted for changes in the number of hours worked per week.

The other approximation involves measuring changes in the rate of use of economic resources by changes in the output of goods and services that results from such use. The GNP (or national income) is one measure, in dollar terms, of the output resulting from economic activity. It has the obvious disadvantage that it may reflect changes in prices as well as in real output. We can try to eliminate the effect of price changes by deflating the dollar figures for the GNP by an appropriate price index, though this procedure is never completely satisfactory. Or we can construct an index number of the physical volume of production that will show, relative to some base period, the average change in many different kinds of production (of services as well as goods). Here again there are difficulties—technical difficulties in constructing and interpreting the index and in including enough different lines of production. There is the further problem

—not very important in measuring month-to-month changes but likely to become quite troublesome over longer periods—that changes in productivity may cause measures of output and of employment to give somewhat different results. Between 1953 and 1959, for example, manufacturing output increased considerably, but employment in manufacturing showed a net decline.

Our first general approach, then, gives us employment, the GNP or national income (corrected for price changes), and production as variables we can measure to discover what is happening to business activity. Let us now look at the other approach to a definition and see what measures it suggests.

THE BUSINESS SITUATION

Let us think of the economic system as a highly complex organism, made up of many separate but interconnected parts. To stretch the analogy a bit, we can say that the organism has a blood stream (the flow of money incomes or, perhaps, total money payments) and a nervous system made up of the financial statements of business.

Just as the human body represents a highly complex set of biological processes, so does the economy represent a complex set of commercial, industrial, and financial processes. A well-functioning economic organism yields desirable results—profits for some, wages and employment for others, goods and services for all, and so on. We can then use "business activity" as a generic term for the general health of the economic organism—to describe the sum total of processes that together make the economic system as a whole function.

Viewed from this approach, business activity (or "the business situation") is not directly measurable. It consists of a congeries of processes, all of which are a part of total business activity—all kinds of production, wholesale and retail trade, banking activities, transportation, and so on. Further, each type of process may have more than one characteristic in which we are interested—for example, not only production but sales, new orders, prices, inventories, credit collections, security issues, etc. Business activity is the total of all these (and other) characteristics of all the different functioning parts of the economy.[1]

[1] This definition of business activity is used explicitly or is implied by a number of writers. It is implicit in the work of the National Bureau of Economic Research. Thus Burns and Mitchell, in reporting on the National Bureau's methods,

Although it is obviously impossible to measure directly total business activity defined in this way, we can rely on short cuts. We can construct a sample of a selected number of series and hope that fluctuations in this sample will tell us what we want to know about the "universe" of total business activity from which the sample is drawn. This, essentially, is what we try to do when we seek to measure changes in the price level. We select particular prices to represent the universe of all prices and then construct an index number of the prices chosen. Similarly, from selected business series, we can construct an "index of general business conditions."[2]

In constructing a general business index, we can choose our sample series with one of two ideas in mind. We may want our business index to be primarily descriptive—to show merely what is happening —or we may want it to emphasize symptoms. If our aim is descriptive, we want a broadly representative index, weighted by the importance (according to some criterion) of the different areas of economic activity represented. Such an index may move rather sluggishly, since it reflects the behavior of all types of business. In the symptomatic type of index, we choose our series to emphasize those kinds of activity that we think are particularly significant as indicators of the future course of business. This type of index is sometimes called a "sensitive index" and sometimes a "business barometer" or a forecasting index. Obviously, both types of indices, the broadly descriptive and the symptomatic or sensitive, can be useful in analyzing current business conditions.

Whichever type of index we decide on for a particular purpose, the limitations inherent in all index numbers—and particularly in those purporting to portray changes in business conditions—should be kept in mind. What is the significance of the series chosen, and

---

describe business cycles in terms of alternating expansions and contractions "occurring at about the same time in many economic activities." A. F. Burns and W. C. Mitchell, *Measuring Business Cycles,* 1946, p. 3; cf. also p. 11. The same idea lies behind the National Bureau's work on business-cycle indicators and diffusion indices. (See pp. 515–522, below.) J. A. Schumpeter has a similar approach to the concept of business activity. See his *Business Cycles,* vol. 1, 1939, chap. 1.

[2] As an alternative to constructing a single index, we may select our series and chart them but not combine them into a single average series. The chart enables us to study interrelationships among the series chosen; and these relationships may be more significant than the *average* behavior of all the series together. Of course, the greater the number of series charted together, the less easily can the eye take in the significant relationships.

what aspects of business behavior do they picture? The number of economic series that tell us something about the business situation is very great indeed. While these series are linked together by fundamental underlying relationships, they do not all behave in precisely the same way. Index numbers are averages, and subject to all the limitations of averages. It is well to remember that the picture presented by an index number is never a direct photograph; it is a synthesis of the "true" picture we cannot see and the distortions inevitably created in the process of constructing the index.

Though this second general approach to the problem of defining business activity suggests that we should average a large number of series, it may be argued that, as a short cut, this concept of business activity can also be measured by a single broad series. Thus, bank debits or bank clearings, which measure total payments made by check, have been frequently used as a business index. It might also be argued that the GNP provides a measure of the average change in all the different types of economic activity, each type weighted according to its contribution to the flow of new goods and services. In addition, the GNP and its components, by indicating current changes in aggregate demand, point to forces at work that may cause the various types of business activity to change.

Which definition of business activity we use does not make too much difference for most purposes. But unless we have some reasonable definition in mind, we are not in a very good position to judge the merits of the various business indices that are currently used.

AGGREGATIVE MEASURES OF BUSINESS ACTIVITY

Let us now look at the actual measures that are currently used to record changes in the "business situation." These can be conveniently classed in three groups: simple aggregative measures, index numbers, and significant partial indicators. We shall briefly describe some of the more important series in each group, indicate their uses and limitations, and see which of our definitions of business activity they seem to fit. (A fourth type—measures that are presumed to have particular forecasting value—might be added here, but we shall defer discussion of this sort of measure until we deal with the general problem of business forecasting in Chapter 17.)

As their name suggests, aggregative measures are series in the form of simple totals that describe the behavior of some important eco-

nomic variable. The most important of such measures are the GNP and national income, bank debits, and employment. The last of these is an estimate of the total number of men at work; the others are dollar totals of expenditures and income. At least in conception, they are simple aggregates of like things and not averages of different kinds of activity.

NATIONAL PRODUCT AND INCOME

The series measuring the gross national product and net national income were described at length in Chapter 3. They are dollar totals representing, respectively, total expenditures on new goods and services produced and the sum of all net incomes earned. The GNP is probably the most widely used measure of *total* business activity.

The value of the GNP as an indicator of changes in business activity is obvious. It measures the result, in terms of the value of goods and services produced, of the use of economic resources or of business activity defined as a complex collection of business processes. Its components—consumers' expenditures, gross investment, etc.—throw valuable light on the forces at work tending to change the course of business. These components give us the best information we have on the critically important saving-investment process.

The GNP, however, has limitations as a measure of business activity. Being a value series, expressed in dollars, it reflects price changes as well as fluctuations in physical volumes. It appears only quarterly and therefore cannot keep us as up-to-date as we should like to be. For the years before 1939, it is available only in annual form. It should also be remembered that the GNP and its components are estimates, subject to an indeterminate margin of error created by deficiencies in the data and the methods used to fill in the gaps in the available information. Finally, the GNP and its main components are broad aggregates. They do not give us detailed information on significant sectors of the economy in which we may be interested; nor do they necessarily reveal significant causal forces at work which may be uncovered by other types of data.

Although estimates of the GNP and total national income are not available more often than quarterly, the Department of Commerce does publish monthly estimates of "personal income," i.e., total incomes paid out to individuals. These estimates show separately wage and salary receipts, other labor income, proprietors' and rental in-

come, interest and dividends paid to individuals, and transfer payments. While subject to inevitable errors in estimating, these figures are an important guide to current changes in the purchasing power of individuals. They also provide a rough measure of the rate at which money incomes are currently being generated.

BANK DEBITS

Before estimates of the GNP became available, bank debits were the most widely used aggregative measure of business activity. Bank debits represent payments made by banks against checks drawn by depositors or against withdrawal slips presented over the counter. Hence they are a direct measure of total money payments made by check. Debits are a better and more inclusive measure than bank clearings, which include only checks that pass through local clearing houses. Debits include, in addition, checks cashed over the counter and those in which the payee deposits the check at the bank on which it is drawn. Such checks do not need to be cleared between banks and hence do not show up in clearings.[3]

What sort of measure of business activity are bank debits? The reader will recall that (in Chapter 2) we expressed the equation of exchange in the form:

$$MV = MV_F + MV_D + MV_Y = PT$$

where the subscripts F, D, and Y refer to financial, "duplicative," and income-generating transactions, respectively. Debits measure, roughly, all money payments of whatever kind except those effected with the use of currency.[4] Hence they include considerably more than merely expenditures on new goods and services (MV_Y). In particular, they are swollen by financial transactions, and operations of the New York money market have a strong influence on the total of bank debits for the country as a whole.

For this reason, it is common practice to exclude New York City

[3] For a detailed discussion of the nature and limitations of both types of data, see the valuable study by George Garvy, *Debits and Clearing Statistics and Their Use*, Board of Governors of the Federal Reserve System, rev. ed., 1959. This source describes in considerable detail the precise composition of both debits and clearings. For historical studies going back before World War I, only clearings data are available. A continuous series on debits is available only for the years since 1919.

[4] Not all banks report debits; and some types of debits are excluded from the current series—for example, debits to time deposits and to United States Government accounts. Data are currently available for 344 "reporting centers."

and to take the debits figures for the rest of the country only. In this form, debits are a somewhat more reliable measure of changes in production and trade than when New York is included. Even so, "outside" bank debits still reflect to some degree changes in financial transactions, and they include also "duplicative" transactions—the buying and selling of goods more than once as they move toward the final consumer. Hence, "outside" debits should be considered as measuring business activity according to our second rather than our first definition: All transactions are included in proportion to the amount of money changing hands without regard to the amount of "real" resources used or the amount of goods or services produced or incomes generated.

In recent years, outside debits have generally been considered to be inferior to the GNP as a measure of business activity—partly because of the effect of financial and duplicative transactions and partly because the GNP and its components tell us more about the economic forces shaping the course of business activity. (Like the GNP, debits also have the weakness that they show the effect of changes in prices as well as in physical volumes.)[5] As far as the broader swings in business activity are concerned, outside debits and the GNP move similarly. The main difference seems to be that the debits series ordinarily shows wider relative fluctuations. With respect to longer-run movement, the volume of financial transactions did not expand after 1933 to the same degree as did the output of goods and services. Hence, after the Great Depression, the GNP recovered to the 1929 level considerably before the debits series. In the period since World War II, however, there has continued to be a close relationship between bank debits (outside the chief financial centers) and the gross national product.[6]

Despite their limitations, the data on outside debits have two great advantages. They are available monthly and are published promptly. Secondly, they are reported separately for a large number of different cities in every section of the country and can therefore be used as regional business indices. Debits for particular cities or re-

[5] Debits can be deflated by a price index (usually the wholesale price index of the Bureau of Labor Statistics), but this procedure is never fully satisfactory. Another minor weakness of debits (and clearings) arises out of the existence of credit transactions, which get into the debit figures only when payment is made and not necessarily when the goods and services are produced or sold.

[6] Cf. Garvy, *op. cit.*, p. 75.

gions are employed by business firms, local chambers of commerce, marketing research agencies, and so on, as measures of changing business activity in the districts covered by the figures.[7]

EMPLOYMENT

Data on employment provide our third important type of aggregative measure of business activity, in addition to being extremely valuable for other purposes. Information on current changes in the volume of employment is available from three main sources. Data from all three sources are published regularly in the *Monthly Report on the Labor Force,* a publication of the United States Department of Labor, and are widely reproduced.[8]

The first of these three bodies of employment statistics provide official monthly data on the total labor force of the United States and on the total number employed and unemployed. Currently, the *Monthly Report on the Labor Force* provides monthly figures on the total labor force, total employment (subdivided into agricultural and nonagricultural employment), and the total number unemployed. The estimates are based on direct interviewing of a scientifically selected sample of the total population, in the manner of the Gallup and other public-opinion polls. The monthly figures are reproduced in, among other publications, the *Monthly Labor Review,* the *Survey of Current Business,* and the *Federal Reserve Bulletin.*

These labor force and employment estimates cannot be classified in detail by type of industry and are thus limited in their usefulness for many purposes. Hence, they need to be supplemented by the series on nonagricultural employment compiled by the Bureau of Labor Statistics. This series is available monthly and has been carried back to 1929.[9] Monthly estimates are made for total nonagricultural employment and for a number of broad industrial groups (manu-

[7] For a discussion of the use of bank debits to analyze regional business conditions, see Garvy, *op. cit.,* pp. 114–116.

[8] Each issue of the *Monthly Report on the Labor Force* contains a brief description of all three of these sources of data. For more detailed description and evaluation, see Universities-National Bureau Committee, *The Measurement and Behavior of Unemployment,* 1957, and P. M. Hauser and W. R. Leonard, eds., *Government Statistics for Business Use,* 2nd ed., 1956, chap. 13.

[9] The BLS estimates of nonagricultural employment are not directly comparable with those in the labor force reports. The former cover only employees and exclude proprietors, the self-employed, and domestic servants. The figures are compiled from reports submitted by a large sample of business firms, in contrast to the procedure of interviewing the workers themselves.

facturing, mining, construction, transportation and public utilities, trade, finance, service, and government). In addition, the data on manufacturing employment are subdivided on the basis of a detailed industrial classification so that separate estimates are available for a long list of different manufacturing industries. The BLS figures for total nonagricultural employment and for employment in the main industrial groups, corrected for seasonal variation, are shown in Figure 19.

A third body of data on employment and unemployment is available for workers covered by unemployment insurance. These figures "furnish a complete count of insured unemployment among the two-thirds of the Nation's labor force covered by unemployment insurance programs."[10] While only insured workers are covered, this source has the advantage of providing a detailed geographical breakdown, both by states and by local labor market areas. Weekly reports, as well as monthly figures, are available.

These official employment estimates provide us with valuable business indicators. They measure the use of our most important economic resource, human labor. In doing this, they yield an index of how effective the economy is in providing jobs for all who want to work. As broad measures of business activity, however, employment figures have some rather obvious limitations. They do not show changes in the average number of hours worked per week and therefore tend to underestimate fluctuations in output and in total man-hours worked.[11] Also, because of technological changes and other factors affecting the ratio of labor to output, even man-hours worked do not reflect perfectly changes in total production.

Along with its estimates of employment, the Bureau of Labor Statistics also compiles figures on payrolls for production workers in manufacturing and for employees in certain other industrial groups. Payroll data form the basis of the Department of Commerce estimates of monthly incomes going to labor, but they are ordinarily less useful than the employment figures as measures of economic activity, particularly because they reflect changes in hourly earnings as well as employment.

[10] *Monthly Report on the Labor Force,* August, 1959, p. 14.

[11] Separate estimates of average hours worked per week are published by the Bureau of Labor Statistics and can be found in the *Monthly Labor Review* and *Survey of Current Business,* as well as in the *Monthly Report on the Labor Force.*

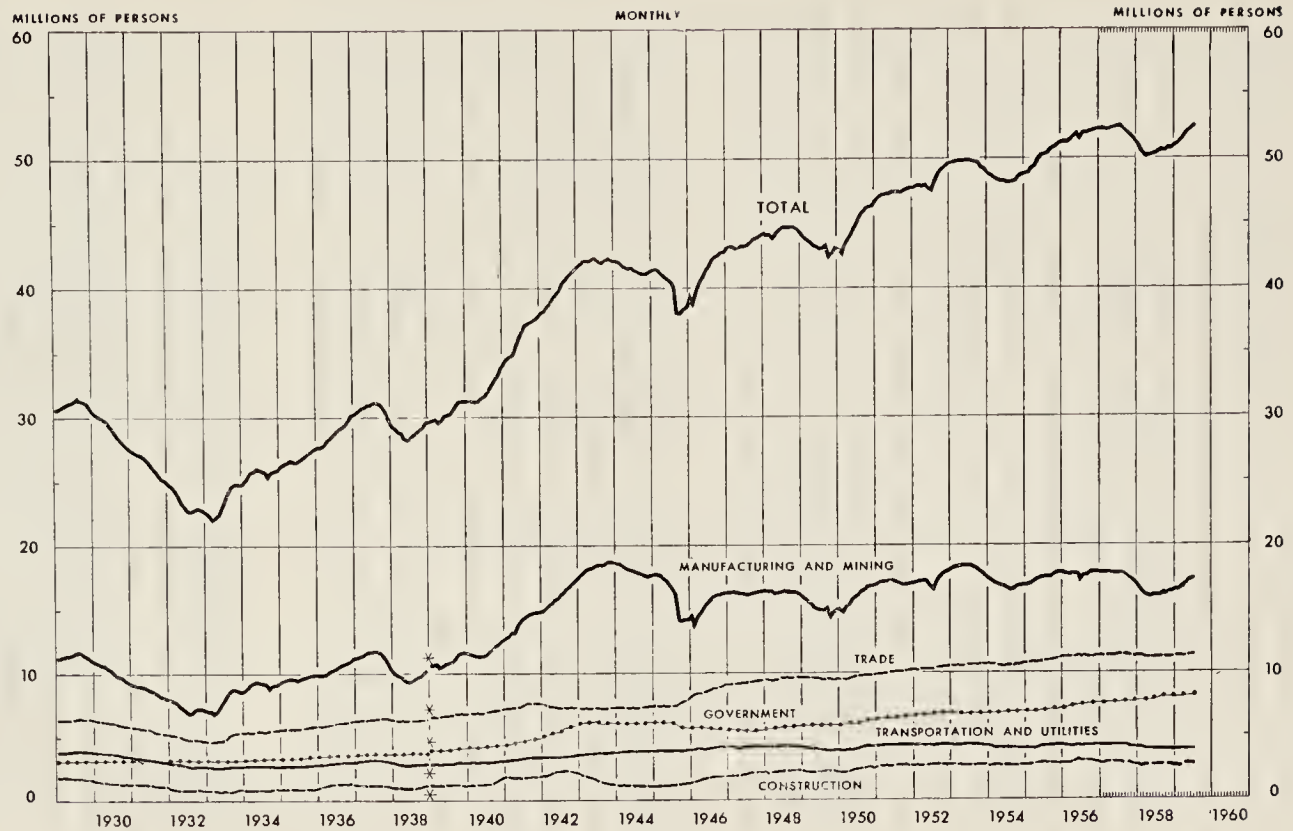


FIGURE 19. Nonagricultural Employment, 1929–1959.

From *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics.* Based on Bureau of Labor Statistics data.

OTHER AGGREGATIVE MEASURES

In addition to the broad measures thus far discussed, a few aggregative series representing a particular type of economic activity have been used as indirect measures of general business. Two of the most widely used are electric power production and railroad carloadings, both of which are available weekly. Since virtually all industry uses electric power, this series does, in a sense, measure changes in industrial activity over a wide area. It has obvious limitations, however, since the series automatically weights various types of activity according to their consumption of electric power rather than on the basis of output, employment, incomes generated, or some other more satisfactory criterion. Data on electric power tell us little or nothing about changes in retail and wholesale trade, construction, finance, and other areas where consumption of electricity is not geared closely to the level of activity. The series has also had a marked upward trend in the past, reflecting not merely the growth of total output but also the spreading use of electric power.

Railroad freight carloadings are a less satisfactory measure of business activity than they were before the advent of the motor truck. In addition, the series for total carloadings is strongly influenced by the movement of coal, although separate series are available for the main types of freight traffic, including "miscellaneous" carloadings (which are closely geared to the volume of manufacturing) and less-than-carlot (l.c.l.) shipments, which represent primarily the movement of goods into wholesale and retail trade. In addition to being used alone, data on carloadings and electric power are frequently included as components of weekly and monthly index numbers of business activity.

INDEX NUMBERS OF BUSINESS CONDITIONS

Monthly index numbers of general business activity, in which a considerable number of series are averaged, are probably somewhat less popular today than they were 20 to 30 years ago. In part, this is because business analysts have come to recognize the value of the data on the GNP and its components and on employment as measures of business activity; in part, there has been a growing awareness of the limitations of index numbers of this type. There is still need, however, for measures of particular types of activity and for fairly

broad measures that appear more frequently or move more sensitively than do the series on GNP and nonagricultural employment.

THE FEDERAL RESERVE INDEX

One index number has met this need so well that it has come to be the most widely quoted business index now being published. This is the Federal Reserve Index of Industrial Production.[12] Strictly speaking, this measure is not an index of *general* business activity at all. Its title describes it precisely. It measures changes only in manufacturing and mining production and in the output of the gas and electric (utilities) industry.[13] Retail and wholesale trade, all forms of transportation, construction activity, finance, foreign trade, and the great variety of service trades, not to mention agriculture, are all specifically excluded. Limited coverage is a characteristic of all business measures of the index-number type. Some indices attempt to be more inclusive than the Federal Reserve index, but no other measure covers equally well the important fields of manufacturing, mining, and utilities.

In its present form, the Federal Reserve index is a weighted average of more than 200 different series, each of which represents the output of a particular product or industry or the man-hours worked in that industry. After adjustment for the number of working days in the month, relatives (i.e., percentages) are computed for each series on the basis of 1957 as 100, and these relatives are weighted in proportion to the value added in production by each industry in the year 1957.[14] The total index and all of its components are available with and without adjustment for seasonal variation. The detailed fluctuations in the Federal Reserve index are traced out in Figures 39 and 40 (pages 400 and 452).[15]

Table 14 tells us something about the composition of the Federal Reserve index. The upper part of the table breaks the index down by major industrial groupings, a classification that brings out how

[12] Published monthly by the Board of Governors of the Federal Reserve System and available currently in the *Federal Reserve Bulletin, Survey of Current Business, Economic Indicators,* and numerous other publications.

[13] Gas and electric output was added to the index in 1959.

[14] The 1957 weights are used only for the period beginning in January, 1953. Since the 1959 revision, the index has been shown both on a 1957 and on a 1947–1949 base.

[15] For a description of the 1959 revision of the index, see *Federal Reserve Bulletin,* December, 1959, pp. 1451–1474.

much manufacturing dominates the total index. Indeed, durable-goods manufacturing alone accounts for roughly half the total weight, a fact that helps to explain why the total index is very sensitive to changes in business conditions.

The classification by type of product in the lower half of the table

TABLE 14. Major Groups in Federal Reserve Index of Industrial Production[a]

Group	Relative Weights[b] (percent)
By type of industry	
Manufactures—total	*86.49*
Durable manufactures	49.66
Nondurable manufactures	36.83
Mining	*8.55*
Utilities	*4.96*
Total index	*100.00*
By type of product	
Final products—total	*46.75*
Consumer goods	*31.13*
Automotive products	3.35
Home goods and apparel	9.60
Consumer staples	18.18
Equipment	*15.62*
Business equipment	12.16
Defense equipment	3.46
Materials	*53.25*
Durable	27.81
Nondurable	25.44
Total index	*100.00*

[a] From *Federal Reserve Bulletin*, December, 1959, p. 1467.
[b] Based on relative importance of the various groups in 1957.

has been available only since 1959. As we can see from this part of the table, a little less than half the index represents final products. The rest are durable or nondurable materials (raw materials, semi-finished goods, building materials, business supplies, business fuel and power, etc.) that are bought by business for further processing. Thus, over half the index represents a "duplicative" or intermediate type of output that does not enter directly into the gross national

product (which is confined to expenditures on final products).[16]

Figure 20 shows the behavior of the major components of the Federal Reserve index during the period 1952–1959. The upper panel of the chart brings out clearly the fact that the output of durable goods is subject to much wider cyclical fluctuations than is the production of nondurables. The lower panel suggests that the production of materials fluctuates more widely than does the output of final products or of consumers' goods.

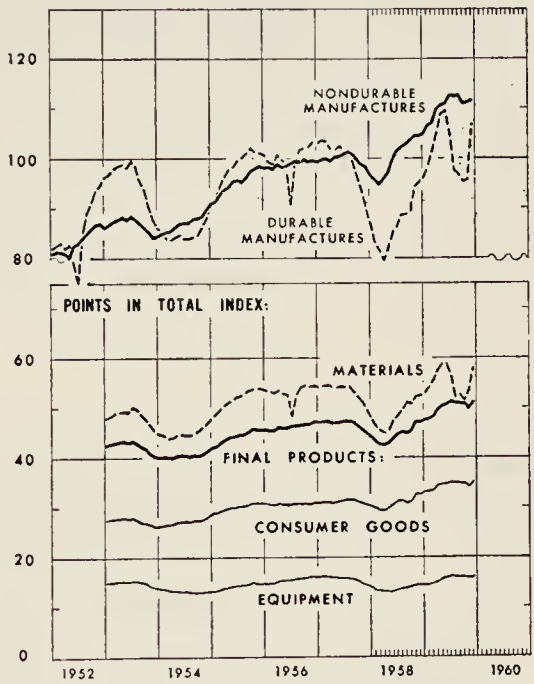


FIGURE 20. Major Components of Federal Reserve Index of Industrial Production, 1952–1959. (1957 = 100)

From *Federal Reserve Chart Book on Financial and Business Statistics,* January, 1960. "Points in total index" refer to the number of points in the total index in any month accounted for by each of the components shown.

Two reasons largely account for the wide use of the Federal Reserve index. First, it is clearly the best index we have of activity in a highly important sector of the economy. It is comprehensive, carefully constructed, and available fairly promptly. It has the limitations inherent in any production index—for example, a downward bias that shows up over long periods, inability fully to take account of new lines of activity and changes in the importance of old lines, failure to reflect changes in quality, limitations inherent in the original data, and so on. But it is much better than any other index of industrial production available.

The second reason for its wide use is its symptomatic value. Manufacturing, mining, and gas and electric power account for about one third of all nonagricultural employment and about the same fraction of the total national income. In large part, the other sectors of

[16] This statement requires one qualification: Any net change in the inventories held of these intermediate products does enter into the GNP.

the economy are engaged in supplying materials to or using the products of manufacturing. Manufacturing and mining are more sensitive to cyclical fluctuations than any other type of business. Thus, industrial production represents a substantial fraction of total output, and it is the part that is most sensitive to changes in aggregate demand. Therefore, though the Federal Reserve index is not an accurate measure of *all* business activity, it does measure changes in a very important and cyclically sensitive part of the total economy.

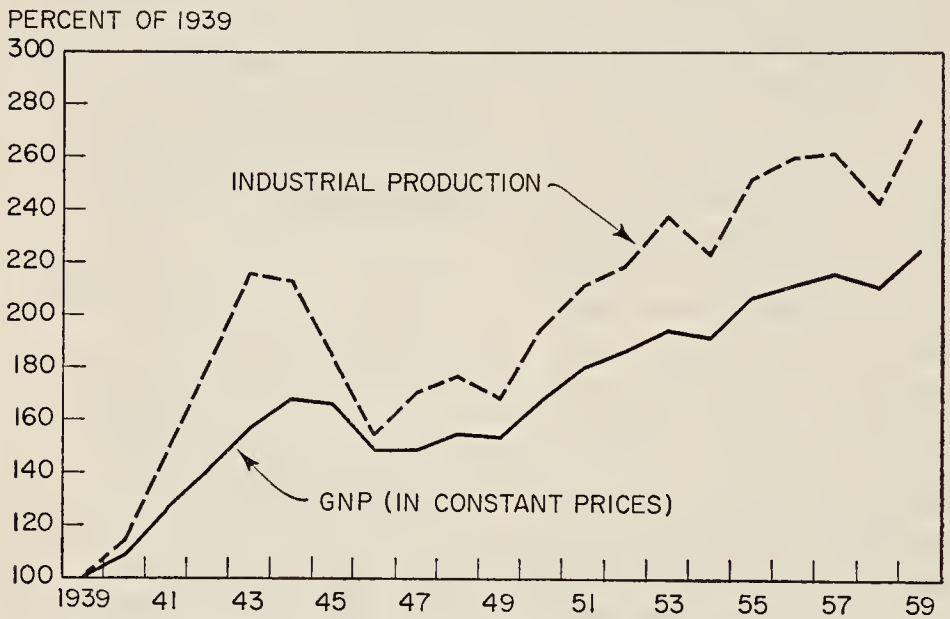


FIGURE 21. Comparison of Industrial Production and GNP in Constant Prices, 1939–1959. (1939 = 100)

Based on Federal Reserve Index of Industrial Production and Department of Commerce data for GNP in constant prices.

Because of its limited coverage, the Federal Reserve index is not a good measure of *total* business activity in either of the two senses discussed earlier in this chapter. It does measure one of the variables (production) we suggested might be used to represent the first of our two concepts of business activity, but it measures only a part of total output. Granted its reliability as a measure of industrial production, its value as a general business index lies in the cyclical importance of the particular area of economic activity that it covers. It should be remembered that changes in the index overstate the magnitude of the cyclical swings in *total* activity as measured, say, by the

GNP corrected for price changes. This can be seen in Figure 21, which compares the Federal Reserve index with a measure of total production of all new goods and services for the period 1939–1959. The latter is merely the GNP expressed in constant dollars, i.e., corrected for price changes. Industrial production rose much more than total output during World War II because of the tremendous expansion in armaments and other manufactured goods for the armed services. The drop in industrial production at the end of the war was much more precipitous than the decline in the deflated GNP series. The greater cyclical sensitivity of industrial production is illustrated by the behavior of the two series after 1945. Notice in particular how sensitive the Federal Reserve index was to each of the mild recessions since World War II. It is also clear that the upward trend in industrial production has been stronger than in the GNP as a whole.

OTHER MONTHLY INDICES

We shall not attempt to review systematically all the other monthly indices which may be found in various business and banking publications. Brief mention of a few, however, will serve to illustrate the types of measures that are available. All of these indices suffer from more or less limited coverage and the need to employ somewhat arbitrary weighting systems. Most of them attempt to measure merely "industrial activity," although this ambiguous concept is frequently interpreted to cover a somewhat broader field than production in manufacturing and mining. Because of their heterogeneous composition, these indices cannot be interpreted in the same clear-cut fashion as can the measures of GNP and employment discussed in the preceding section.

The Cleveland Trust Company publishes a monthly index of American business activity which has been widely reproduced. Although called an index of business activity, this measure in recent years has actually been an index of industrial production, with no wider coverage than that of the Federal Reserve index. For the period since 1901, only data for manufacturing and mining production are used; for the period since 1919, the Federal Reserve index is utilized. Like many other business indices, the Cleveland Trust Company measure is corrected for trend. The index has been carried back to 1790, and a variety of series were used for the decades before 1901.

The Cleveland index is chiefly useful as a measure of business activity for this earlier period.

Although not publicly available, one of the better indices is that prepared by the American Telephone and Telegraph Company for the use of company officials. Its title, "Index of Industrial Activity," indicates its limited coverage. At present, it is based on 25 series, chiefly representing output or man-hours in manufacturing but also including data on electric power. Each of the 25 series is corrected for seasonal variation and weighted according to its "representativeness as a measure of industrial activity." The resulting index, with and without adjustment for trend, has been carried back to 1899.[17] It is charted in Figure 29 on page 255.

Barron's, a weekly financial magazine, publishes a monthly "Index of Production and Trade," which combines the Federal Reserve index for manufacturing and mining with series on freight carloadings, building activity, and electric power production. The index is adjusted for trend (and population growth) ; but it is also published without trend elimination. As can be seen from the series included, the index covers primarily changes in industrial production, although it has a somewhat broader coverage than the Federal Reserve index.[18]

WEEKLY INDICES

Like most of their monthly counterparts, the weekly measures of business activity currently available reflect primarily changes in industrial production rather than in total business activity. The number of reliable series that appear promptly on a weekly basis and at the same time measure changes in activity in significant sectors of the economy is fairly limited. Most of the weekly indices, therefore, rely on some or all of the following series, which do meet these criteria: production data for steel, automobiles, lumber, coal, petroleum, and electric power; railroad carloadings; and construction.

Three weekly indices that reach a fairly wide public are those published in *Business Week,* the Sunday *New York Times,* and *Barron's.*

[17] Information regarding this index has been kindly supplied by the American Telephone and Telegraph Company.

[18] For detailed description of the index, see Edwin Frickey, *Barron's Index of Business Since 1899,* 1943, reprinted from *Barron's* magazine for various dates in 1942–1943.

The *Business Week* index, for example, is based on series for residential and engineering construction, railroad carloadings, and the production of electric power, steel, automobiles and trucks, machinery, transportation equipment, crude oil, and paperboard. The series are corrected for seasonal variation, converted to a 1947–1949 base, weighted, and averaged. No correction is made for trend.[19] The *New York Times* index comprises six series. The *Barron's* weekly index attempts to cover industrial production, electric power production, and carloadings. (Construction is apparently omitted from the weekly index.) Only a few series are used for the industrial production component, but a rough adjustment is made to allow for incomplete coverage. The index is published both with and without correction for trend.

Although weekly business indices of the sort described are perhaps useful for some purposes, they are of doubtful value as measures of total business activity. Their coverage is highly limited; and, since relatively few series are used, fluctuations in one or two series can have an undesirably large effect on the total index. While it is convenient to have a single index to look at, much more can be learned by examining the behavior of the individual series that enter into the index. This is not an overwhelming chore, since only a relatively few series are involved. As a matter of fact, examination of the component series is essential if one is properly to interpret current changes in a weekly business index. If this safeguard is taken, a well-constructed weekly index may help one to anticipate by a few weeks the approximate level of one of the more reliable monthly indices.[20]

SIGNIFICANT PARTIAL INDICATORS

Broad measures of business activity, whether of the aggregative or index number type, have the defect of concealing more than they reveal. By suppressing details, they yield a single figure to represent the amount of change in the level of business or economic activity (however this concept is implicity defined by the measure being used). This information is valuable as a first step. But to understand what

[19] The *Business Week* index is described in the January 16, 1954, issue of the magazine.

[20] For a valuable compilation of national and regional business indices currently available, see Arthur H. Cole, *Measures of Business Change*, 1952. Cf. also the compilation in R. M. Snyder, *Measuring Business Changes*, 1955, pp. 198–199.

is happening, and particularly to appraise what is likely to happen next, we obviously need to know much more. If our general business index is rising, which lines of activity are expanding more rapidly than others? How are various types of expenditure behaving? Is there reason to believe that the rise is speculative and likely soon to terminate? To what extent is the expansion being financed by new bank credit? What is happening to prices, profit margins, wage rates, etc.?

It is too early for us to try to decide which statistical series are likely to have the greatest symptomatic value for the student of economic fluctuations. Suggestions on this score will emerge automatically out of the analysis of later chapters. It will suffice at this point to indicate the general types of information that are available and how varied are the facets of what we call the "business situation."

One useful source of data on the current economic situation is *Economic Indicators,* prepared for the Joint Economic Committee of Congress by the President's Council of Economic Advisers.[21] This monthly publication brings together a large collection of tables and charts under the following headings:

1. *Total output, income, and spending:* various series for the GNP and its components (both spending and income).
2. *Employment, unemployment, and wages.*
3. *Production and business activity,* including industrial production, weekly indicators of production, construction, sales and inventories, and foreign trade.
4. *Prices,* including both consumer and wholesale prices and also prices received and paid by farmers.
5. *Currency, credit, and security markets,* covering currency and deposits, bank loans and investments, the volume of consumer credit, bond yields and interest rates, and stock prices.
6. *Federal finance,* showing the receipts, expenditures, and deficit or surplus of the federal government.

These are by no means the only indicators that the economic analyst might want to examine on particular occasions. One of the

[21] Published monthly by the United States Government Printing Office. See also the latest *Historical and Descriptive Supplement to Economic Indicators,* which describes the series included and gives historical tables carrying the data back a number of years.

most useful magazines for the student of business fluctuations is the *Survey of Current Business,* published monthly by the United States Department of Commerce. In addition to a review of the current business situation and articles on special topics, each issue contains some 40 pages of statistics giving monthly or quarterly figures for the past year for hundreds of different series. A review of the kinds of information in this statistical section provides a good introduction to the kinds of partial indicators that can be used to supplement the information provided by the broader measures of business activity.[22]

MEASURING CHANGES IN THE PRICE LEVEL

Students of economic instability must be concerned with changes in prices for two reasons. First of all, and quite obviously, price changes and price relationships are important aspects of the general business situation. Changes in the general level of prices and in the relationships among groups of prices affect business and consumer expectations, have an influence on costs and profits, help to shape the environment of wage negotiations, and otherwise influence the level of spending and output.

But there is also a second reason for stressing the importance of price changes. We are interested in the general price level for its own sake, because the price level measures the purchasing power of the dollars that we have to spend. The higher the level of prices, the lower is the purchasing power of the dollar, and the less will a given money income buy in terms of goods and services. The price level in the United States and most other countries has been rising almost without interruption since the middle 1930s; in terms of consumers' goods a dollar in 1960 was worth less than half as much as it was in 1939. As a result, changes in the official index numbers of prices are closely followed in business, labor, and government circles. The behavior of the cost-of-living is front-page news in the daily press and the subject for frequent editorial comment.

[22] Another useful and equally inexpensive source of current statistical information is the *Federal Reserve Bulletin,* which is particularly valuable for data in the field of money and banking but which also includes a convenient selection of other statistics. The Board of Governors of the Federal Reserve System also issues monthly a useful compilation entitled *Federal Reserve Chart Book on Financial and Business Statistics.* Also useful are the "Historical Supplement" issues of the Federal Reserve Chart Book, which provide charts going back to the 1920's.

The two best known American price indexes are both published by the United States Bureau of Labor Statistics.[23] One is the Wholesale Price Index; the other is the widely cited Consumer Price Index. The first, as its name suggests, measures prices at wholesale—or, more accurately, at the level of the first commercial transaction for each commodity. The Consumer Price Index (frequently referred to as the CPI) measures changes in the prices of goods and services purchased by typical urban wage-earner and clerical-worker families. It is probably the most widely used price index available, in part because of "escalation clauses" in wage contracts affecting several million workers.

The Consumer Price Index measures the total cost of a "market basket" of about 300 goods and services.[24] The kinds and quantities of goods and services going into this market basket are based on the amounts purchased by a representative sample of city wage-earner and clerical-worker families in a past year. Prices for the items included are collected in a large number of cities—monthly in the largest cities, less often in the smaller. (Prices for food and a few other items are obtained monthly in all cities.) The prices used are retail prices charged in the kinds of establishments in which wage-earner families would normally buy the kinds of items included in the index. Technically, the index is of the weighted aggregate type with fixed weights. An average of the years 1947–1949 is used for the base period; the weights at the time this was written were for the year 1952.

The BLS Wholesale Price Index refers to prices paid by businessmen rather than by consumers—that is, to prices on organized commodity exchanges and to those charged by producers in their usual channels of distribution. Nearly 2000 commodities, all precisely specified, are included. This index is also basically of the weighted aggregate type, with fixed weights (for the year 1954) and, like the

[23] For more detailed description of these series, see the following sources, among others: U.S. Bureau of Labor Statistics, *Techniques of Preparing Major BLS Statistical Series,* Bulletin No. 1168, 1954, chaps. 9–10; H. E. Riley, "The Price Indexes of the Bureau of Labor Statistics," in Joint Economic Committee, *The Relationship of Prices to Economic Stability and Growth* (85th Congress, 2nd. sess., 1958), pp. 107–116; Snyder, *op. cit.,* pp. 91–131. For more recent information, see BLS Bulletins No. 1256 (on consumer prices) and No. 1257 (on wholesale prices), both published in 1959.

[24] The description given here of the consumer and wholesale price indices applies to the situation as it existed early in 1960.

Consumer Price Index, with an average of the years 1947–1949 as the base period.

As we should expect, the Wholesale Price Index is somewhat more sensitive cyclically than the Consumer Price Index. This sensitiveness has been somewhat reduced over the years as the index has been revised to include more commodities and to give less

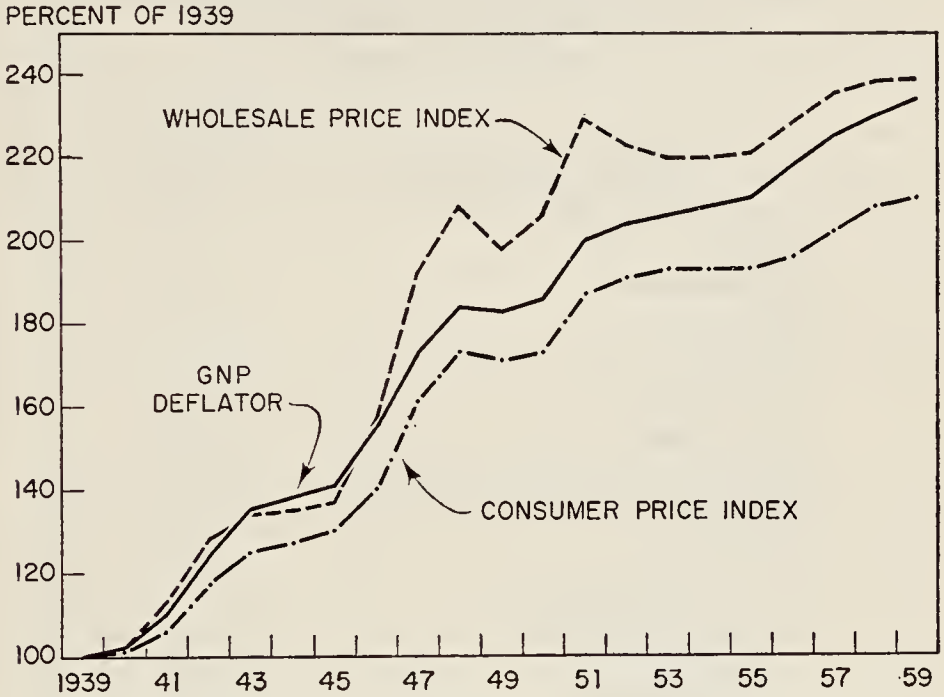


FIGURE 22. Wholesale and Consumer Price Indices and Implicit GNP Deflator, 1939–1959.

Data are from *Economic Report of the President*, January, 1960. All three indices have been converted to a 1939 base.

weight to relatively volatile raw-material and agricultural prices. The broad movements of the two indices are similar, however, as Figure 22 suggests. The wholesale price index fell further in the depression of the 1930's and rose more in the subsequent recovery. Since 1939, both indexes have more than doubled, the net increase being somewhat greater in the case of wholesale prices. During the decade of the 1950's alone, the net increase was somewhat greater in consumer than in wholesale prices. (See Table 15 and Figure 22.)

Both index numbers are broken down into a variety of group in-

dices, and a study of the behavior of these group index numbers can frequently be very enlightening. Farm and food prices, for example, follow a pattern of their own, and it is often necessary to study these prices separately. For example, the wholesale price index showed very little change between 1952 and 1955. This was the joint product of falling agricultural prices and rising industrial

TABLE 15. Selected Components of Consumer and Wholesale Price Indexes and of GNP Deflator, 1959 [a] (1947–1949 = 100)

Price Index and Component	Value of Index in 1959
Consumer Price Index	
All items	124.5
All commodities	116.5
All services	145.6
Wholesale Price Index	
All items	119.5
Farm products	89.1
Processed foods	107.0
All items except farm products and foods	128.2
GNP Deflator	
Entire GNP	129.9
New construction	141.6
Producers' durables	147.8
Consumers expenditures	123.6
Durable goods	116.9
Nondurable goods	116.4
Services	138.0

[a] From *Economic Report of the President*, January, 1960. GNP deflators have been converted to a 1947–1949 base to be comparable with the wholesale and consumer price indices.

prices. When the fall in farm prices ceased, the overall wholesale price index rose significantly to reflect the continued rise in industrial prices. Similarly, it is important to look at the major components of the Consumer Price Index if we are to understand the complex of forces operating on consumer prices as a whole. Thus, the most marked increases in consumer prices during the decade of the 1950's were in services (housing, transportation, medical care, etc.) rather than in commodities. Examples of some of these dis-

parate movements in both consumer and wholesale prices, for the period 1947–1958, are shown in Table 15.

Table 15 lists changes in not only the consumer and wholesale price indices but also in the "GNP deflator." This last is the most comprehensive price index available since it is, in effect, a weighted index of the prices entering into all the expenditures on new goods and services that make up the gross national product. The GNP deflator is a by-product resulting from the attempt of the Department of Commerce to express its GNP estimates in terms of constant dollars—i.e., to deflate for price changes. This comprehensive price index is available only quarterly and is therefore of limited usefulness for analysis of current price changes. As Figure 22 suggests, the total GNP deflator and the Consumer Price Index move in broadly similar fashion, which is not surprising since consumers' expenditures constitute some two thirds of the total gross national product. The chart also suggests, however, that, over the whole 20-year period, the GNP deflator rose somewhat more than the CPI. This is because the increase in the prices of capital goods and government services was greater than that in consumer goods prices.

Numerous other price indices are available besides the ones that we have described. Businessmen to whom short-run price fluctuations are important are likely to make use of one of the indices of "sensitive" prices available—for example, the BLS daily Index of Spot Market Prices, the Dow-Jones Indexes of Commodity Futures Prices and of Spot Prices, and so on. The Department of Agriculture publishes index numbers of prices received and paid by farmers, which are of obvious importance to the agricultural sector of the economy. And there are others. But much the most widely used are the BLS wholesale and consumer price indexes that we have described.

The inherent limitations in even the best price index numbers should always be kept in mind. The sampling problems are always serious. Thus, the CPI represents only a sample of all the varieties of all possible commodities and services that might be included; it is possible to collect only a sample of all the prices at which a given commodity was sold on a particular date; the weights are based on the amounts bought by a sample of families. A particularly serious problem is created by technological change and changes in taste.

Old commodities become obsolete and new ones take their place. What creates an even more difficult problem, qualitative changes are continually being made in the commodities that remain in the price index. It is virtually impossible fully to take account of these quality changes. As a result, we are never able to say with certainty what part of a rise in, say, the Consumer Price Index represents a "true" increase in prices and what part reflects improved quality of the goods included in the index. Because price indices are slow to drop old commodities and substitute new ones and because of the changes in quality that have occurred in many goods, it is fair to say that price indices generally have an upward bias. For these reasons it is believed that the official price indices exaggerate the degree of price inflation that has occurred since World War II.[25]

[25] For further discussion of this range of problems, see, for example, Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959), pp. 106–109; Richard and Nancy Ruggles, in Joint Economic Committee, *The Relationship of Prices to Economic Stability and Growth, op. cit.*, pp. 298–299.

CHAPTER 8

DYNAMIC CHARACTERISTICS OF THE AMERICAN ECONOMY

STATISTICAL series representing the course of economic activity over a considerable period present a dual picture of gradual change, particularly growth, and of short-run instability. A study of economic dynamics has to do with both of these types of movement. To understand the reasons for changes in the level of economic activity, we must know something about the forces inherent in, or acting upon, the economy that make for both long-term growth and short-period oscillations. The factors responsible for the longer-term movements are in part separate from those that create the shorter-run pattern of booms and depressions. To an important degree, however, the two sets of factors are interrelated, and neither can be studied in isolation. Long-term forces operate on an economic structure containing certain elements of short-term instability; the interaction of these two sets of forces creates simultaneously the fluctuations that we call business cycles and the longer-run movements that we observe beneath the short-run changes. "Trends" do not occur in isolation but only as the gradually emerging product of continuous change.

The purpose of the present chapter is to describe the more salient characteristics of the American economy that make change inevitable. We shall first consider the primary factors making for secular changes, particularly the kind of change that we call economic growth, and then turn to those characteristics of modern economic society that lead to the familiar pattern of alternating prosperity and depression. This chapter begins the search for the causes of economic instability that will occupy us through the larger part of this

book. Although confined primarily to the United States, our comments will also be applicable to other countries, the more so as their ways of doing business resemble those in this country.

THE RECORD WITH RESPECT TO GROWTH

Economic stability does not mean stagnation. Though elimination of the extremes of instability should be one of the primary goals of economic policy, we want stability in the form of continuous enlargement of the national output, viewed in both total and per capita terms. In short, we want both economic stability *and* a satisfactory rate of economic growth. Table 16 suggests that, what-

TABLE 16. Growth of Output and Population in the United States, 1839–1959 [a]

Period	Annual Rate of Increase (percent)		
	Output [b]	Population	Output Per Capita [b]
Entire period, 1839–1959	3.66	1.97	1.64
1839–1879	4.31	2.71	1.55
1879–1919	3.72	1.91	1.76
1919–1959	2.97	1.30	1.64

[a] From statement presented by R. W. Goldsmith in *Employment, Growth, and Price Levels*, Hearings before the Joint Economic Committee, Part 2 (86th Congress, 1st sess., 1959), p. 271.

[b] Output is GNP in constant prices.

ever the record with respect to short-run instability, the United States has so far experienced a gratifyingly rapid rise in both total output and production per capita. (See also Figure 1 on page 4.)

Over the past century, the output of the American economy (GNP in constant prices) has grown at an average annual rate in the neighborhood of 3.5 percent per year. This growth has not been constant, however, being higher in the nineteenth century than it has been, for example, since World War I. But the cumulative effect of this growth has been enormous. In 1959, for example, the GNP in constant prices was about 10 times what it had been 70 years before in 1889.

While a substantial fraction of this growth in total output has been absorbed by the expansion in population, output per capita has also risen substantially—at an average annual rate of about

1.6 percent over the last century. (See Table 16.) Thus, at the end of the 1950's, real GNP per capita was about five times what it had been 100 years earlier.[1]

These are impressive figures which help to explain why the American standard of living is so much higher than it was even a few decades ago. Actually, material progress has been even greater than these figures suggest, because this large increase in output has been accompanied by a much reduced work week. The great increase in leisure that has occurred is not included in our measurements of total output. Nor do these measurements fully reflect the variety of new products that have become available, the great advances in public health and medical care, and so on. On the other side, these measurements make no allowance for the price that we have had to pay for some aspects of our rapid growth: water and air pollution, traffic congestion, the need for urban redevelopment, and the other growing pains with which most of us are familiar.

The years since World War II have made us familiar with the contrast between the "advanced" and the "backward" (or underdeveloped) countries. The former are economically "advanced" because for a century or more they all have experienced substantial rates of growth in output per capita and hence have been able to enjoy rising standards of living. This has not been true of the underdeveloped countries. They are "backward" precisely to the degree that they have not been able to generate a significant rate of growth in output per capita. As a result, during the nineteenth and the first half of the twentieth centuries, the gap in well-being between the advanced and backward countries widened significantly. Since World War II there has been a revolution in aspirations in these underprivileged parts of the world—which take in all of Asia except Japan, most of Latin America, Africa, and eastern Europe.[2]

Table 17 provides some estimates of the different rates of growth experienced by a number of countries for various periods in the

[1] For recent evidence on growth rates in the American economy, see the papers presented by R. W. Goldsmith and Solomon Fabricant in *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee (86th Congress, 1st Session, 1959); also Simon Kuznets, ed., *Income and Wealth of the United States: Trends and Structure,* 1952.

[2] There has been since World War II an enormous expansion in the literature on economic development, with particular reference to the underdeveloped parts of the world. For an introduction to this literature, see the text by Benjamin Higgins, *Economic Development: Principles, Problems, and Policies,* 1959.

past. These are all advanced countries, which have enjoyed substantial rates of growth in both total output and output per capita during the last half-century or more. As can be seen from the table, the American rate of growth, whether measured in terms of total output or output per capita, has been matched or exceeded in a few other countries. But rapid growth has been going on longer, and started from a higher level, in the United States than in these other

TABLE 17. Comparative Rates of Growth in Selected Countries for Various Periods [a]

Country	Period Covered [b]	Percent Change Per Decade		
		Population	Output Per Capita	Total Output
United Kingdom	1860–1953	8.0	12.5	21.5
France	1841–1953	1.3	13.8	15.3
Germany	1860–1954	10.1	15.1	27.4
Sweden	1861–1954	6.6	27.6	36.0
Italy	1862–1954	6.9	10.4	18.0
Russia	1870–1954	13.4	15.4	31.0
United States	1869–1954	17.4	20.3	41.2
Canada	1870–1954	18.3	19.3	41.3
Japan	1878–1954	12.7	26.3	42.3

[a] From Simon Kuznets, *Six Lectures on Economic Growth*, The Free Press, 1959, p. 20.

[b] The initial interval in each period generally covers a decade or slightly less, beginning with the first year shown. The ending interval is usually a five-year period terminating in the final date shown.

countries, so that the American output per capita is higher than in any other country in the world.[3] Since World War II, a considerable number of countries have experienced more rapid rates of growth than the United States. The American rate of growth was particularly disappointing in the latter half of the 1950's.

Of particular significance is the rapid expansion in Russian output during the last 30 years. Soviet expansion has concentrated on industrial output—particularly industrial materials, capital goods, and electric power—with the result that Russia is now the second industrial nation of the world.[4] The cold war and military implica-

[3] Compare the more complete data given in the source for Table 17; also United Nations, *Per Capita National Product of Fifty-Five Countries*, 1952–1954, Statistical Papers, Series E, No. 4, 1957.

[4] See, for example, Joint Economic Committee, *Comparisons of the United States and Soviet Economies* (86th Congress, 1st sess., 1959).

tions of this expansion would carry us too far afield, but merely to refer to them is to suggest one of the several reasons that public discussion today places so much importance on the maintenance of a satisfactory rate of growth for the United States. But there are other obvious reasons, also: the general desire for a higher standard or living, the need to provide employment for a steadily expanding labor force, and the opportunity that a rising national income provides to finance increasingly urgent social needs.

THE CAUSES OF SECULAR CHANGE

As we have already seen (pages 76–79), economic growth depends on the balanced expansion of aggregate demand and aggregate supply. It is common, in discussions of long-term growth, to concentrate on aggregate supply—on the rate at which productive capacity expands. But it would be a mistake to ignore the demand side. As we have seen, if aggregate demand expands more rapidly than supply, inflation will be the result. If demand rises too slowly, productive resources will be unutilized and growth of output will to that extent be retarded. Demand and supply also interact. If aggregate demand expands fast enough but not too fast, growth in productive capacity will be stimulated, so that aggregate supply rises faster than it otherwise would.

The most important influences operating to bring about long-run expansion in aggregate demand and aggregate supply are suggested by a consideration of the following simple relationships. (Actually, they are no more than identities.)

The first relationship we can write as follows: Productive capacity or aggregate supply (let us call it Y_s) is equal to population (N) times the proportion of the population in the active labor force ($\frac{L}{N}$, where L is the number of workers in the labor force) times the average number of hours worked per year per worker (H) times the potential output per manhour $\left(\frac{Y_s}{\mathrm{LH}}\right)$. Or,

$$Y_S = N \cdot \frac{L}{N} \cdot H \cdot \frac{Y_S}{LH}.$$

Consideration of this expression immediately suggests the most important factors that contribute to growth in productive capacity.

First, there is the size of the population and, second, the proportion of the population in the active labor force (which changes only very gradually). A third influence is the number of hours worked per year, which has declined markedly in the last century. And finally we come to the most important element of all, output per man-hour or labor productivity. This, in turn, depends primarily on four factors: the amount of capital per worker, the level of technology, the quality of the labor force (education and training,

TABLE 18. Three Estimates of Potential Economic Growth in the United States, 1959–1975 [a]
(All figures are in percent per year)

	Rate of Growth 1909–1958	Projected Potential Growth Rates, 1959–1975		
		High	Medium	Low
1. Total labor force	1.4	1.9	1.7	1.5
2. Annual hours of work	−0.6	−0.4	−0.5	−0.6
3. Total man-hours	0.9	1.6	1.2	0.9
4. Output per man-hour [b]	2.0	3.0	2.8	2.6
5. GNP in constant prices [c]	2.9	4.6	4.0	3.5

[a] Derived from J. W. Knowles, *The Potential Economic Growth in the United States*, Study Paper no. 20, prepared for the Joint Economic Committee *Study of Employment, Growth, and Price Levels* (86th Congress, 2nd Session, 1960), p. 40.

[b] Not shown in original source; derived by subtracting line 3 from line 5.

[c] Potential growth rates are measured from potential rather than actual output in 1959.

health, motivation, etc.), and the skill and drive of those who manage productive enterprises. The first two of these—capital and technology—are most subject to marked change with the passage of time. Different countries vary widely in their endowments of the other two factors, but these do not within any one country change as rapidly as do the first two. (At least the change is not so readily observed.)

Our equation for potential aggregate supply is frequently used in making long-term projections of the future potential growth of output. An example of such a projection for 1959–1975 is given in Table 18. Actually, three estimates or projections are presented, each based on a somewhat different set of assumptions about future trends in the labor force, hours worked, and labor productivity. As a result, the three projections range from a low of 3.5 percent rate of

growth to a high of 4.6 percent. All three projections give a rate of growth higher than that experienced during the preceding 50 years.

Our other basic relationship is simply our old friend, the aggregate demand equation, which we can write as:

$$Y = C + I + G.$$

Consumption will expand with income and population, and will also reflect changes in tastes. The volume of investment is limited by the supply of saving but also depends on the forces determining the marginal efficiency of capital that were discussed in Chapter 6. Among the most important of these from the point of view of long-term growth are technological change, the growth of output itself, and change in population. The expansion in government expenditures depends on a variety of economic, political, sociological, and technological factors. In general, G will rise as income and population expand; it reflects the increasing desire or need for a variety of new public services; and, as we know too well, it is sensitive to every change in the degree of international tension.

If we examine the array of long-run forces operating on the trend of total output, from the side of both demand and supply, we are likely to conclude that the most important factors are population growth, the amount of savings and investment, technological change, and war and the threat of war. In the following pages we shall also pay some attention to changes in the supply of natural resources (which, like technology and the supply of capital, affects labor productivity), the trend in the supply of money, and important changes in the institutional framework within which economic activity is carried on. Changes in the supply of money help to determine the trend in prices, as well as having some influence on secular changes in output.

Some of these factors have their chief effect on aggregate supply; others influence primarily aggregate demand; the most important ones have some effect on both demand and supply. This is best illustrated by the role of saving and investment. The larger the fraction of the national income saved and invested, the more rapidly can productive capacity expand. At the same time, a rising level of investment, through the multiplier, causes aggregate demand to expand correspondingly. The simple Harrod-Domar

growth model discussed in Chapter 6 (pages 150–153) illustrates this dual effect of saving and investment on growth. Similarly, to cite another case, technological change increases productive capacity and, at the same time, by stimulating investment, causes aggregate demand to increase.[5]

POPULATION AND LABOR FORCE

The first of our secular forces is increase in population, including change in the size of the labor force. The great increase in the population of the Western world in the nineteenth century, which also extended into the twentieth, was one of the outstanding characteristics of the rapid economic expansion that began with the Industrial Revolution.

Although the *absolute* rate of growth continued to increase through the 1920's, the *percentage* rate of population growth in the United States declined steadily from the 1850's to the 1930's. Since World War II there has been a new population upsurge, not only in the United States but also in other countries; and this acceleration in population growth has been of great significance in the expansion that has occurred since the war in both aggregate demand and aggregate supply.[6]

Population changes impinge on the course of economic activity in a number of different ways.

1. First, we may mention the effect on the labor supply. In the United States, the total labor force has grown from about 13 million in 1870 to over 70 million in 1960. During the 1950's, the labor force increased on the average by about 800,000 persons per year. During the 1960's, this rate of increase will be substantially greater, reflecting the high birth rates roughly 20 years earlier. One estimate puts the labor force in 1975 at more than 90 million.[7]

The proportion of the total population in the labor force is the

[5] This and the following discussion concentrate on the American economy. For the European case since World War I, see Ingvar Svennilson, *Growth and Stagnation in the European Economy*, 1954.

[6] For a discussion of past and present population growth, see Conrad and Irene Taeuber, *The Changing Population of the United States*, 1958; J. S. Davis, "The Population Upsurge and the American Economy, 1945–80," *Journal of Political Economy*, vol. 61, October, 1953, pp. 369–388; and D. J. Bogue, *The Population of the United States*, 1959.

[7] Cf. Bogue, *op. cit.*, p. 423.

product of a number of factors: the age distribution of the population, the proportion of school-age children who remain in school, the age at which older workers retire, and the extent to which married women seek jobs. Over the last 50 years, the second and third of these have worked to reduce the ratio of labor force to population; the last has operated strongly to increase this ratio. The result of these offsetting factors is that the fraction of the population in the labor force has not shown any marked upward or downward trend over the last half-century or so, although it was somewhat higher in 1960 than in 1900. This also seems to have been true in a number of other countries.[8]

While the number of workers in the American labor force has increased as indicated, the number of hours actually worked has increased less rapidly. This is because Americans have chosen to take part of the rising trend in productivity in the form of more leisure—primarily through a shorter work week but also through longer vacations and more holidays.

2. Population growth has an important effect on the volume of investment and therefore on the level of aggregate demand. The larger the increase in population, the greater the amount of new homes, factories, other productive equipment, and public investment needed to maintain a given per capita production. The correlation between increase in population and investment need not be perfect, and it may exist only with a lag; but the fundamental nature of the relationship is obvious.

3. Population growth has an effect on the total amount of consumers' expenditures and also on the distribution of these expenditures among different types of goods and services. For one thing, the higher the rate of population growth for a given rate of increase in total national income, the less will be the rate of increase in income *per capita;* and, other things being equal, the lower income *per capita,* the higher will be the ratio of *total* consumption to *total* disposable income. Also, acceleration in the rate of population growth due to higher birth rates and earlier marriages, as has occurred since World War II, means that there are relatively more children in the population, which tends to increase the consumption-income ratio for a given level of family incomes. An upsurge

[8] Cf. C. D. Long, *The Labor Force under Changing Income and Employment,* 1958, chap. 12.

in the number of marriages also creates a heavy demand for consumers' durables, and many young couples go heavily into debt to equip their homes. The effect on the *pattern* of expenditures is obvious—increased spending for baby food, children's clothing and toys, household appliances, etc.

One offsetting factor must, however, be mentioned. People save for their old age during their working years and dissave when they retire. For many families, net saving over a lifetime is close to zero. To the extent that this is true, the higher the ratio of those in the working age groups to those who have retired, the higher will be the ratio of saving to income. This means that the higher the rate of population growth (for a given rate of increase in income per capita), the higher will be the saving ratio.[9] But in the first years after a large increase in the rate of growth (if due to an increase in the birth rate), the effect is likely to be in the opposite direction. The accelerated growth will show first in the youngest nonworking age groups, which will tend to increase the consumption-income ratio. Only later, after the higher rate of growth has remained constant for a number of years, will the saving-income ratio rise as the fraction of the population in the working (and saving) age groups increases.

4. Particularly in the United States, geographical shifts have accompanied changes in total population—shifts between regions, from farm to city, and from city to suburbs. Such population movements stimulate investment, especially construction, in the new areas, create a demand for additional means of transportation, lead to new competitive relations between producing regions, and otherwise create changes that gradually are absorbed by the economy.

5. We must not forget to mention, in this connection, the impact of population growth on government expenditures, both federal and particularly state and local. Population growth increases the need for schools, streets and roads, and other public facilities. When population growth is concentrated in metropolitan areas, as it has increasingly come to be, additional needs arise for slum clearance

[9] This will be true after the higher rate of growth has been maintained for a generation. Cf. M. J. Farrell, "The New Theories of the Consumption Function," *Economic Journal*, vol. 69, December, 1959, pp. 685–687; also the reference given there to the work of Modigliani and Brumberg.

and urban redevelopment, control of air and water pollution, improved public transportation, and so on.

6. One final point about population growth needs to be made. The trend in the percentage rate of growth of population has been downward, and even the population upsurge since World War II has left the percentage rate of population increase much lower than it was in the nineteenth century. This means that, even if the percentage rate of growth in output per capita remains constant, the rate of growth in total output will tend to decline because of the retardation in the rate of growth in population. We can see from Table 16 (page 184) that something like this has been happening in the United States over the last century or more. This is a subject to which we shall return in the next chapter.

SAVING AND INVESTMENT

The growth in population in the United States and other Western countries has been accompanied by a steady flow of savings and by a rapid accumulation of real wealth, in the form of both capital goods and durable consumers' goods. The steady increase in the stock of real capital has provided the productive equipment necessary for an expanding labor force, and the continuing supply of saving has provided the means of financing the investment opportunities created by the expanding population, technological change, and the opening up of new markets and new natural resources.

Table 19 gives some idea of the growth of the capital stock in the United States. After deducting depreciation, the total capital stock (in 1929 prices) rose from less than 30 billion dollars in 1869 to more than 400 billions in 1955. This increase was enough to supply each worker in a growing labor force with a steadily increasing amount of capital.

On the average, it requires several dollars of capital formation to create the facilities needed for a dollar's increase in output. This capital-output ratio, which remained relatively constant in the United States in the late nineteenth and early twentieth centuries, is now significantly lower than it was before World War II. (Table 19.) A given amount of growth requires less net investment today than was the case 30 or 40 years ago. This trend can probably be considered one of the characteristics of a relatively mature coun-

try.[10] It also reflects the fact that recent technological change has not, on the whole, required as much capital as did some of the major innovations of a century or more ago—particularly the railroads. Thus, the amount of capital investment required per dollar of

TABLE 19. Growth of Net Capital Stock in the United States, 1869–1955 [a]
(In 1929 prices)

Year	Net Capital Stock [b]		
	Total (billion dollars)	Per Member of the Labor Force (thousand dollars)	Net Capital-Output Ratio [c]
1869	27	2.11	—
1879	42	2.49	3.5
1889	68	3.06	2.9
1899	108	3.79	3.4
1909	165	4.41	3.4
1919	227	5.46	3.6
1929	306	6.33	3.5
1939	319	6.04	3.9
1949	391	6.43	2.5
1955	442	6.74	2.4

[a] From Simon Kuznets, *Capital in the American Economy: Its Formation and Financing*, Princeton University Press for the National Bureau of Economic Research, 1961, chap. 3.

[b] Net after allowance for capital consumption.

[c] Ratio of net capital stock to net national product (Department of Commerce concept). Figures are, in effect, averages for the decade preceding the year shown. The last figure is an average for 1949–1955.

growth in output depends on the stage of a country's development (whether it needs expensive railroads, dams, etc.), on the kinds of goods or services the output of which is to be expanded, and on the state of technology.

Growth, then, requires investment, and investment must be fi-

[10] A relatively mature economy can continue to grow with a lower capital-output ratio for several reasons. (1) It already has much of the "overhead" capital it needs: roads, bridges, river and harbor improvements, railroads, dams, various types of buildings, etc. (2) Through replacement, it can substitute more efficient units of capital for those that wear out or become obsolete, so that some increase in output is possible with little or no increase in the net capital stock. (3) As the rate of population increase declines, further growth in output requires less capital-using forms of investment. (This third point is closely related to but is not identical with the first.)

nanced by saving out of income. Given the capital-output ratio, how fast an economy can grow depends on the proportion of the national income saved. During the 75 to 100 years preceding World War II, the combination of a capital-output ratio of about 3.5 and a net saving percentage of around 12 percent or a bit more permitted the 3.5 percent rate of growth to which we have already referred.[11]

While saving is a necessary condition for growth, the attempt to save more than the economy is prepared to invest will hamper growth. As we saw in Part I, an excess of *ex ante* saving over *ex ante* investment causes aggregate demand to fall. In virtually all countries today, however, the limitation on growth is of the opposite sort; i.e., growth could be faster if there were more saving available for investment.[12] This is particularly true of the underdeveloped countries.

Another feature of the relation between saving and growth calls for comment. A substantial fraction of gross capital formation is financed out of charges made by business firms for depreciation (including obsolescence). The faster the rate of growth in the capital stock, all other things being equal, the smaller will be ratio of depreciation to gross capital formation and GNP.[13] Also, the more durable the capital stock, the smaller will be the percentage of gross investment financed by depreciation. In the United States, the ratio of depreciation to gross capital formation has risen significantly in the last half-century for two reasons: the rate of growth in the capital stock has declined, and the average life or durability of the capital stock has also diminished. This latter tendency stems primarily from the fact that machinery and equipment (i.e., producers' durable goods) have been a rising fraction of total investment, and construction has become a smaller fraction.[14]

As we saw in earlier chapters, what the economy invests it also saves (*ex post*). Hence, saving has grown with capital formation. The ratio of gross saving (including depreciation) to GNP has re-

[11] These are Kuznets' figures. For the 1950's both the saving rate and the capital-output ratio would be lower.

[12] There is some doubt that this statement applies to growth in the United States in the late 1950's. At least, it does not apply without important qualifications.

[13] Cf. E. D. Domar, "Depreciation, Replacement, and Growth," reprinted in his *Essays in the Theory of Economic Growth*, 1957, pp. 154–194.

[14] On all this, see Kuznets, *op. cit.*, chaps. 3–4.

mained roughly constant. As we have seen, however, the ratio of depreciation to gross capital formation has shown a rising trend, and net saving (personal, corporate, and government) as a fraction of net national income is now lower than it was in the half-century preceding 1929.[15] There have been some important changes in recent decades in the composition of saving and in the channels through which savings find an outlet in investment: the increasing importance of "contractual savings" (through life insurance, pension funds, etc.), a somewhat greater tendency for corporations to finance their capital needs from retained gross earnings, the increased importance of government saving (especially if social security funds are included), the growing role of "financial intermediaries" (commercial and savings banks, life insurance companies, savings and loan associations, pension funds, etc.) in channeling savings into investment, and so on.[16]

TECHNOLOGICAL DEVELOPMENTS

Along with the growth and shifting of population and the accumulation of capital, the most important secular force operating on the American economy in the past century and a half has been technological change. The concentration during this period, on both sides of the Atlantic, of innovations involving new sources of power, new productive processes, new methods of transportation, and new products has no parallel in any previous period of history. The array of technological developments has created the modern factory system, brought forth a dazzling array of new products, revolutionized transportation and communication, and in other ways contributed to the vast economic and social changes of the last 200 years.

The introduction of these technological changes, while always going on, has not been smooth and continuous; and one wave of innovations has been succeeded in time by another. We have had not one but a series of industrial revolutions, each overlapping the preceding one, of which the following are particularly to be noted.

[15] *Ibid.*, chap. 3. Cf. also R. W. Goldsmith, *A Study of Saving in the United States*, vol. I, 1955. Net saving here, as throughout our discussion, excludes expenditures on consumers' durables.

[16] For detailed analysis of these and other trends in saving, see Goldsmith, *op. cit.*

1. To the century from 1750 or earlier to the beginning of the Railway Age, and beginning in England, belong the path-breaking innovations in the textile industry and the introduction of steam power for industrial use, which together ushered in the modern era of factory production.[17] Important improvements in iron manufacture belong to this period, which, in the United States, dates roughly from the beginning of the nineteenth century.

2. With the application of steam power to transportation began the Railway Age, which in the United States covers roughly the last 60 years of the nineteenth century. Railway development was the main stimulus to private investment in this period, both directly and also indirectly through its effect in reducing the cost of transportation. It was the railroad, particularly from the 1850's, that filled the great spaces of the West with farms and factories.

3. More or less coterminous with the Railway Age is the Age of Steel. The Bessemer furnace dates from the 1860's, and steel rapidly became the basic raw material of our burgeoning productive system. Steel, in turn, required vast quantities of coal, as did the widening use of steam power. This was also in the United States the period of the triumph of mass production techniques in manufacturing, centering around the production of standardized units and the assembly of interchangeable parts. This development had its beginnings well before the Civil War, in the machine shops, armories, and clock factories of New England and Pennsylvania, but it assumed large-scale proportions in the 50 years following that conflict. An unending array of technological improvements reduced costs in one industry after another, as machinery and modern methods of management and marketing were introduced to capitalize on the mass market created by the growth of population, the railroad, and the steady migration westward.

4. From roughly the turn of the century, electric power and the automobile, and also chemical industries, added their stimulus to the technological forces already at work; so also did the telephone and later the radio. Earlier developments continued, particularly the spread of mass-production techniques and innovations in management and marketing. The "scientific management" movement belongs primarily to this period.

[17] For some stimulating comments, particularly on the early part of this period, see J. A. Schumpeter, *Business Cycles,* 1939, vol. 1, chap. 6, especially pp. 253–254.

5. Beginning more or less with World War II, a new technological revolution has swept over the United States and other industrial nations. The pace of advance in the basic sciences has accelerated with breathless rapidity; and, more than in any previous period, improvements in industrial technology have come to depend on scientific research, both basic and applied. During the last generation we have had the rise of a completely new electronics industry, the development of nuclear energy (and the hydrogen bomb), the rapid expansion of commercial aviation, the commercial application of jet propulsion, the development of plastics and other chemical products and of new synthetic fibers and fabrics, the expansion in light metals, the discovery of a long list of new "wonder drugs," the revolution in managerial methods being brought about by electronic data-processing equipment and control devices, and so on.[18]

Two particular (and related) features of recent technological change have important implications for the growth and stability of the economy. First, technology has stimulated investment in a variety of products and processes over wide sectors of the economy; no one or two innovations play the dominating role that, for example, the railroads did in the nineteenth century. Secondly, innovations are more likely to come out of a research laboratory today than formerly. Technological change is increasingly the product of systematic, large-scale research effort—by industry, government, and the universities. This is reflected in the sharp upward trend in private and public expenditures on research and development. It has been suggested that this may make for a more regular flow of innovations than in the past, with the result that economic growth in the future might possibly be somewhat smoother than during the last century or more.

As we saw earlier in this chapter, the growth of aggregate supply depends both on the supply of labor (labor force and hours worked) and on the trend in labor productivity (output per man-hour). Insofar as it is legitimate to make this distinction, improvement in labor productivity depends more on technological change (including improvements in managerial effectiveness) than it does on increases in the amount of capital used per worker. Since World

[18] See, for example, Alexander King, "Science and the Changing Face of Industry," *Impact of Science on Society*, vol. 7, March, 1956, pp. 3–33; Francis Bello, "The 1960's: A Forecast of the Technology," *Fortune*, January, 1959, pp. 74 ff.

War I, output per man-hour has increased at an annual average rate of about 2.3 percent. Various statistical tests indicate that only a small part of this increase in labor productivity can be explained by the contribution of the increasing amount of capital used per worker.[19] The trends in output per man-hour, per unit of capital, and per unit of labor and capital combined are traced out in Figure 23.

Improvements in methods of agricultural production have paralleled developments in industry. A steadily declining proportion of the population in agriculture has been able to furnish the nation with its food and agricultural raw materials. Technological changes in agriculture have included new and improved implements, the tractor as a greatly improved source of power, gains in plant and livestock breeding and insect control, better soil conservation and use, improvement in market organization, and so on. The improvements in agricultural productivity have been quite spectacular since World War II. Indeed, labor productivity in agriculture has risen significantly more rapidly than in the rest of the economy.

Technological advance operates to create both long-period and short-period changes, in particular industries and in measures of aggregate economic activity.[20] The initial effects of important innovations are usually spurts of expansion, followed by competitive tensions and structural and financial maladjustments that create temporary lulls. Then additional expansion ensues, and further changes follow the earlier ones. Gradually there emerge the secular growth patterns in both particular industries and total activity. The long-run course of total production may even give evidence of gradual wavelike movements around an underlying trend, reflecting the changing nature and intensity of technological change in

[19] Cf. Solomon Fabricant, *Basic Facts on Productivity Change,* National Bureau of Economic Research Occasional Paper 63, 1959, pp. 11, 18–22; also John W. Kendrick, *Productivity Trends: Capital and Labor,* National Bureau of Economic Research Occasional Paper 53, 1956. For a more sophisticated analysis leading to the same conclusion, see R. M. Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics,* vol. 39, August, 1957, pp. 312–320.

[20] The most stimulating work on the effect of technological changes on the contours of economic development has been done by J. A. Schumpeter. See his *Business Cycles,* 2 vols., 1939, and the earlier *The Theory of Economic Development* (English translation, 1934).

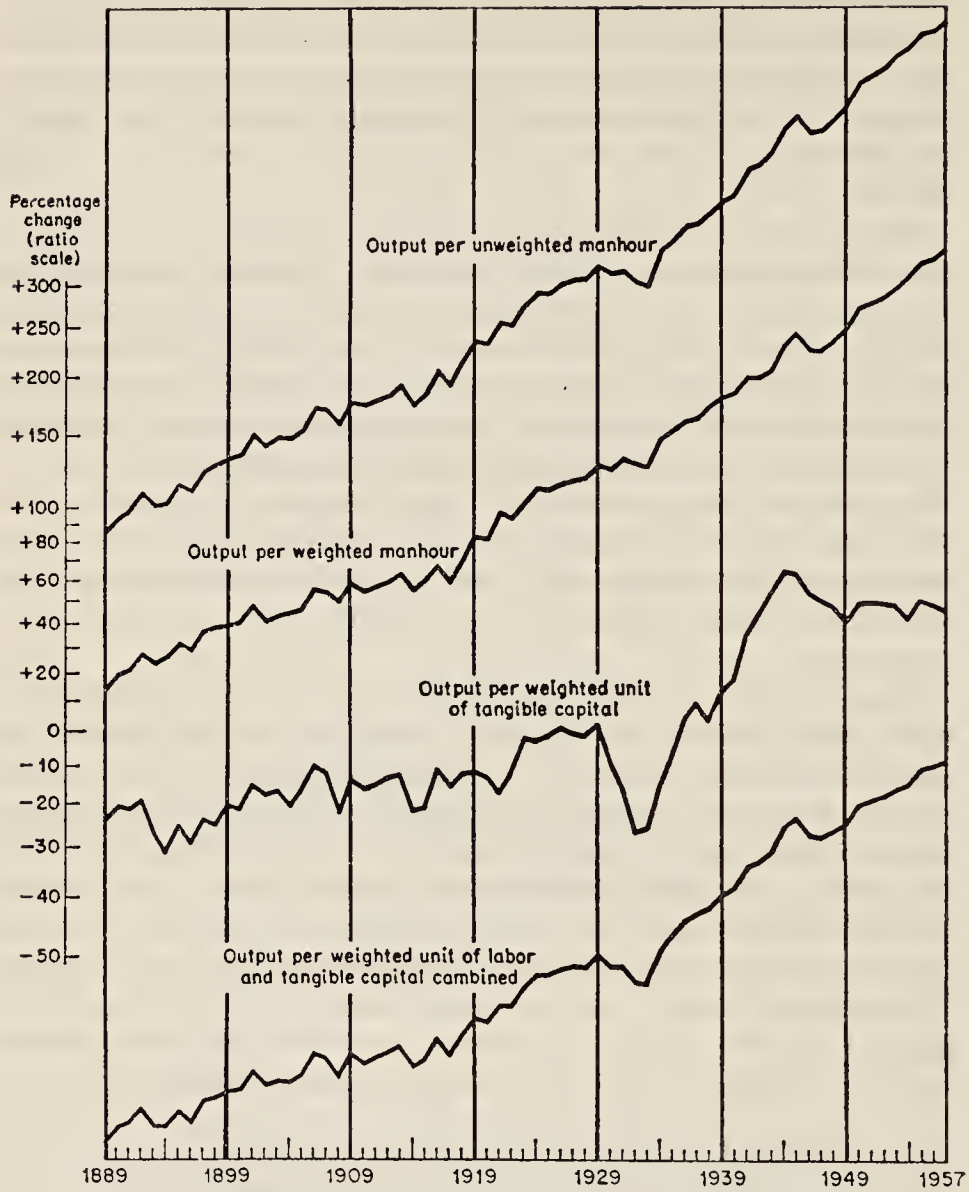


FIGURE 23. Indices of Productivity in the United States Private Economy (Excluding Government), 1889–1957.

Reproduced from Solomon Fabricant, *Basic Facts on Productivity Change,* National Bureau of Economic Research Occasional Paper 63, p. 13.

different periods. We shall examine in a later chapter the evidence bearing on the existence of such "long waves."

The expansion in a particular industry engendered by technological developments and widening markets does not continue unabated forever. Eventual retardation in the rate of growth is a universal characteristic of all expanding industries. We shall have more to say about the shape of growth curves later. In the meantime, we need merely record the fact that industries eventually "mature"; there usually sets in fairly early a decline in the percentage rate of increase in production, and eventually the absolute increments also begin to decline. The consequences of this typical pattern of growth have already been mentioned in connection with the acceleration principle and will come up again in our later analysis. We have also commented on it in connection with the possibility of a declining rate of population growth.

NATURAL RESOURCES

Secular change also results from new discoveries of natural resources, including land itself. The long-run consequences of the discovery of America need no elaboration, and we have already commented on the exploration and settlement of the West. During the last century and a half, man has pried loose from nature the tremendous quantities of raw materials, fuel, and other power resources required to feed the expanding needs of industry: water power, lumber, coal, iron ore, copper, lead and zinc, petroleum, and a variety of metals and minerals of lesser quantitative importance. The availability of natural resources has been a stimulus to industrial expansion in this country and has been related as both cause and effect to that expansion. The gradual exhaustion of some of these resources will also have important secular effects. While the United States has had to rely increasingly on foreign sources for certain raw materials, a shortage of natural resources is not likely to impede growth in the closing decades of the twentieth century. Indeed, some shortages may spur technological change and investment and possibly even stimulate growth.[21]

[21] For a useful recent discussion of this range of problems, see J. L. Fisher and Edward Boorstein, *The Adequacy of Resources for Economic Growth in the United States,* Study Paper No. 13, prepared for the Joint Economic Committee *Study of Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959).

MONEY

Long-period changes in the money supply are another secular force that must be included in our list. This factor operates particularly on the level of prices; but, since it affects total spending, it may affect also production and employment. The influx of precious metals into Europe in the sixteenth and seventeenth centuries can be seen in the trend of prices during that period, and the gold discoveries of the 1840's and again at the close of the century had similar effects. Negatively, the failure of the money supply to rise sufficiently rapidly as population and production expand may generate a downward drift in prices, such as Europe and the United States experienced in the last quarter of the nineteenth century. These secular monetary forces may arise for other reasons, which have in recent decades been more important than changes in the quantity of the precious metals. Forces operating on the volume of bank credit or on the velocity of circulation may have similar effects; and interest rates and other variables may be affected as well as prices.

While secular changes in the money supply have their most important effect on trends in prices and interest rates, they can also influence, directly or indirectly, the rate of growth in output. As we have already seen and as the equation of exchange tells us, changes in the money supply can affect total spending (aggregate demand). From the equation of exchange, we know that $MV = PO$. Thus, if we want output (O) to grow at any given rate while prices remain constant, M must grow at the same rate, modified for any secular trend in velocity. Over the last century, the trend in velocity has been downward. It has been suggested that a 3 percent rate of growth in output with price stability requires a steady increase in the money supply of about 4 percent per annum.[22]

WARS

The effects of wars must be included in our list of secular forces. The short-run consequences of wars are obvious. The immediate impact of a major war on the volume and composition of total out-

[22] Cf. Milton Friedman's testimony in *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee, Part 4 (86th Congress, 1st Session, 1959), p. 628; also E. S. Shaw, "Money Supply and Stable Economic Growth," in American Assembly, *United States Monetary Policy,* 1958, pp. 49–71.

put, on the volume of currency and bank credit, on prices, and on other economic variables is fresh in our memory. These short-run effects are reflected in both wartime expansion and later contraction of the variables mentioned. Some of the readjustment that occurs after wars takes place only gradually, and the economic effects may be felt for long periods afterward. A swollen supply of bank credit may raise prices to very high levels during the war; the first postwar readjustment may leave them much higher than prewar; and there may follow a period of downward drift in prices as an expanding economy grows up to its money supply. This seems to have occurred in the United States after the Civil War and World War I and in various countries after the Napoleonic wars, although there has been no downward trend in prices since World War II. Wars may stimulate technological developments in particular lines, thus affecting growth trends in the postwar period. In the opposite direction, some war-expanded industries may need to go through a long and painful process of contraction which extends over several business cycles. The destruction and economic distortions that result from war may affect the functioning of the economy and the trends of many economic variables for years after the conflict is over.

Wars, then, must be listed as having important effects both cyclically and secularly. In large part, wars influence trends through the forces previously mentioned—that is, through affecting population, technological change, the money supply, and so on. However, they also operate through other channels; and, particularly in view of the importance and tragic frequency of wars in man's history, it seems wise to list them as a separate force, which occurs sporadically but the effects of which may be felt over long periods.

We should include here not only actual war but also the fear of war. The cold war and American policies with respect to national security have obviously affected economic trends in the United States since World War II. The effect has been on both aggregate demand and aggregate supply. The former has been influenced by the level of government spending and taxation. Supply conditions have been affected particularly through the government's role in stimulating technological change in a nuclear, rocket, and electronic age in which national security has come to be in good part in the hands of scientists and engineers.

INSTITUTIONAL CHANGES

There are other causes of secular change that occur sporadically—political changes, for example, or a change in government regulations affecting business (of which tariffs are a good illustration). In addition, economic history records continuous alterations in the institutional framework within which economic activity goes on: the development of large-scale enterprise, the proliferation of financial institutions, the growth of labor organizations, and so on. These impinge on the cyclical process and also affect the course of secular change. They might be called "structural changes," which affect the environment, or "rules of the game" within which cyclical forces work themselves out and partly as a result of which secular movements emerge. Some of these structural changes are themselves, in part at least, the resultant of the basic underlying forces first mentioned—for example, population changes or technological developments.

Most of the forces mentioned in this section—for example, population growth, technological change, alterations in the money supply, wars, and institutional changes—would be classed by many writers as "external forces" or "exogenous" factors, which are not inherent in the operation of a capitalist economy, in contrast to "endogenous" ones, which are. Of those mentioned, population change has at least in part an economic origin; the same is true of saving and investment; and Schumpeter has made innovations the endogenous factor *sui generis* explaining the contour of economic evolution. This distinction is not one that need detain us. It is doubtful that we learn much by emphasizing it, and it has been raised chiefly in connection with the search for causes of cyclical fluctuations that can be said to be inherent in the functioning of a free-enterprise, capital- and credit-using economy.

THE SOURCES OF SHORT-RUN INSTABILITY

In addition to undergoing gradual change in response to the long-run forces enumerated in the preceding pages, measures of business activity trace out a restless path of continuous short-period fluctuations, the more important of which we call business cycles. All economic systems with certain characteristics generate an apparently never-ending succession of upswings and downswings in business,

and these alternations are so marked, and recur so regularly, that we use the word *cycles* in referring to them.

This is not the point at which to offer a precise definition of the business cycle. The essential thing to emphasize here is that cyclical fluctuations in business activity display three salient characteristics that need to be fitted into any definition that is formulated. First, cyclical movements, whether upward or downward, are cumulative. Change in a given direction generates further change in the same direction. Second, they take the form of oscillations, in which the cumulative movement in one direction gradually builds up pressures that eventually lead to a movement in the opposite direction. Finally, cyclical movements are all-pervasive. They infiltrate into all parts of the economy and even spread internationally.

Thus, the economy displays certain characteristic responses to whatever factors begin a movement in a given direction. Even in the absence of external stimuli—such, for example, as wars—modern business economies seem to generate forces capable of initiating cyclical swings. These swings, once they begin, feed on themselves until the piling up of resistances or the occurrence of some new disturbance leads to a movement in the opposite direction, which again becomes cumulative and eventually is in turn reversed.

THE CHARACTERISTICS OF A BUSINESS ECONOMY

For a century, thoughtful men have been asking: Why? The possible answers are complex and not yet fully understood, but it is clear that much of this instability arises out of the basic characteristics of a money-making and money-using economy. There is some evidence that business cycles did not begin until our ways of living came to be dominated by certain of these characteristics, the nature of which we must now investigate.

Part of the problem, perhaps the major part of it, has been succinctly summarized by an outstanding authority on business cycles in the following manner:

"Business cycles do not become a prominent feature of economic experience in any community until a large proportion of its members have begun to live by making and spending money incomes. On the other hand, such cycles seem to appear in all countries when economic activity becomes organized predominantly in this fashion. These observations suggest that there is an organic connection be-

tween that elaborate form of economic organization which we may call 'business economy,' and recurrent cycles of prosperity and depression."[23]

At least three aspects of a money-using business economy tend to make for economic instability.

First, and most important, economic activity is geared to profit-making. "In our analysis of business cycles . . . we must recognize that profit-making is the central process among the congeries that constitute the activities of a business economy."[24] Production takes place, job opportunities are offered, and money incomes are created as a result of the decisions of businessmen—and these decisions depend on the expectations of profit. Businessmen stand at the center of the productive process, and the level of economic activity depends on their evaluation of profit possibilities. Because they act in response to anticipations, human psychology inevitably plays an important role in the course that the business cycle takes. (To say this, however, is not to advance a psychological theory of cyclical fluctuations.)

Second, businessmen operate in a price system marked by extensive division of labor and an infinitely complex network of buying and selling relationships. Several results ensue. Profits depend on the relations between different groups of prices; the sales price to the seller represents cost to the buyer. Then, because all parts of the economy are linked through a network of buying and selling relationships, changes in the profit opportunities in one firm or industry are rapidly communicated to other parts of the economy. Further, since production is for sale in markets more or less distant from the producer, it must be undertaken and costs must be incurred in anticipation of a demand that may or may not be correctly evaluated in advance.

Finally, a modern business economy relies on the use of money, including bank credit. The existence of money means that income receivers are free to use their income as they wish. Since payments to the factors of production are made in money rather than in kind, consumers can allocate their expenditures without regard to the pattern of production. They have the choice of spending or not spending; and, if they spend, they have discretion as to the objects

[23] W. C. Mitchell, *Business Cycles: The Problem and Its Setting*, 1927, p. 182.
[24] *Ibid.*, p. 183.

of their expenditure. At every stage of the circuit flow, the recipients of money payments, businessmen as well as householders, have discretion as to whether and when to spend what they have received. As a result, income velocity, and therefore aggregate demand, will vary in accordance with the spending decisions that are made.

This is not all. Because of the institution of private commercial banking, with its corollary of deposit creation in response to the profit motive, the circuit flow is continuously expanding or contracting as new money is injected or withdrawn. Thus M, as well as V and V_T, may fluctuate. If business firms, spurred by attractive profit opportunities, want to borrow, and if commercial banks can and are willing to lend newly created deposits, business spends more than it receives; and the circular flow expands. The usual definitions of "capitalism" or "free enterprise" do not require that private, profit-making institutions provide the larger part of our means of payment. As a matter of history, however, this is what has happened. Business cycles would occur even if banks lost their power to create and destroy money, but the fact that they have that ability adds another element of instability to a money-making, money-using economy.

Let us combine these three elements of instability—production for profit, a network of price-cost relations stemming from economic specialization, and the use of money—with certain other characteristics that are likely to hold for any economy. Of these, two are particularly important here. First, all production takes time, and for some goods the time involved in production is considerable. The second factor also is characteristic of any type of economy, but it is particularly important in an economic and political system based on individual liberty and private property. All factors of production are to some extent immobile. Productive equipment that can no longer be profitably used in one line of production cannot immediately and without friction be shifted to the production of another commodity—in a different industry and perhaps under different ownership and in another part of the country. If workers can no longer be employed at a profit by one firm, they do not automatically and without time lag find employment elsewhere, particularly if other jobs that may be available entail movement to other areas or the acquisition of new skills.

THE RESULTING TENDENCY TOWARD FLUCTUATION

In an economy with these characteristics, business fluctuations are inevitable.[25] Any significant stimulus or disturbance will set in train a cumulative process of expansion or contraction. An initial improvement in profit expectations will lead some producers to expand production; incomes therefore rise; so does consumers' spending and therefore aggregate demand (the multiplier process); the stimulus is now felt in other industries; and so on. Inevitably, also, the expansion will come to an end. In some lines, demand will be overestimated; not all prices behave similarly, and some producers will be stimulated by widening profit margins that will not continue forever; some markets, particularly those for durable goods, will eventually be saturated for the time being; expansion will presently bring higher costs as the factors of production are more intensively utilized; the banking system, through credit creation, may encourage a rate of business expansion that it cannot support forever. Other possibilities can be added, including disturbances arising outside the normal functioning of the economic system—such as government action, a crop failure, or some unfavorable development abroad. Given the sort of economic system we have postulated, an upward movement will eventually be reversed.

When the reversal begins—centering first in a particular part of the economy—the previous accumulation of potential difficulties, the tendency of businessmen to react strongly when optimistic expectations are disappointed, and the existence of varying degrees of price rigidity and immobility of the productive factors prevent the restoration of a balanced, stable position before a cumulative downward movement has gained momentum. On the downswing, as we shall see when we examine the cumulative process in greater detail in Chapter 11, the contraction first spreads and feeds on itself, and then gradually generates forces that lead, as a minimum, to a stabilization of the level of economic activity. Since, with the sort of economy we have described, businessmen are likely to contract production faster than consumers do their spending, and since

[25] This and the following paragraphs do no more than suggest how a private-enterprise economy may generate fluctuations in business activity. For a more fully developed explanation, the reader is referred to Chapter 11.

various of the maladjustments described will be eventually (although not necessarily completely) corrected, a new upswing will begin in time. Recovery may begin sooner than this implies if new expansionary stimuli are introduced into the economic system—for example, through additional government spending or a large increase in foreign demand.

Whatever the strength of the expansionary forces that automatically emerge from the process of contraction, other forces will eventually operate to begin a new period of expansion. The most important of these stimuli are contained in the secular forces described in the preceding section. To repeat what we said at the beginning of this chapter: Long-term forces operate on an economic structure containing certain elements of short-term instability; the interaction of these two sets of forces creates simultaneously the fluctuations that we call business cycles and the longer-run movements that we observe beneath the short-run changes.

Of the secular forces mentioned, change in population is the one most likely to be absorbed gradually by the economy without setting in motion a set of cumulative cyclical swings. Even in this case, however, cyclical instability may result. Changes in the rate of increase of population can start a cumulative movement in building which will be transmitted to other parts of the economy. Geographical shifts in population might also be sufficient to begin a cyclical movement, as might tendencies toward greater urbanization. Each of these forces, if operating over a sufficient period, might set in motion a series of cycles. The underlying stimulus engenders a cumulative expansion that is overdone; the ensuing downswing corrects sufficient of the maladjustments arising from the upswing for the underlying stimulus to be felt again.[26]

Such external stimuli as wars, various types of government intervention, a secularly expanding money supply, and the discovery of new natural resources are all capable of setting a cumulative process in motion. Some of these forces may be important only sporadically; others may operate on the economy strongly but gradually over con-

[26] Even a constant rate of population growth can affect the course of cyclical fluctuations. If a downswing occurs for other reasons, the continued steady growth of population may be one of the most important forces leading eventually to a new cyclical upswing.

siderable periods. In each case, they are capable of providing forces making for cyclical instability that reinforce those which may be inherent in the typical functioning of the economic system.

THE ROLE OF SAVING AND INVESTMENT

So far we have said little about the phenomena of saving and investment as forces making for instability in a private-enterprise economy. This has been deliberate. The emphasis in recent years on saving and investment and their relations to a few other broad aggregates has tended to obscure the complex variety of interacting forces that give rise to business fluctuations. In this connection, it can be shown that all fluctuations in aggregate demand arise out of changing relations between saving and investment when the latter are defined in *ex ante* terms. Total spending on new output will rise when investment injects into the income stream more than is withdrawn through saving, and spending falls when the reverse is the case. Thus, the cyclical process described in the preceding paragraphs can be translated into terms of changing saving-investment relationships.

Even if consumption normally accounted for virtually all of the national income, so that saving and investment (including government receipts and expenditures) were much less important than they actually are today, a profit-making, money-using economy would still generate cyclical fluctuations. They would undoubtedly, however, be much milder than the cycles we have experienced during the last century and a half. A cumulative expansion could still arise, and the characteristics of a business economy previously described would lead eventually to a cumulative movement in the opposite direction. *Some* investment and saving would have to occur in this process. If, for example, some development led to anticipations of higher prices, businessmen would attempt to accumulate inventories, which is a form of investment. We could say that the rise in production and incomes that would follow was due to an excess of planned investment over saving. If, later, disappointed expectations led businessmen to try to reduce output and liquidate inventories, such action would represent a decline in investment.

In fact, of course, saving and investment make up a significant fraction of the national income—which is another way of saying

that aggregate demand does not depend exclusively on consumers' expenditures. Thus, another element of instability is added to those we have already listed. Investment is the most volatile part of total spending and will inevitably fluctuate quite widely—in part because of those characteristics of a private-enterprise economy that have already been mentioned. Since consumers' spending is relatively passive, being tied fairly closely to the level of income, consumption and saving will not change automatically by the amount necessary to offset the fluctuations in investment. Rather, the reverse occurs. A change in investment leads to a change in income which induces a change in consumption (the multiplier process), so that variations in investment have a multiplied effect on income.

Thus, a high-saving, high-investment economy is potentially very unstable, provided it is also a private-enterprise economy. If income receivers tend to save a large proportion of their incomes, investment must remain at a high level or else aggregate demand will decline. In a private-enterprise economy with large savings, we must rely to an uncomfortable extent on a component of aggregate demand that seems incapable of sustaining itself continuously at just the right level. But, to repeat, this is a problem that arises only in private-enterprise economies. If the saving-investment process is completely under government control, the mere fact that saving and investment are a large part of the national income does not thereby lead to instability.

In Chapter 6 we took a preliminary look at the reasons for the wide fluctuations in the volume of private investment. First of all, investment depends on profit expectations. These expectations are subject to very marked fluctuations, in part because of the exaggerated way in which businessmen react to favorable and unfavorable developments. Apart from the purely psychological element, which in a sense arises out of the fact that investment represents a gamble on what will happen in the more or less distant future, the expected profitability of investment depends primarily on technological change and the rate of expansion of output in different industries. The rate of expansion periodically engendered by new products and new productive methods cannot be sustained forever. This alone is enough to make investment occur in spurts. In addition, a variety of other factors can accelerate or slow down the rate of expansion in particular industries or the whole economy. The

demand for capital goods, over and above replacement needs, depends on continuing *expansion* in the volume of production. Leaving aside technological change, an unchanging level of output does not require any *net* investment at all.

The instability of investment is accentuated by the relative durability of capital goods. As was brought out in our analysis of the acceleration principle in Chapter 6, the more durable the equipment required, other things being equal, the greater must be the capital expenditures necessary to take care of a given increase in output. To accommodate today's increase in demand, the businessman must invest in durable equipment that will also be available to satisfy demand next year and the year after. The fact that capital goods are durable also makes replacement expenditures unstable. Worn-out equipment need not be replaced immediately, if output declines; and frequently, if times are bad, capital goods may be kept in use beyond their normal life. Later, when conditions are better, there will be a surge of investment representing accumulated replacement needs. The same sort of fluctuations occur in the case of consumers' durable goods.

As this discussion suggests, it is investment rather than saving that is the active destabilizing force. Saving would also constitute an active destabilizing element if it showed marked tendencies toward independent fluctuations of its own. We saw in Chapter 5, however, that this is not ordinarily the case. Saving tends to vary more or less passively with income. There is some tendency for the consumption-income ratio to fall during cyclical upswings and to rise during downswings, but this sort of behavior makes for less rather than more instability. In a sense, the difficulty is that (*ex ante*) saving does not fluctuate widely enough; consumption does not automatically change in such a way as to offset the variations in investment.

Since the early 1930's, and particularly since the "revolution" in economic thought stirred up by the late Lord Keynes, the specter of "underemployment equilibrium" has been added to the observed fact of cyclical instability. If a national income large enough to generate a satisfactorily high level of employment regularly creates a large amount of *ex ante* saving, a correspondingly large amount of investment is continuously necessary if that level of employment is to be maintained. If for any reason—say, a decline in the rate of population growth combined with a slowing up of technological progress—businessmen are regularly unable to invest an amount

equal to that which would be saved when the level of income corresponded to "full employment," then the economy is unable to generate the desired level of aggregate demand. Such a permanent deficiency in aggregate demand could be equally well attributed to "oversaving" or to "underinvestment." This is the problem of full employment to which so many economists in Great Britain and the United States have called attention in the last generation. And it is this that is meant when some writers have expressed concern over the possibility of "economic maturity" or "secular stagnation" in the United States.

Since World War II, less has been heard of the problem of secular stagnation. The heavy increase in government expenditures has added a stimulus that may well be permanent, and private investment in the postwar years has shown more sustained buoyancy than many observers anticipated. In both Europe and the United States the danger in recent years has been inflation, not the reverse. In addition, if private investment should eventually show signs of not keeping up with the rate at which the economy wants to save, a variety of things could be done to prevent chronic unemployment. The propensity to consume could be raised; private investment could be stimulated in a number of ways; and there are a wide array of highly useful and socially desirable projects toward which public investment could be directed.[27]

While the phrase *secular stagnation* has not been widely used in recent years, essentially the same issue in another dress was being debated in the United States as the decade of the 1960's opened. Questions were being raised as to whether the rate of growth of output in the late 1950's was as rapid as it should have been, and some concern was being expressed over the fact that, despite inflation and apparent prosperity, average annual unemployment had not fallen below 4 percent during the years 1954–1960.[28]

THE FINANCIAL SUPERSTRUCTURE

Mention must be made of one additional characteristic of a modern business economy, which is related to the saving-investment process and also to the system of credit creation to which we have

[27] This subject is pursued further in Chapters 18–21, which deal with questions of policy.

[28] See, for example, the voluminous hearings and staff reports published by the Joint Economic Committee under the general title of *Employment, Growth, and Price Levels* (86th Congress, 1st and 2nd sessions, 1959–1960).

already referred. This additional destabilizing force arises from the relative importance of financial transactions in total business activity. On a base of real wealth and the production of real goods and services, modern economic systems have erected a considerable superstructure of paper wealth and of financial institutions to trade in such wealth. Thus we have stocks, bonds, and mortgages, as well as a wide variety of short-term credit instruments. We have also stock exchanges, investment as well as commercial banks, brokerage houses, insurance companies, mortgage companies, bill brokers, and so on. We have corporations, holding companies, and investment trusts. We have, in short, all the paraphernalia of what some writers have referred to as "finance capitalism." One indication of the growing importance of financial activities is the fact that the percentage of the labor force accounted for by the financial industries has increased very greatly over the last century.

This financial apparatus has enormously facilitated the flow of savings into the channels required by the investment needs of an expanding economy. But it has also added to the instability that goes with such expansion. In practice, the prices of securities not merely reflect changes in the profits of business firms but magnify those changes. Rising security prices expand the collateral on which new bank loans can be secured. Gains and losses from security trading affect consumers' expenditures and the ability and willingness of businessmen to undertake investment expenditures. Such institutions as the investment trust and holding company—and indeed the corporation itself, with its negotiable securities, particularly common stock—increase the probability that during booms investment will expand too rapidly to be maintained for long. Some weak firms and ill-advised projects are likely to secure financial encouragement they do not deserve. During the period of contraction, business failures and declining profits are reflected in exaggerated form in security prices; failures of financial institutions are added to commercial failures; and commercial banks that have loaned heavily on paper collateral become embarrassed, with the result that a growing unwillingness to lend even for legitimate business needs becomes accentuated.[29]

[29] For a trenchant discussion of the destabilizing role played by financial activity, one can still turn with profit to Thorstein Veblen. See, for example, *The Theory of Business Enterprise,* 1904.

The destabilizing force inherent in our financial institutions showed itself particularly clearly in the boom of the 1920's and the severe depression that followed. The stock-market boom and the wide use of the investment trust and holding company certainly led to much unwise investment. The violence of the collapse in the stock market, and the difficulties experienced by banks and other financial institutions as security prices declined, played a significant role in accentuating the severity of the depression. Indeed, these financial developments had much to do with the final collapse of the American banking system in 1932–1933. The part played by financial activities and speculation in real estate and securities has loomed large in other periods also—for example, in the "panics" of 1837, 1873, 1893, and 1907, to mention some of the more familiar ones. Fortunately, the financial and banking reforms of the 1930's have significantly reduced this element of instability in the American economy.

STRUCTURAL CHANGES MAKING FOR GREATER STABILITY

There is widespread agreement that, since the 1930's, important structural changes have occurred that significantly reduce the tendencies toward instability described in this chapter. While we shall confine ourselves to the United States, similar stabilizing changes have been occurring in other countries. The more important of these structural changes have been the following.[30]

1. *Banking and financial reforms,* which have greatly strengthened the banking system, reduced the extent of financial speculation, and otherwise improved financial practices. These reforms include deposit insurance, regulation of security issues and security trading, the development of amortized mortgages (and of guaranteed and insured mortgages), and so on. We should also include here improvement in the conduct of monetary policy, so that we can safely count on the Federal Reserve System to follow an active program of monetary ease if a business recession should develop.

[30] For a useful discussion of these changes, see A. F. Burns, "Progress Towards Economic Stability," *American Economic Review,* vol. 50, March, 1960, pp. 1–19; also B. G. Hickman, *Growth and Stability of the Postwar Economy,* 1960. Various of these stabilizing changes will be discussed further in later chapters. For a review of European developments, see Angus Maddison, "The Postwar Business Cycle in Western Europe and the Role of Government Policy," *Banca Nazionale del Lavoro Quarterly Review,* June, 1960, pp. 99–148.

2. *The increased importance of the "automatic stabilizers,"* which help to hold up disposable income and consumers' expenditures when the GNP declines.[31] These "stabilizers" include social security payments (particularly unemployment compensation), corporation and personal income taxes, undistributed corporate profits, and, to some extent, subsidy payments to farmers. Thus, if a decline in investment causes the GNP to fall, undistributed profits and income taxes will fall more than in proportion, and unemployment compensation, other social security payments, and farm subsidies will rise. A substantial fraction of the decline in GNP may be offset in this way, so that the decline in disposable income and consumption is much less than it otherwise would be. The result is to reduce the size of the multiplier during contractions and thus make business recessions milder than would otherwise be the case.

3. *Government spending is a much larger fraction of GNP than before World War II,* and private investment (when measured in constant prices) is a somewhat smaller fraction of GNP than it was before the 1930's. Thus, as long as government spending is maintained, a given percentage change in private investment represents a smaller fraction of GNP than was once the case, and the destabilizing influence of changes in investment is further weakened by the effect of the automatic stabilizers.

4. *The structure of employment has been changing in a stabilizing direction.* Government, the service trades, and white-collar jobs generally are not subject to the wide cyclical swings in employment characteristic of manufacturing, mining, construction, and freight transportation; and it is these relatively stable kinds of employment that have shown the most rapid increases in recent decades.

5. *The government's commitment to maintain a high level of employment,* which was given concrete expression in the Employment Act of 1946, has undoubtedly influenced business and consumer expectations in a way that is favorable to stability. In particular, business firms and consumers are not so likely to rush to contract their expenditures if they are confident that government intervention will stop a recession before it has gone very far.

6. *Business seems to be managing its investment planning and its inventories more wisely*—for example, through long-term capital

[31] The automatic stabilizers are discussed in greater detail in Chapter 19.

budgeting and through economizing in the use of inventories. Plant and equipment expenditures can still be (and have been) cut back in recessions, and inventory investment is still highly volatile; but better business planning has probably made some contribution to greater stability.

7. *For a variety of reasons, prices and costs seem to be less variable over the cycle than was once the case.* Businessmen do not defer expenditures in recessions or try to liquidate inventories simply because they expect prices to fall rapidly. At the same time, the steady upward pressure on wages may help to support workers' incomes during business contractions, although the evidence on this point is by no means clear.

8. Finally, *the secular forces mentioned earlier* in this chapter have operated in the postwar period to create strong booms and mild depressions. These include the rapid rate of population growth and the acceleration of technological change, to which should probably be added the steady rise in prices. It *may* be, also, that, with the greater emphasis on systematic laboratory research, technological change will proceed more smoothly and steadily than in the past.

These various changes have undoubtedly made the American economy more stable than it was before World War II; and, because of them, we are not likely to experience again a catastrophic depression like that of the 1930's. But some degree of short-run economic instability will continue to be with us, and there has been no interruption in the steady recurrence of the milder type of business cycle. The factors that cause boom periods to end are still largely with us. What these stabilizing changes mean primarily is that fewer mistakes are likely to be made during booms and that the process of readjustment during cyclical contractions should be briefer, milder, and less painful than during the more severe depressions of the past.

THE NEXT STEPS

This chapter does not present a theory of economic fluctuations. What we have tried to do is to pick out those characteristics of a private-enterprise economy that tend to make for secular and cyclical change. Two further steps are necessary before we can attempt a detailed analysis of how business fluctuations are generated.

We must first examine the record of economic change and try to separate the various types of long- and short-run movements that have occurred. This is done in Chapter 9, and the results should tell us what kinds of fluctuations we must try to explain. Then, in Chapter 10 we shall examine in some detail how the more significant economic series behave as total business activity moves up and down. With this background of what does happen during business cycles, we shall then be ready for the detailed analysis of causes that follows in Chapter 11.

CHAPTER 9

THE KINDS OF ECONOMIC CHANGE

IN EXAMINING the dynamic characteristics of the American economy in Chapter 8 we saw that, broadly speaking, two sets of forces were at work to create constant change—one leading to gradual movements over long periods and the other creating short-period fluctuations. We emphasized that these two sets of forces react on each other and together create the actual pattern of change that we see in statistical series.

Our task in this chapter is to see if we can resolve this total pattern of change, with its mixture of short- and long-run movements, into its basic elements. Can we separate the gradual secular movements from the short-run fluctuations? Can we resolve the latter into purely cyclical and possibly other types of fluctuation? In particular, can we mark off that part of the total movement that represents "the business cycle"?

There is general agreement among economists and statisticians that the dynamic forces operating on the economy create at least the following types of movement: long-run or secular movements, seasonal variations, cyclical fluctuations, and random or sporadic changes that do not fit into the other three categories. This is approximately the classification that will be found in any elementary text on economic statistics. But the agreement ends at this point. In particular, uncertainty exists as to the precise nature of secular movements and of cyclical fluctuations.

SEASONAL VARIATION

Let us begin, however, with the simplest of the types of fluctuation just mentioned—purely seasonal movements that repeat them-

selves in more or less the same fashion every year. Unlike secular and cyclical movements, this type of change does not arise directly out of the dynamic forces discussed in Chapter 8. Seasonal fluctuations are directly related to the changing seasons of the year; and, like the seasons, they tend to repeat themselves, though the precise seasonal pattern may change with the passage of time. A number of methods are available for measuring that part of the total fluctuation in economic series which is due to purely seasonal causes, and there is general agreement that the seasonal element should be removed from statistical series whose cyclical fluctuations we may wish to study.[1]

The need to identify, measure, and eliminate seasonal movements is well illustrated by Figure 24, which portrays the behavior of department store sales both before and after adjustment for seasonal variation. The seasonal movement is striking. The rise in sales at Easter and, particularly, in the Christmas shopping season dominates the "unadjusted" series (i.e., before correction for seasonal variation). The effect of removing the seasonal element can be seen in the "adjusted" series plotted in Figure 24. Now the cyclical movement is much more clearly observable. We can, for example, trace out the mild recessions in 1953–1954 and 1957–1958. We can see that department store sales were declining in the latter half of 1953 and the closing months of 1957, although this is not clearly apparent in the original unadjusted data.

Although the business cycle is the chief cause of instability in our economy, the burden imposed on society by seasonal fluctuations should not be minimized. Sharp seasonal swings in production usually mean that some surplus labor must exist to meet peak seasonal needs, but it will be unemployed at other times of the year. Seasonal unemployment is particularly to be found in agriculture—especially where hired labor is important in harvesting crops —but it exists also in various lines of industry and trade. Seasonal

[1] We shall make no effort in this section to describe the statistical methods that are used to measure and eliminate seasonal variation. The student will find the simpler of these methods adequately described in any of the better textbooks on economic statistics. For a more technical discussion, including a description of how electronic computers are being utilized in making seasonal adjustments, see Julius Shiskin and Harry Eisenpress, *Seasonal Adjustments by Electronic Computer Methods,* National Bureau of Economic Research Technical Paper 12, 1958.

instability also adds to the cost of doing business. Seasonal fluctuations in demand, for example, require either extra equipment which is idle part of the year or the accumulation of inventories in periods of slack demand which can be sold when the seasonal peak occurs. Either procedure adds to the cost of doing business. Where

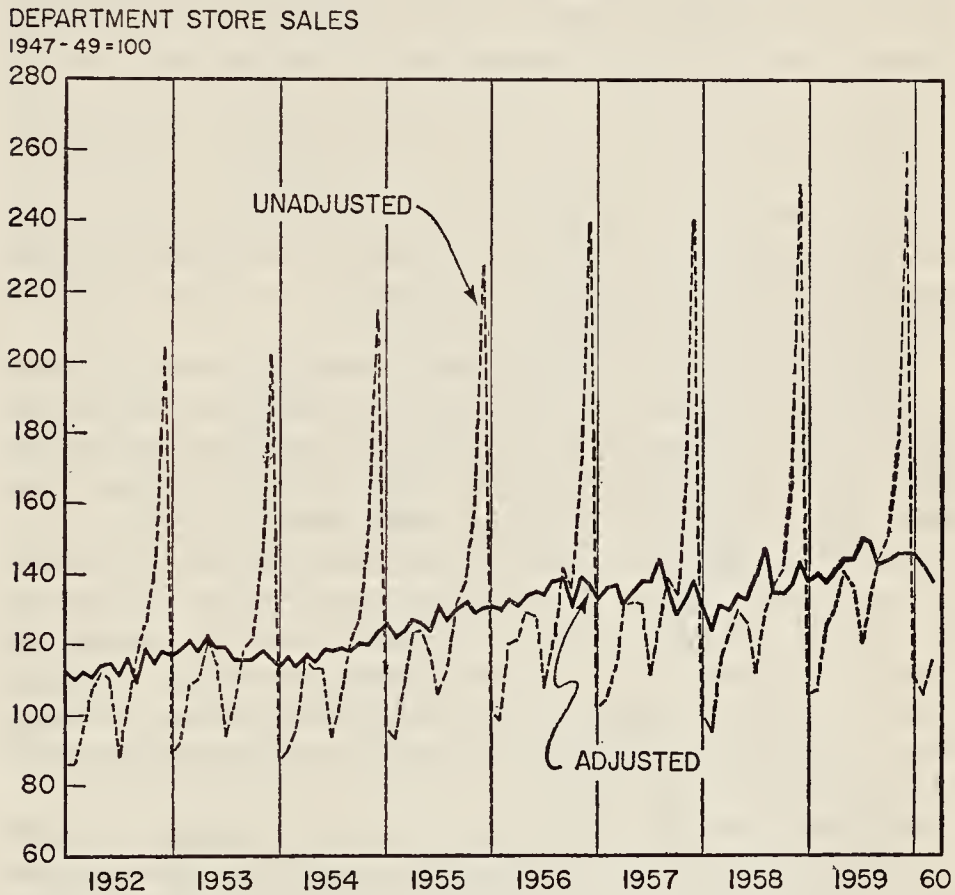


FIGURE 24. Department Store Sales, with and without Seasonal Adjustment, 1952–1960.

From data in the *Federal Reserve Bulletin,* December, 1957, and later issues.

production is seasonally unstable, wage rates may be higher for this reason (as has been the case in the construction industry), labor turnover is greater, and other extra costs must be incurred because of the recurring expansion and contraction of operations.

The causes of seasonal variation are fairly clear. The most important factor is the weather, but social custom is also of great

importance. The seasonal patterns in agricultural production and in construction are examples of seasonals attributable to the weather. The seasonal pattern in department store sales is a joint product of weather and social custom. Our customary method of recording the months, as represented by the Gregorian calendar, introduces a seasonal element into sales and production series, since not all months contain the same number of days. Business holidays and five-Sunday months create the same sort of complication.

Seasonal patterns have an uncomfortable habit of changing, sometimes very gradually and sometimes abruptly. While the seasonal pattern of the weather is subject to only minor and erratic change, custom may alter; and the development of technology and business practices may significantly change the way buyers or sellers typically react to differences in the seasons.

Gradual changes in seasonal patterns are fairly common. The development of an elastic credit supply and centralized banking reserves under the Federal Reserve System steadily reduced the amplitude of the seasonal swings in short-term interest rates until they virtually disappeared in the 1930's. The spread of air conditioning has significantly increased the residential use of electricity in the summer months. Increasing emphasis on style and fashion has altered seasonal patterns in numerous lines—for example, in various branches of the shoe and clothing industries. Changing consumers' tastes, aided by advertising, have significantly reduced the contrast between the winter low and the summer high point in the production of ice cream.

Sometimes the pattern of seasonal variation changes abruptly. During World War II, when an all-out production effort was being made, no seasonal variation was recognized at all in many lines of production, although a seasonal pattern had existed before the war and re-established itself after the war. Changes in the time of introducing new models has on several occasions led to abrupt changes in the seasonal pattern of automobile production.

The fact that seasonal patterns change creates some technical problems for the student of cyclical fluctuations. We can eliminate seasonal fluctuations only if we can isolate them. If the character of the seasonal variation is changing, we may not be able to measure

precisely the "true" seasonal pattern. The very concept of seasonal variation implies typical or average behavior. Techniques exist for computing changing seasonals, but these methods in turn assume that the seasonal pattern is changing in a regular and systematic way that can be surmised from past experience. The less this is true, where we know from general observation that the seasonal movement is changing, the more does arbitrary judgment enter into our measurement of the seasonal pattern.[2] Although this difficulty tends to blur somewhat the distinction between seasonal and other short-period fluctuations, cyclical movements are ordinarily so pronounced that they show up distinctly when the conventional methods are used to eliminate seasonal movements. But difficulties sometimes do arise in interpreting month-to-month movements in periods too short to reflect the full impact of cyclical forces.

While on the subject of changing seasonals, we should pay brief attention to a problem concerning which not a great deal is known—the possibility that seasonal and other types of fluctuations are causally related. The usual assumption is that the absolute amplitude of the seasonal movement varies with the level of the series in question—thus we usually express the seasonal adjustment factor as a percentage of the nonseasonal part of the series—but beyond this it is ordinarily assumed that cyclical and secular movements do not affect the seasonal pattern. This assumption is not always warranted. The changing phases of the business cycle may cause the seasonal pattern to change. For example, when sales and prices are declining, businessmen may adopt a "hand-to-mouth" buying policy; and this may affect the seasonal behavior of production and inventories. Cyclically poor business may affect the timing of new models and the introduction of new styles. Not only may the cycle affect the seasonal pattern, but the reverse may also be true. For example, it is at least plausible that a cyclical revival can begin more easily when a seasonal upturn occurs than when sales and production are seasonally declining. To some extent, the mere existence of seasonal instability throughout the economy creates conditions that affect the course of the business cycle. For

[2] The Board of Governors of the Federal Reserve System has had to face this problem in making the seasonal adjustments for many series in their production index. For description of the method used and of some of the problems involved, see H. C. Barton, Jr., "Adjustment for Seasonal Variation," *Federal Reserve Bulletin*, June, 1941, pp. 518–528. Cf. also Shiskin and Eisenpress, *op. cit.*

example, the larger stocks are for seasonal reasons, the more purchases can be reduced if business declines. Seasonal patterns may also be related to secular movements. For example, a secular rise or fall in prices may make businessmen more or less willing to carry inventories, and larger or smaller inventories may affect the seasonal pattern of production.

THE NATURE OF SECULAR MOVEMENTS

Let us now turn to the gradual, long-run movements that can be traced in various aspects of economic activity. The secular movement or "trend" in any series is the underlying movement that persists over a period which is long relative to the business cycle. The existence of such secular movements is readily established by simple observation of economic series over periods encompassing a number of business cycles.

As in the case of seasonal variation, the economist and the businessman have a twofold interest in the long-run movements in economic series. There is the need to study these movements for their own sake—to see what they are like, what causes them, and whether they are likely to continue in the future—and we also want to be able to eliminate them in order to study the shorter-run fluctuations in isolation.

In Chapter 8, we described the more important causes of secular change, particularly of long-run growth in total output. In seeking to isolate the secular movement in a series, we try to measure the gradual change that results from the working of these secular forces. Strictly speaking, however, we can never do this precisely, because cyclical influences modify the effect of the long-run forces at work. Secular movements result, but in an important sense they are the joint product of both short- and long-run influences operating on the economy. Over long periods, however, it is the long-run forces mentioned in Chapter 8 that predominate in shaping the path of the secular movements.

Underlying secular trends may be observed in many different aspects of economic activity—in total production and in the production of individual industries, in index numbers of prices and in individual prices, in the value of production, trade, etc. (which reflect the movement of both physical volumes and prices), and in other economic variables. The secular movements observable

over the last century or so in the volume of production and trade are different from those that show up in price series; interest rates or the volume of employment will reveal yet other types of secular movements. In this chapter we shall be concerned particularly with secular movements in the physical volume of production and the level of prices.

When we deal with periods of considerable length, it is useful to make a distinction between two types of secular movements. One we may call the "primary trend" and the other the "intermediate trend." Both types of movement may be illustrated by the long-run movement of GNP shown in Figure 25. This chart plots the course of the GNP in constant prices from the 1870's to the 1950's. We have drawn through the original data an undulating line (based on a five-year moving average) that suggests that the growth of output in the United States has been subject to long swings, considerably longer than the ordinary business cycle, during which output alternated between periods of particularly rapid growth and periods of somewhat slower growth. These long swings can be taken as a representation of the intermediate trend in this series.

Underlying this intermediate trend, one can visualize a primary trend that would rise smoothly throughout the whole period and around which the intermediate trend would undulate. Such a primary trend might be represented by a smooth curve, possibly fitted mathematically. We have not fitted a primary trend to this series for reasons that we shall mention later.

The dividing line between primary and intermediate trends is not a precise one, and not all economists are agreed that the distinction should be made. The most gradual type of change in the entire period being studied, which underlies all the other types of movement discernible, is the primary trend. There may possibly be, however, other gradual movements, longer than the business cycle, which are shorter in duration than the primary trend. These are intermediate trends—intermediate between the primary trend and the business cycle. Since the term *secular trend* is loosely used in practice and may refer to either intermediate or primary trends, we shall use it, along with *secular movement,* only as a generic term to refer to any type of gradual change longer than the business cycle.

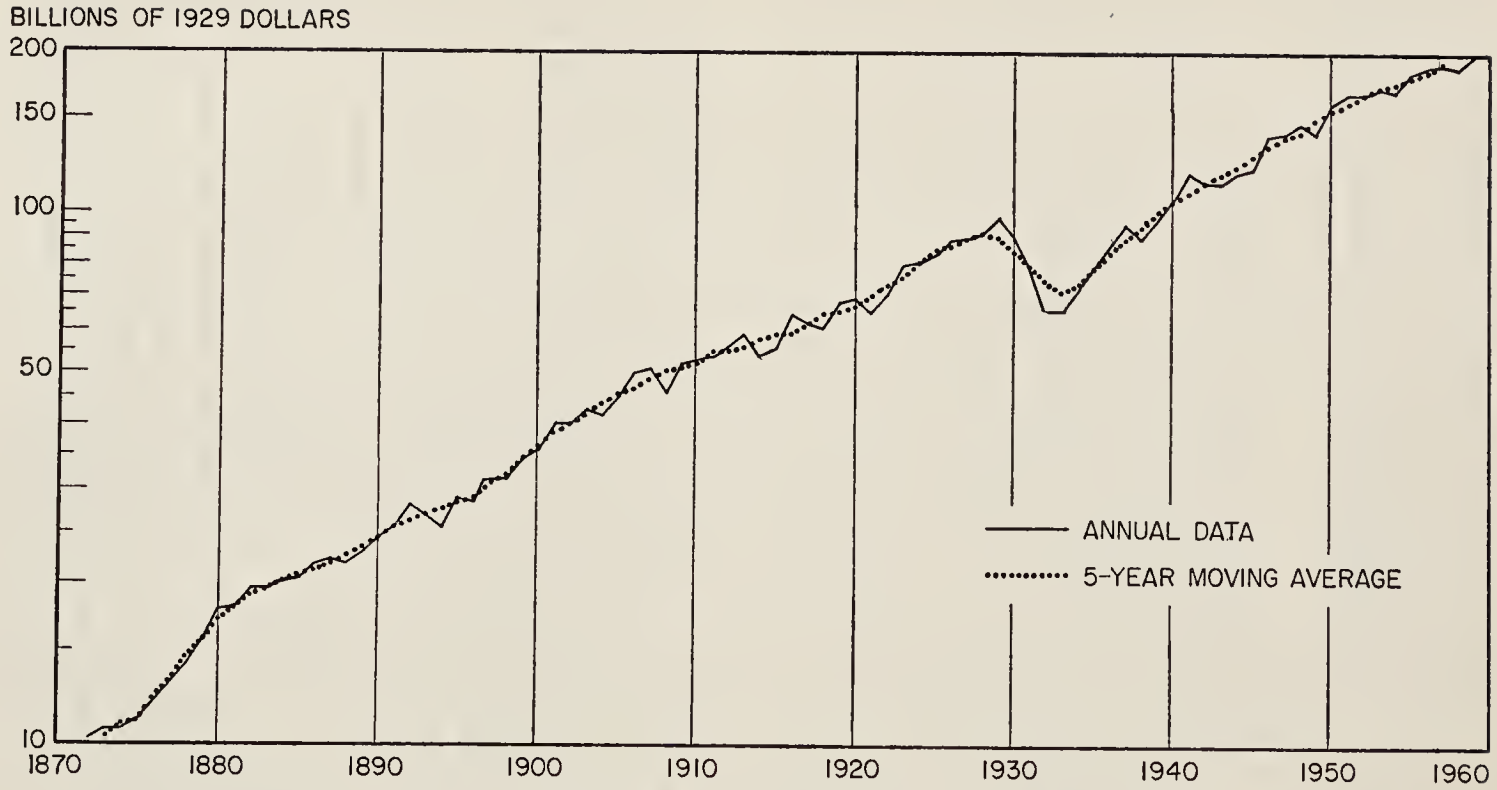


FIGURE 25. Gross National Product, in 1929 Prices, with Five-Year Moving Average, 1869–1959.

Annual data are those of Simon Kuznets through 1953, to which have been spliced Department of Commerce figures for 1954–1959.

Although the distinction between "primary" and "intermediate" trends is a useful one, it should be emphasized that both types do not necessarily show themselves in all series. Index numbers of wholesale prices for the 150 years before World War II, for example, go through alternating periods of secular rise and fall (see Figure 28 on page 235), but there is no evidence of a primary trend extending over the entire period. Other series may show a primary but no intermediate trend. When a relatively short period is taken—say, 20 or 30 years—it is impossible to distinguish between possible primary and intermediate trends. A "secular trend" fitted to such a short period may reflect chiefly an intermediate secular movement which may change significantly in a decade or two, and the primary trend may not be discernible unless a much longer period is taken.[3]

When we try to identify and measure the secular movements in a series, we face such questions as the following: What is our understanding of the secular *and cyclical* forces at work during the period in question? We do not want the cycle distorted by elimination of the wrong sort of trend, nor do we want the trend to be unduly influenced by short-run cyclical forces. Should we recognize the existence of both primary and intermediate trends? If so, how can we best represent the two types of trend? Should the primary trend be represented by a straight line or by a more complicated mathematical function? Should any breaks (i.e., abrupt changes) in the primary trend be recognized? And how shall we portray the intermediate trend movements? Some of the most troublesome questions arise when the secular movement follows an undulatory course and the nature of the primary trend is not obvious from visual examination and economic analysis. In such cases, we may be unable to decide whether the total secular movement is better

[3] Unfortunately, there is no generally agreed classification and nomenclature for economic movements of longer duration than the business cycle. Many writers make no distinction between possible types of secular movements and refer to all such movements as secular trends. Others do make a distinction. Some of these use the term *secular trend* for what we have called the primary trend, and at least one author (A. F. Burns) applies the term *secular trend* to what we have called the intermediate trend. However, Burns uses *primary trend* for the very long-run type of movement in the same way that we do. Cf. *Production Trends in the United States Since 1870,* 1934, pp. 44–45. Cf., also, the distinction between primary and secondary secular movements made by Simon Kuznets in *Secular Movements in Production and Prices,* 1930, p. 72.

described by a succession of broken straight lines or curves joined together or by some form of smooth primary trend on which is superimposed some type of intermediate trend.

In considering all of these questions, we should keep in mind what has been called the "principle of economic meaning." That is, what we do must make economic and historical sense. There is no way of determining trends in blind, mechanical fashion. The secular movements we mark off must show the gradual effect of the secular forces we believe to be at work (as modified by cyclical influences), and elimination of these secular movements must not distort or conceal the cyclical fluctuations that we believe exist in the series in question.[4]

SECULAR TRENDS IN PRODUCTION

TRENDS IN TOTAL OUTPUT

Over the past two centuries or so, *growth* has been the dominant secular characteristic of production and trade in the countries of the Western world. This has been true for the whole economy of each country and also for most industries taken individually. This growth has been due primarily to the secular forces described in Chapter 8. These forces have created long-run patterns of growth in total production similar to that shown in Figure 25. If we were to fit a primary trend to the series in Figure 25, it would rise strongly but also turn out to be slightly concave downward on a semilogarithmic scale. This would be true even for the period 1880–1930, after which there seems to have been what looks like a break in the trend. A curved line that is concave downward on a semilogarithmic scale reflects a declining percentage rate of growth. Figure 25 seems to suggest that the rate of growth of total output in the United States before World War II was subject to some degree of retardation, although the retardation was not marked and hardly showed up at all over considerable periods.

We shall have more to say later in this chapter about the long swings traced out by the intermediate trend in Figure 25. It is

[4] For a stimulating discussion of a number of the issues raised in this section, as well as a painstaking examination of the nature of secular and cyclical movements in the United States between the Civil War and World War I, see Edwin Frickey, *Economic Fluctuations in the United States,* 1942. Cf. also J. A. Schumpeter, *Business Cycles,* 1939, vol. 1, pp. 199–200.

clear that, even apart from the ordinary business cycle, growth in some periods has been more rapid than in others; and it is these alternating periods of accelerated and retarded growth that are marked off by the intermediate trend.

This and other trend studies suggest that, depending on the period covered and the measure of output used, the primary trend in the total output of the United States has been rising at an average annual rate of between 3 and 4 percent per year; and there is some further evidence that the trend on a logarithmic scale, at least before World War II, had been flattening out somewhat i.e., that the percentage rate of growth had been tending to decline. As we should expect, the average percentage rate of increase has been greater in manufacturing, mining, and trade than in agriculture. The degree of retardation—that is, the extent to which the percentage rate of increase declines—seems to have been greatest in the extractive industries, e.g., agriculture, mining, and lumbering.[5]

It is not surprising that the growth of total output in the United States should have shown some retardation over the last century.[6] As Table 16 (page 184) suggests, this has been due chiefly to the decline in the rate of population growth. The rate of growth in population declined steadily from about 35 percent per decade in the 1840's and 1850's to 7 percent in the 1930's. Although we have had a population upsurge since then, the rate of increase is still lower than it was a century ago. Growth in output per capita, however, has shown no retardation. (See Table 16.) Nor has there been any retardation in the growth of output per man-hour. If anything, there is some evidence of acceleration rather than retardation. (See Figure 23, page 200.) Thus, retardation in the rate

[5] Cf. A. F. Burns, *Production Trends in the United States,* 1934, chap. 6; and Simon Kuznets and Raymond Goldsmith, *Income and Wealth of the United States: Trends and Structure,* 1952, pp. 99–106. For discussion of retardation tendencies in the growth of the total national income (corrected for price changes), see Simon Kuznets, *National Income: A Summary of Findings,* 1946, pp. 34–36, and Kuznets and Goldsmith, *op. cit.,* p. 51. Cf. also Table 16, p. 184, above. Cf. also Edward Ames, "Trends, Cycles, and Stagnation in U. S. Manufacturing Since 1860," *Oxford Economic Papers,* vol. 11, October, 1959, pp. 270–281. Ames believes that, for manufacturing output, the tendency toward retardation disappears if we include the period since 1930.

[6] Such retardation has been evident in other countries also. Cf. Simon Kuznets, "Population, Income, and Capital," in L. H. Dupriez, ed., *Economic Progress,* 1955, pp. 34–35, 44.

of population growth (combined with the decline in the length of the work week) has brought about significant retardation in the rate of increase in the total number of hours worked per year by the American population, but output per hour worked has continued to rise at an undiminished, even a slightly accelerating, rate. In the next several decades, the rate of increase in the labor force and in total hours worked will be higher than during the last 50 years; the rate of increase in output per man-hour may accelerate still further; and, hence, it is quite possible that the rate of growth in total output during the 1960's and 1970's will be higher than during the preceding several decades. (This will depend in part on whether aggregate demand rises rapidly enough.) But for the time being, until later decades supply us with more evidence, we cannot say how much of the growth in output since World War II reflects a change in the primary trend and how much represents the gradual undulations of the intermediate trend.[7] That there has been some break in the primary trend seems highly probable.

FITTING SECULAR TRENDS

There are various mathematical curves that may be fitted to a series to yield a declining percentage rate of growth. Each makes certain assumptions about the nature of the secular changes taking place. One such curve is illustrated in Figure 26. Here a logistic growth curve has been fitted to an index of industrial production for the half-century or so preceding World War II. This particular curve is frequently used to describe the "law of growth" of population and of production in certain industries, and is also extensively used in the field of biology. Its character, however, makes it inappropriate as a representation of the secular trend of *total* production. For example, the curve approaches but never exceeds an upper asymptote, which is marked on the chart as the "saturation level." While this may on occasion be a valid assumption for a particular industry, it is clearly inappropriate for any large segment of total production, with the continued expansion possible

[7] For a projection to 1975 of labor force, hours worked, output per man-hour, and potential total output, see Table 18 and the source there cited. Cf. also Simon Kuznets, *Capital in the American Economy: Its Formation and Financing* (in press), chap. 10.

through new wants and new products. The logistic curve also makes the absolute amount of change (the slope of the trend on an arithmetic scale) depend on the absolute level of the curve and its distance from the "saturation level." To the left of the "critical point" on the chart, the absolute amount of change is increasing; to the right of this point, the amount of increase is decreasing.

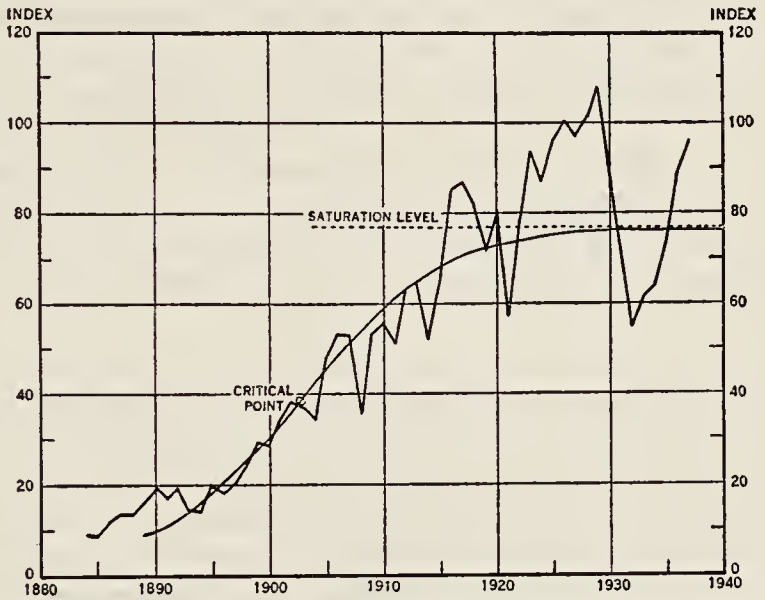


FIGURE 26. Logistic Curve Fitted to Index of Industrial Production.

From *The Analysis of Economic Time Series* by H. T. Davis, based on the production index of the Standard Statistics Company (Cowles Commission for Research in Economics Monograph No. 6, p. 255) by permission of Cowles Commission for Research in Economics. Copyright, 1941, by the Cowles Commission for Research in Economics.

The *percentage* rate of increase decreases throughout.[8] The role of the upper limit in this curve makes the logistic have little economic meaning when applied to total production. We can also see that, apart from its economic meaning, the particular curve drawn in

[8] For further description of the properties of the logistic, see, for example, Simon Kuznets, *Secular Movements in Production and Prices,* p. 64, and the references there cited. For a more detailed mathematical discussion, see H. T. Davis, *The Analysis of Economic Time Series,* 1941, pp. 247 *ff.* The weaknesses of the logistic, as a representation of the primary trend in total production, is shared by the Gompertz curve, which has similar properties.

Figure 26 provides a relatively poor fit. It flattens out too rapidly and at too low a level in the latter part of the period; and it also does not give a particularly good fit for the period before World War I.

The primary trend in total production in the United States during the last century or so seems to be best described by the so-called logarithmic parabola. The equation of this curve provides for a given percentage rate of increase which is subject to a constant rate of retardation.[9] The curve is concave downward on a semi-logarithmic chart, the degree of concavity depending on the constants in the equation.[10] The properties of the curve, as a description of the primary trend in total production, are reasonable in terms of economic analysis. They conform to our finding that total output has been growing at close to a constant percentage rate, with possibly some slight tendency for the rate of increase to decline.

This is the place for a word of caution. A primary trend, however well it may fit the original data for some past period and however reasonable it may seem in terms of economic analysis, cannot safely be projected for any long period into the future. The underlying forces creating the trend may change. Or the data for future years may suggest that, though the underlying secular forces are continuing without drastic change, the additional evidence requires that our fitted trends be recomputed. And, if intermediate trends exist, they may for several decades move the curve of output away from the line that best describes the very long-run primary trend. We have already encountered this range of problems in connection with the primary trend that might be fitted to the growth of output in Figure 25. The forces that tended to make for retardation in output growth in the century before World War II may well have changed, particularly because of the upsurge in the

[9] The logarithmic parabola can also provide for a constant rate of acceleration, which would be the case if the underlying data tended to bend upward (rather than downward) on a logarithmic scale.

[10] The equation of this curve may be written in the form $\log y = a + bt + ct^2$, where b is the logarithm of the ratio of growth in the year of origin and $2c$ (i.e., the second derivative) is the logarithm of the constant retardation ratio. If, for example, the retardation ratio is 0.998, this means that the ratio of growth in any year is 99.8 percent of what it was the preceding year. By ratio of growth we mean merely one plus the percentage rate of increase. For an excellent nontechnical explanation of the properties of the logarithmic parabola, see Frickey, *op. cit.*, pp. 144–146; also Burns, *op. cit.*, pp. 97–98.

rate of population growth. A logarithmic parabola fitted, for example, to the period 1880–1930 would not fit very well the data for the three following decades.

PRODUCTION IN INDIVIDUAL INDUSTRIES

Thus far we have been considering secular movements in production aggregates, covering a large part or all of the economy. Similar considerations apply when we attempt to identify the secular movements in the production of particular industries, though in some respects the problem is somewhat simpler than in the case of aggregate output.

Single industries tend to follow a typical pattern of development that can be called, very loosely, a "law of growth." Rapid expansion at an increasing absolute rate follows the birth of the industry. Eventually the rate of increase begins to decline; and ultimately, unless there are important new developments in technology or on the demand side, the industry reaches maturity and ceases to expand much further and perhaps begins to decline. This type of secular behavior can be described by several mathematical curves, all of which trace out an S-shaped pattern such as that shown by the fitted trend in Figure 26. Of these, perhaps the most widely used is the logistic. The main properties of the logistic curve were mentioned earlier. Perhaps its chief weakness is that it approaches an upper asymptote and eventually moves virtually horizontally. Provision for a secular decline after a period of expansion must be made by fitting a new curve with a downward inclination. Some growth curves—for example, the logarithmic parabola—are not subject to this difficulty.

Growth curves that are S-shaped when drawn on an arithmetic scale rise at a continuously decreasing rate on a semilogarithmic scale, as, for example, does the trend fitted to coal production in Figure 27. As noted, these curves may either approach more and more closely to an upper limit, or they may eventually begin to decline. The curve drawn in Figure 27 reaches a peak in the neighborhood of World War I, and projection of the trend shown implies a forecast that the primary trend in coal production will be downward in future decades.[11]

[11] For other examples of growth curves fitted to individual industries, see E. R. Dewey and E. F. Dakin, *Cycles: The Science of Prediction*, 1947. A much more scholarly but earlier book is Kuznets, *Secular Movements in Production and*

The fact that industries go through typical growth patterns should be considered in connection with the discussion of technological change and the acceleration principle in Chapters 6 and 8. New industries first go through a period of rapid growth, and this is when they provide the greatest stimulus to net new investment. As industries approach maturity, there is less need to expand capacity further; investment is primarily then for replacement and modernization. These secular changes in investment opportunities

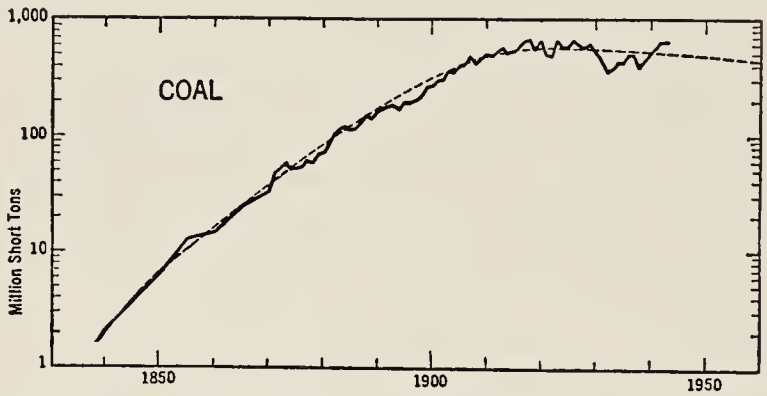


FIGURE 27. United States Coal Production, with Fitted Primary Trend.

From *Cycles: The Science of Prediction,* by Edward R. Dewey and Edwin F. Dakin (p. 40). Copyright, 1947, by Henry Holt and Company, Inc.

created by the birth and growth of important industries at different times cannot help but play an important role in the shaping of intermediate secular movements and the major cyclical fluctuations in business activity.

Two additional points should be made about growth curves for individual industries. First, curves such as those in Figures 26 and 27 may need to be given an upward tilt to reflect the fact that a steadily rising national income may lead to a more or less corresponding rise in the "saturation level" for a particular industry. Secondly, the growth trend for a particular industry may be subject to more or less abrupt shifts—due, for example, to important new technological changes, the discovery of new sources of supply

Prices, op. cit., in which there are a large number of growth curves fitted for various periods up to about 1925.

(as in mining), or the sudden appearance of new forms of competition.[12]

SECULAR MOVEMENTS IN PRICES AND OTHER ECONOMIC SERIES

The long-term movements in commodity prices are entirely different from those to be observed in production data. This is clearly evident from an examination of Figure 28. (See also Figure 1 on

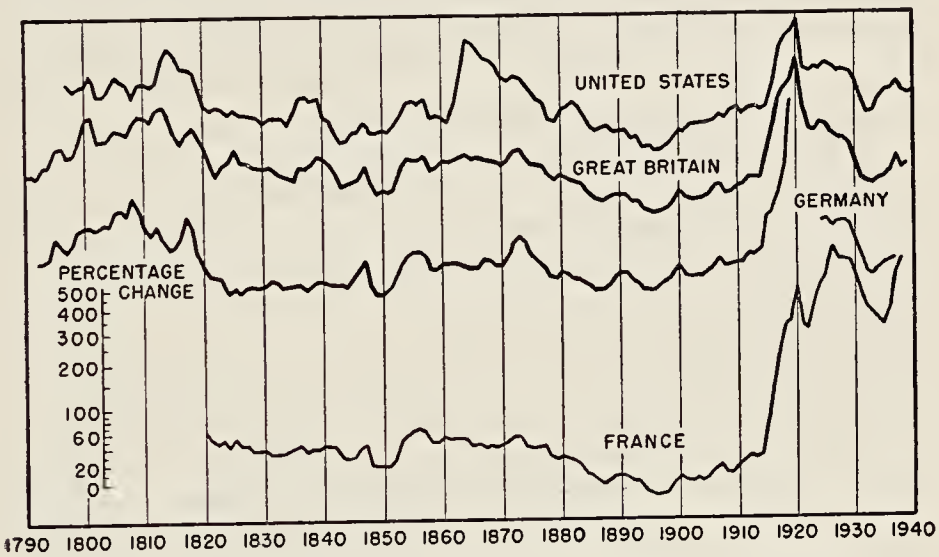


FIGURE 28. Wholesale Prices in Four Countries, 1790–1940.

From *Measuring Business Cycles* by A. F. Burns and W. C. Mitchell (p. 439), by permission of National Bureau of Economic Research. Copyright, 1946, by National Bureau of Economic Research.

page 4.) In none of the four countries included is there any persistent drift in prices, upward or downward, that would provide the basis for a single primary trend fitted to the entire period from 1790 to 1940. There have been secular movements in prices, but these movements have regularly reversed themselves, thus creating an impression of "long waves" or cycles. The question whether these intermediate secular movements are best interpreted as cycles that inevitably repeat themselves or whether they should be considered merely as a succession of trends, each more or less independent of

[12] For an interesting analysis of the growth forces operating on one important group of industries, see M. J. Ulmer, *Capital in Transportation, Communications, and Public Utilities: Its Formation and Financing*, 1960.

those that precede and follow, is still being debated. We shall look into this issue further in the next section. Here we shall consider merely the direction of the secular price movements shown in Figure 28.

Examination of the chart suggests that there were three complete swings of secular rise and fall in prices from about 1790 to the 1930's. In the United States, the first secular rise ended with the War of 1812. After a sharp drop to about 1820, prices drifted downward until the 1840's. The secular rise that then followed is associated with the gold discoveries of that decade, and it culminated in the marked inflation of the Civil War period. Then ensued a marked downward secular movement which reached a low point in the 1890's, and again important gold discoveries were associated with a new secular expansion which reached a peak in 1920. Prices then tended downward until the early 1930's. Since then we have had a new long upswing in prices, the fourth in the last 175 years, which, unlike previous long price upswings, did not reverse itself at the end of a major war.

Thus, prior to World War II, the peaks in these long-term price movements in the United States were all associated with wars. If we could exclude the periods of wartime inflation and of downward readjustment after these inflationary periods, the secular swings in prices would appear much less pronounced than they do in Figure 28. This is not to deny that factors other than wars have created secular movements in prices. There is no question that an expanding monetary base led to an upward movement in prices in the 1850's; and there is no doubt that the tremendous expansion in world production and trade from the 1870's on, taken together with a money supply that expanded much less rapidly, led to a marked downward trend in the price level which did not end until the 1890's. There then followed nearly 20 years of secularly rising prices before the inflationary effects of World War I began to be felt.

We shall have a bit more to say about secular movements in prices in the next section, where we shall consider whether the available evidence supports the hypothesis that there are "long waves," in production as well as prices, that inevitably repeat themselves. Before turning to that discussion, we may merely note that secular movements are to be found in many other economic series

besides those dealing with production and prices. Some of these series will reflect secular forces making for growth. Thus, such series as money in circulation, bank deposits, bank clearings, value of foreign trade, etc., will display marked primary trends reflecting various of the secular forces making for long-run expansion in the total volume of business activity, as well as, perhaps, special secular influences of their own. Some series may show a primary trend that is associated chiefly with long-run factors more or less peculiar to the variable in question. Superimposed on these primary trends —whether reflecting chiefly the growth of total activity or other secular forces—may be intermediate trends.[13]

ARE THERE LONG WAVES IN BUSINESS ACTIVITY?

The pronounced long swings in prices and the apparent existence of intermediate trends in various production series raise an intriguing question: Are there waves or cycles in business activity that are longer than the ordinary business cycle? We have not yet defined the business cycle or considered the period it typically covers. We may, however, anticipate the next section and define business cycles as recurring alternations of cumulative expansion and contraction in economic activity, the total period of rise and fall typically ranging between 2 and 11 years. Are there also long swings in business that last longer than 11 years?

The possibility that such long waves exist has received the attention of economists, particularly during the last 40 years or so. At first, attention was concentrated on the possibility of a long wave of 50 to 60 years. More recently, the work of several investigators suggests that, if there is a long cycle in economic activity, it is more likely to have a shorter duration—say, of 15 to 20 years.

THE KONDRATIEFF LONG WAVE

Belief in the existence of a long wave of about 50 years is largely the result of the work of a Russian economist, N. D. Kondratieff,

[13] It should be noted that *value* series result from multiplying physical volume by prices and will therefore reflect secular movements in prices as well as forces making for expansion in the physical volume of activity. Thus the trend in the national income, uncorrected for price changes, could be expected to reflect the intermediate secular movements in the price level as well as the primary and possibly intermediate trends in the physical volume of total production.

who published his results during the 1920's.[14] J. A. Schumpeter, certainly one of the most brilliant of the students of business cycles, attempted to integrate these long waves into his study of economic development and cyclical fluctuations; and various other writers have accepted the fact of alternating secular movements in economic activity that roughly correspond to the long swings in prices previously discussed.[15] In the last 10 or 15 years, however, the existence of these "Kondratieff" waves has been increasingly questioned. Let us look at some of the evidence.

The first piece of evidence is the long "cycles" in prices previously described. There is considerable doubt, however, as to whether these secular price movements should be described as cycles. The peaks of these long swings are chiefly associated with periods of war inflation, and the amplitude of the swings is much less pronounced if these inflationary periods are omitted. In addition, between wars, there have been in some countries considerable periods during which prices have shown no persistent tendency either to rise or to fall, and this is scarcely consistent with the picture of long cycles some writers have attempted to build.[16]

Kondratieff believed that these long waves existed not only in prices but also in production, and he advanced the thesis that capitalist evolution breeds long cycles or waves in business activity that can be seen as deviations from the primary trend. Table 20 gives the approximate dates for the turning points of the long waves discovered by Kondratieff and compares them with the dates of the secular movements in prices previously discussed. It will be seen that the dates Kondratieff gives for his long waves are essentially the dates of the turning points in prices. He was much less successful in finding long cycles in production series than

[14] See N. D. Kondratieff, "The Long Waves in Economic Life," *Review of Economic Statistics*, vol. 17, November, 1935, pp. 105–115. This is the only report by Kondratieff available in English. For a bibliography and summary of his work, as well as a critical evaluation of his results and an account of the reaction of Russian economists to his findings, see the highly useful article by George Garvy, "Kondratieff's Theory of Long Cycles," *Review of Economic Statistics*, vol. 25, November, 1943, pp. 203–220.

[15] For a summary of some of these views, see A. H. Hansen, *Business Cycles and National Income*, 1951, chap. 4. A recent empirical investigation is that of L. H. Dupriez, *Des Mouvements Économiques Généraux*, 1947. See also his more recent *Philosophie des Conjonctures Économiques*, 1959, chaps. 10–11.

[16] Cf. A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, 1946, p. 440.

he was in the price data he examined. In about half of the production series he studied, no long cycle at all was discernible; and in most of the others he had data covering only one long cycle. Only eight of his series went back far enough to include as much as two long waves, and six of these were price series. It is clear that Kondratieff did not establish the existence of long swings, with the turning points summarized in Table 20, for other than price move-

TABLE 20. Turning-Point Dates of Kondratieff's Long Waves and Secular Movements in Prices [a]

Turning Point	Kondratieff's Long Waves	Price Movements United States	Great Britain
Trough	1785–1795	1789	1789
Peak	1810–1817	1814	1813
Trough	1844–1851	1843	1849
Peak	1870–1875	1864	1873
Trough	1890–1896	1896–1897	1896
Peak	1914–1920	1920	1920
Trough		1932	1933

[a] Based on a table in A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, 1946, p. 432.

ments. The existence of the long swings in prices has already been recognized, though, as we have seen, there is reason to doubt that they should be viewed as cycles.[17]

SCHUMPETER'S POSITION

Although Schumpeter has not presented additional statistical evidence to show that there are clearly marked long waves in measures of the physical volume of activity, he has used the long-wave hypothesis in a suggestive way in interpreting the course of economic development.[18]

Schumpeter believes that business cycles of different durations occur simultaneously, and that the course of economic activity can be satisfactorily described by a pattern of three different cycles super-

[17] For a more detailed discussion of the weaknesses in Kondratieff's empirical work, see Garvy, *op. cit.*, esp. pp. 206, 216–219, on which a large part of this paragraph is based.

[18] Schumpeter's analysis is most fully presented in his two-volume work, *Business Cycles*. For a briefer statement see "The Analysis of Economic Change," *Review of Economic Statistics*, vol. 17, May, 1935, pp. 1–10.

imposed on each other. These three are a short cycle of 40 months, a longer cycle of about 10 years, and a long wave or "Kondratieff cycle" of some 50 years. In Schumpeter's view, "innovations" (including territorial expansion and improvements in marketing and management as well as technological change in the narrow sense) are responsible for all three types of business cycles. The long waves are associated with the succession of "industrial revolutions" of the last 150 years or so—i.e., they result from the most revolutionary and wide-sweeping types of technological change. Thus, he associates Kondratieff's first long wave (roughly 1790 to 1842) with the so-called Industrial Revolution, which affected particularly the textile and iron industries. The second wave (1842–1897) covered the age of steam and steel, with particular emphasis on the revolutionary effects of the development of railroad transportation. The third wave (from 1898 on) was the age of automobiles, electric power, and chemistry.[19]

Schumpeter's dates for these long waves do not altogether agree with Kondratieff's, and in this connection it is important to distinguish between the movement of production and the movement of prices. Schumpeter assumes that a new cycle begins only after some period of recovery from the preceding downswing. After such a revival, with prices already rising, important new innovations stimulate investment and lead to an expansion of bank credit. This is the prosperity phase of the long wave—during which prices are rising but total output is not necessarily expanding (relative to the primary trend) since the new methods of production result in increased output only with a lag. The peak of the cycle marks the beginning of the decline in prices and bank credit (the declines may be merely relative to an underlying primary trend), but production now rises rapidly as the new methods pour out an increased flow of goods. During this "recession" phase, rapidly expanding output goes with gradually falling prices. Eventually, produc-

[19] Alvin Hansen also emphasizes the role of technology (as well as the opening up of new territory and new resources and changes in the rate of population growth) in explaining the fact that there seem to have been in the past prolonged periods of accelerated or retarded growth, associated with secular movements in prices. However, he does not look on these intermediate secular movements as constituting cumulative, self-generating cycles. See *Business Cycles and National Income*, chap. 4. In this respect, his views resemble those of Wicksell and Spiethoff.

tion may also decline (relative to the primary growth trend) if the maladjustments created during the Kondratieff downswing are serious enough. This is the "depression" phase. Thus, in Schumpeter's scheme, the peaks of the long waves in production should not coincide with those of the long cycles in prices. Paradoxical as it may sound, output should rise during at least part of the "downswing" in the Kondratieff cycle.

Schumpeter did not satisfactorily demonstrate that measures of production, prices, and other economic variables actually have behaved as his theory of long cycles suggests that they should. Nonetheless, his analysis of the effects of waves of major innovations does throw some light on the reasons for intermediate secular movements, particularly in production. To put it another way, Schumpeter's analysis sheds light on some of the possible causes of intermediate secular movements in production, prices, and other variables without proving that these movements are part of a common cyclical pattern which tends to repeat itself in approximately the same form.

SHORTER SECULAR MOVEMENTS AND THE KUZNETS CYCLE

A considerable amount of empirical work during the last few decades suggests that there may well be swings in economic growth averaging perhaps 20 years in duration, or less than half the period assumed for the Kondratieff cycle. The most important work suggesting the existence of these intermediate swings has been done by Arthur Burns, Brinley Thomas, Moses Abramovitz, and particularly Simon Kuznets.[20] Because of the latter's pioneering work, it has been suggested that these swings be called "Kuznets" cycles.[21]

[20] See A. F. Burns, *op. cit.;* Brinley Thomas, *Migration and Economic Growth,* 1954; Moses Abramovitz, *Resource and Output Trends in the United States Since 1870,* National Bureau of Economic Research Occasional Paper 52, 1956, and statement in *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee, Part 2 (86th Congress, 1st Session, 1959); Simon Kuznets, "Long-Term Changes in National Income of the United States . . . ," in Kuznets and Goldsmith, *op. cit.,* pp. 49 ff.; Simon Kuznets and Ernest Rubin, *Immigration and the Foreign Born,* National Bureau of Economic Research Occasional Paper 46, 1954; and Simon Kuznets, *Capital in the American Economy: Its Formation and Financing,* chaps. 2, 7, 8. See also Kuznets' earlier volume, *Secular Movements in Production and Prices, op. cit.*

[21] Cf. W. A. Lewis and P. J. O'Leary, "Secular Swings in Production and Trade, *1870–1913,*" *The Manchester School,* vol. 23, May, 1955, pp. 113–152.

These swings show up primarily as alternations in the *rate of growth,* not in the absolute magnitudes, of such variables as output, population and the labor force, the supply of money, the stock of capital, productivity, and so on. In some highly volatile series, such as building activity, we find swings in the absolute magnitudes themselves.

Thus Arthur Burns, in his careful study of production trends in the United States, found that there were apparently cycles in the percentage rates of growth of industrial production over successive overlapping decades.[22] A common wavelike pattern in percentage rates of growth showed itself in the production of most nonagricultural commodities and in indices of total production. Thus, the decades 1875–1885, 1895–1905, 1910–1920, and 1920–1929 were periods of particularly rapid expansion; the decades 1885–1895, 1905–1915, and 1915–1925 were periods of particularly low rates of growth. Burns did not attempt to date the peaks and troughs in the original data that would correspond to these cycles in rates of growth, and the fact that his data are expressed in terms of decade rates makes precise dating impossible.

Kuznets and Abramovitz have found that these intermediate swings stand out clearly in the growth of GNP and capital formation, in immigration and in the growth of population, and in a number of other variables. Brinley Thomas obtained similar results for the migration of labor and capital from Europe to the United States, and evidence of these intermediate swings has been found in a number of other countries.[23]

Close examination of the intermediate trend in Figure 25 (page 226) permits us to provide approximate dates for these long swings in GNP since 1870. The periods of particularly rapid growth (upswings of the "Kuznets" cycles) came from the mid-1870's to about 1882, from the mid-1890's to about 1907, from about 1916 to the middle or late 1920's, and after the mid-1930's. The periods of retarded growth came from 1882 to around 1895, from 1907 to about 1916, and from the mid or late 1920's to the

[22] Burns, *op. cit.,* chap. 5. Burns expressed his data in the form of average annual rates of change during overlapping decades—i.e., 1870–1880, 1875–1885, 1880–1890, etc.

[23] See the references cited in footnotes 20 and 21.

middle of the 1930's.[24] It is possible that another period of retarded growth came in the latter half of the 1950's.

Economists have not yet developed a generally accepted explanation of these intermediate swings, nor is there full agreement that these swings constitute a separate order of cycles distinct from business cycles. One uncertainty arises from the fact that these "cycles" are obviously related to the severe depressions of the past century. It is not surprising that expansion should be particularly rapid as the economy comes out of a deep depression, and the "downswings" of these long cycles may reflect in part the fact that we have experienced severe depressions. It is significant, however, that in the past, deep depressions have been associated with substantial retardation in the rate of growth of output.

In the past, particularly before the 1920's, some significant sequences seem to have been contained in these intermediate swings. Following a serious depression, growth of output would accelerate; this was usually followed by accelerated population growth and migration (including immigration). This, in turn, helped to bring on an upswing in residential and other population-induced construction and then in other types of capital formation. Eventually retardation in the growth of output would set in, and at some point the building boom would come to an end. And usually a severe depression soon followed.

BUILDING CYCLES

One thing that seems to be clear is that these intermediate movements are related to the wide swings in building activity that have come to be called building cycles. There is widespread (although not universal) agreement that building activity in the past has been characterized by long cycles of 15 to 20 years in duration which

[24] These are roughly the same dates given by Kuznets in *Capital in the American Economy, op. cit.*, chap. 7. See also Abramovitz, *op. cit.*, p. 21. In his more recent work, Abramovitz has dated these intermediate swings back to the early part of the nineteenth century. His dates, following Burns' earlier procedure, are the years of maximum or minimum rate of growth and thus do not correspond, for the period following 1870, to the dates that we have given in the text. Abramovitz also recognizes more intermediate swings than we do, with the result that he gets a shorter average duration than do we. See his testimony before the Joint Economic Committee cited in footnote 20.

have attained wide amplitudes.[25] These building cycles have shown up not only in the United States but also in other countries.[26]

The American economy experienced four of these long building cycles between the Civil War and the Great Depression of the 1930's. (See Table 21.) After 1933, there was another upswing in building activity which reached a peak in 1941, followed by a low in 1944. This trough was obviously due to wartime restrictions on

TABLE 21. Turning Points of Long Cycles in Building, 1853–1933 [a]

Peak	Trough	Duration (in years) Peak to Peak	Trough to Trough
1853	1862		
		18	16
1871	1878		
		19	22
1890	1900		
		19	18
1909	1918		
		16	15
1925	1933		

[a] From A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, 1946, p. 422. By permission of the National Bureau of Economic Research. See also the dates for the cycles in different types of building construction given in C. D. Long, Jr., *Building Cycles and the Theory of Investment*, 1940, pp. 135–136.

building and should perhaps be excluded from consideration.

After 1944, total private building activity (measured in constant prices) expanded almost continuously until 1955 and remained not far from the 1955 level during the remainder of the 1950's. Residential building reached a peak in 1950, in terms of new housing starts, which was virtually duplicated in 1955 and 1959. Dollar expenditures on private residential building (in constant prices)

[25] Cf. Burns and Mitchell, *op. cit.*, p. 418. For more detailed discussion of long cycles in building, see C. D. Long, Jr., *Building Cycles and the Theory of Investment*, 1940, chaps. 7–12; N. J. Silberling, *The Dynamics of Business* 1943, Chap. 9; and R. C. O. Matthews, *The Business Cycle*, 1959, chap. 6. For a skeptical view regarding the existence of long building cycles, see M. L. Colean and Robinson Newcomb, *Stabilizing Construction: The Record and Potential*, 1952, Appendix N. It should be noted that this discussion of building cycles refers only to the construction of buildings and excludes other types of construction such as roads, bridges, etc.

[26] See G. F. Warren and F. A. Pearson, *World Prices and the Building Industry*, 1937, chaps. 6–7.

seemed to have reached a peak in 1955, but this was exceeded in 1959. In general, the expansion in private building activity (residential and nonresidential combined) flattened out during the latter half of the 1950's, but it was not clear at the close of the decade that a peak in a building cycle had been reached. Thus, the building cycle, if it still exists, did not stand out clearly in the first 15 years after World War II.[27] It would be premature, however, to jump to the conclusion that wide swings in building activity are entirely a thing of the past.[28]

The prewar building cycle was much more marked in residential than in nonresidential building, but it did show up in the latter also.[29] Even with respect to residential building, however, there is at least one important reason for taking a second look before jumping to the conclusion that the building industry inevitably generates a regular, long cycle of 15 to 20 years. This is the role that wars have played in stimulating and retarding building activity. Two of the five troughs in Table 21 occurred during major wars, and another trough which is not shown in the table occurred during World War II. In addition, war-created shortages help to explain the booms that followed the Civil War and both world wars.

With this warning in mind, we may formulate our own conclusions regarding the long building cycle as follows. There have been these long swings in (particularly residential) building activity, and in the past the amplitude of these movements has been very wide. Further, the character of the building industry would lead us to expect long cycles of some sort. Buildings are durable; construction in any one year is a small percentage of the total existing stock of buildings; and a considerable period must elapse between

[27] The behavior of residential building in the first postwar decade and its relation to prewar building cycles are dealt with in Leo Grebler, D. M. Blank, and Louis Winnick, *Capital Formation in Residential Real Estate,* 1956, pp. 37–43, 309–311.

[28] A number of factors have operated to sustain residential building: the high rate of family formation, continued stimulation by the federal government, the higher rate of demolitions in the last few years, and the continued migration to the suburbs. These and other factors have operated to sustain private nonresidential building—including the migration to the suburbs (with the attendant need for stores and other commercial buildings), a new wave of office building (the first since the 1920's), a good deal of decentralization of industry, the general expansion in output, and accelerated technological change.

[29] Cf. Long, *op. cit.,* pp. 130–131.

the initial decision to build and final occupancy of the completed structure. In addition, the financial organization of the industry has been such as to encourage speculative building and, in the past, much unwise lending. As a result of these factors, any strong stimulus (such as a wave of immigration or a war-created housing shortage) will lead to a boom of considerable duration. A number of years may pass before it becomes evident that the boom has gone too far. Once a decline begins because of an excess stock of buildings, construction can fall to very low levels and remain depressed for a long period. Gradually, the continued growth of population and the slow wearing out of old buildings will create a situation favorable to a new upswing.

This is the sort of response mechanism that tends to create long building cycles. In addition, however, we must take account of the external forces that impinge on this response pattern. Of these, the most important are wars and changes in the rate of population growth. Government intervention may also be an important factor. These stimuli and disturbances do not necessarily occur at fixed intervals, and therefore we need not expect that building cycles will always be of approximately the same duration. In the absence of wars, building cycles would probably be milder and also perhaps longer than those we have experienced in the past. A variety of external forces and institutional changes operated to create an unusually long building cycle in the period following World War II.

In view of the size of the building industry and the amplitude of the long cycles in building, it would not be surprising if we found somewhat corresponding movements in the course of general business activity. As we saw on page 243, there does seem to be a relation between the swings in the rate of growth of output and building cycles. In the past, when the growth in output accelerated, so did population growth; and a boom in population-induced investment followed. Significantly, also, the peaks of building cycles antedated by a few years the severe depressions of the 1870's, 1890's and 1930's.[30]

[30] See, for example, Kuznets, *Capital in the American Economy, op. cit.*, chap. 7; also Matthews, *op. cit.*, p. 212. Walter Isard has argued that there is a relation between important transportation developments, building activity, and resulting long cycles in business. See the *Review of Economics and Statistics,* vol. 32, November, 1950, pp. 347–351, and his earlier articles cited in this source.

THE NATURE OF BUSINESS-CYCLE FLUCTUATIONS

In addition to experiencing seasonal variation and long-term secular movements, economic activity moves continuously through alternations of rise and fall that we call business cycles. If the secular movements are "eliminated" from seasonally adjusted data, these cyclical movements will show up as wavelike fluctuations around a horizontal line (as in the lower section of Figure 29 on page 255).

If we examine these cyclical fluctuations in important monthly indicators of business activity, we note immediately that the cyclical pattern is anything but smooth. Irregular month-to-month fluctuations seem to be superimposed on the underlying sweep of the business cycle; and on occasion these brief interruptions in the general cyclical pattern can be quite marked. Removal of secular movements and seasonal variation, then, leaves us with two types of short-run fluctuations to discuss—business cycles and what we may call "irregular fluctuations."

IRREGULAR FLUCTUATIONS

The concept of business cycles apart from irregular fluctuations is in part an abstraction.[31] There is no reason to believe that the cyclical forces at work operate to create a perfectly smooth path of expansion and contraction and that, therefore, the month-to-month irregularities in the pattern of short-term fluctuations are entirely the result of noncyclical forces. In all probability, the business cycle should be considered as including many of these very short-term irregular movements. The cyclical forces at work create the latter at the same time that they generate the underlying cumulative swings that we have chiefly in mind when we speak of business cycles.

These short-term irregular movements are primarily of two types. There are, first, what we may call minor random movements. These relatively unimportant fluctuations occur so frequently that they give a slight "saw-tooth" effect to the cyclical movement in most

[31] Cf. the discussion of irregular fluctuations in W. C. Mitchell, *Business Cycles: The Problem and Its Setting*, 1927, pp. 249 ff. For an attempt to develop a statistical method to iron out irregular fluctuations, see Julius Shiskin, *Electronic Computers and Business Indicators*, National Bureau of Economic Research Occasional Paper 57, 1957.

monthly series. They may arise for many different reasons, and frequently the causes are not known. At any rate, these random movements are relatively so unimportant and occur so frequently that it is not worthwhile compiling a list of their possible causes. It is best to consider them merely as part of the total cyclical swing that we observe in practice.

The second type of irregular fluctuation is more important. Included here are the sharper breaks in the underlying cyclical sweep, which can usually be traced to specific causes. We may call these sporadic changes. Examples of such sporadic movements would be a sharp temporary drop in steel production due to a strike; the effect of an earthquake, flood, or other natural catastrophe on a variety of series; a temporary spurt in prices due to a war scare; or a sudden drop in various monetary series resulting from a temporary closing of the banks (such as occurred in 1933).

What is the dividing line between these very short-term sporadic movements and the fluctuations we call business cycles? Both types involve a rise and fall in business activity, and these sporadic changes may be of considerable magnitude. The answer rests on two characteristics of the business cycle, one basic and the other merely an empirical rule of thumb. It is of the essence of business-cycle fluctuations that they are cumulative or self-reinforcing in nature. A movement in a given direction feeds on itself and creates further movement in the same direction. Therefore, any short-term change that is not cumulative—that is immediately followed by a reverse movement back toward the previous level—is considered an irregular fluctuation. Some irregular movements, however, are cumulative for a short period. A war scare, for example, may cause a short-lived speculative boom that feeds on itself for several months before it collapses. In this case, we fall back on our empirical rule of thumb. To be business cycles, fluctuations not only must be cumulative but the period of rise and fall must cover some minimum period of time. The National Bureau of Economic Research, in formulating its definition of business cycles, sets this minimum period as more than one year.

Although it is important to note that irregular fluctuations exist, and to have a working rule for distinguishing the more marked of such fluctuations from business cycles, there is no satisfactory way

of removing these random and sporadic movements from a series.[32] Even if there were, it would not ordinarily be desirable to do so. Dynamic forces do not operate smoothly to create business cycles that look like sine curves, and the forces generating the irregular movements are part of a total composite of forces that create the ragged cyclical contours we experience in practice. The chief reason for discussing irregular fluctuations is to emphasize the fact that not every combination of rise and fall constitutes a business cycle and to insure that we mark off those particular alternations of expansion and contraction that reveal the underlying cyclical forces that are at work.

BUSINESS CYCLES

How, then, shall we describe those fluctuations that we do want to call business cycles? Business cycles consist of recurring alternations of expansion and contraction in aggregate economic activity, the alternating movements in each direction being self-reinforcing and pervading virtually all parts of the economy. Let us look at this definition further.[33]

We speak of recurring alternations of expansion and contraction —i.e., of cycles in economic activity. The economy seems to be incapable of remaining on an even keel, and periods of expanding activity always and all too soon give way to declining production and employment. Further, and this is the essence of the problem, each upswing or downswing is self-reinforcing. It feeds on itself and creates further movement in the same direction; once begun, it persists in a given direction until forces accumulate to reverse the direction. Finally, business cycles are pervasive in their effects. They

[32] Moving averages are sometimes used for this purpose, particularly if the irregular movements are very marked. See the reference to Shiskin in the preceding footnote.

[33] The reader may be interested in comparing this definition with that advanced by the National Bureau of Economic Research: "Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own." (Burns and Mitchell, *op. cit.*, p. 3.)

show up as approximately concurrent movements in many different series; they affect virtually all industries and all parts of the country; and they tend to spread internationally, so that periods of prosperity and depression have tended to occur at about the same time in capitalist countries bound together by trading relationships.

As a working rule, we may also specify that business cycles generally last from two to eleven years, counting from peak to peak or from trough to trough. While these time limits are empirical and rule of thumb, they are apparently related to the dynamic processes that create business cycles. Fluctuations must be cumulative to be business cycles; and the development of a cumulative movement, either upward or downward, takes time. Hence a full cycle is not likely to be of less than two years' duration, although on occasion we may want to recognize a cyclical movement that lasts a few months less than this. At the other extreme, the elements of instability (and of recuperation) inherent in the economy seem to prevent complete business cycles from lasting longer than 10 or 11 years (and usually much less than this).

Although business cycles represent *recurring* alternations of prosperity and depression, virtually all authorities are agreed that there is nothing *periodic* about these movements. There is no evidence that business cycles tend to recur over and over again in virtually the same form, with the same duration and the same amplitude of movement. Some cycles are mild and others severe; some last two or three or four years, and others eight or ten years. In some the upswing is longer than the downswing; in others the reverse is the case.

In this connection, it is interesting to examine in Table 22 the durations of American business cycles that have been identified by the National Bureau of Economic Research. Over the period covered in the table, business cycles have ranged in duration from 28 months (1919–1921) to 99 months (1870–1879). Only during the period 1885–1914 was there any marked tendency for the durations of business cycles to cluster around some typical period (about 40 months), and even during these years successive cycles differed significantly from each other in the relative lengths of their expansion and contraction phases. The last five cycles shown in Table 22 varied in length from 44 months (1954–1958) to 88 months (1938–1945). In all of these five, the expansion phase lasted

much longer than the contraction, but they were immediately preceded by the cycle of 1927–1933, in which the contrary was true.

The degree of diversity between successive business cycles is

TABLE 22. Turning Points and Durations of American Business Cycles, 1854–1958[a]

Dates of Turning Points		Duration in Months		
Peak	Trough	Expansion[b]	Contraction[c]	Full Cycle
—	Dec., 1854	—	—	—
June, 1857	Dec., 1858	30	18	48
Oct., 1860	June, 1861	22	8	30
Apr., 1865	Dec., 1867	46	32	78
June, 1869	Dec., 1870	18	18	36
Oct., 1873	Mar., 1879	34	65	99
Mar., 1882	May, 1885	36	38	74
Mar., 1887	April, 1888	22	13	35
July, 1890	May, 1891	27	10	37
Jan., 1893	June, 1894	20	17	37
Dec., 1895	June, 1897	18	18	36
June, 1899	Dec., 1900	24	18	42
Sept., 1902	Aug., 1904	21	23	44
May, 1907	June, 1908	33	13	46
Jan., 1910	Jan., 1912	19	24	43
Jan., 1913	Dec., 1914	12	23	35
Aug., 1918	Mar., 1919	44	7	51
Jan., 1920	July, 1921	10	18	28
May, 1923	July, 1924	22	14	36
Oct., 1926	Nov., 1927	27	13	40
Aug., 1929	Mar., 1933	21	43	64
May, 1937	June, 1938	50	13	63
Feb., 1945	Oct., 1945	80	8	88
Nov., 1948	Oct., 1949	37	11	48
July, 1953	Aug., 1954	45	13	58
July, 1957	April, 1958	35	9	44

[a] Dates are taken from G. H. Moore (ed.), *Business Cycle Indicators*, vol. 1, 1960, p. 670. Reproduced by permission of the National Bureau of Economic Research and Princeton University Press.

[b] Measured from trough on preceding line to peak.

[c] From peak to trough on same line.

greater than Table 22 suggests. Cycles differ in amplitude as well as in duration, and the interrelationships among different types of economic activity vary from cycle to cycle. In addition, there is some evidence that the American economy has experienced at least

two types of business cycles. These two types may be called major and minor cycles. Though no precise line can be drawn between them, the former tend to be relatively long (six to eleven years), to have a relatively wide amplitude, and to be associated with important changes in business expectations regarding the profitability of long-term investment. The minor cycles are shorter (usually two to four years) and of smaller amplitude; and they tend to be associated with changes in short-term business expectations and with minor maladjustments, but not with significant alterations in long-term investment opportunities.[34]

Sometimes these two types of cycles occur simultaneously. Thus, the course of business activity in the United States between 1921 and 1933 seems to trace out a major cycle covering the entire period, on which were superimposed minor cycles with peaks in 1923 and 1926 and troughs in 1924 and 1927. At other times only the minor cycles seem to show up, as during the 1950's; and on yet other occasions, only the major cycle stands out. Table 22 marks every peak and trough, whether they are associated with major or minor cycles. As a result, the cycle durations shown reflect in part the different combinations of major and minor cycles that may have occurred at various times in the past.[35]

It should be noted that we include *all* alternations of expansion and contraction in our concept of business cycles, provided these fluctuations are cumulative in nature, affect most of the economy, and (as a rule of thumb) have a duration of at least one and preferably two years. Thus a war boom followed by a postwar depression constitutes a business cycle, even though there is a definite, noneconomic cause which obviously is chiefly responsible for initiating the cyclical movement. The important point here is that a cumulative movement results and represents the operation of the

[34] The distinction between major and minor cycles is developed further in Chapters 10 and 11. The reader should be warned, however, that not all economists accept this distinction. For example, such impressive authorities as Burns and Mitchell hold that the available statistical evidence does not thus far clearly establish the existence of major cycles distinct from minor cycles. See *Measuring Business Cycles,* pp. 440–465. Hansen, on the other hand, emphasizes this distinction, as do a number of other economists.

[35] Ruth Mack has suggested that in addition to business cycles (major and minor), the American economy may also have experienced "subcycles" with an average duration of about 18 months. "Notes on Subcycles in Theory and Practice," *American Economic Review,* vol. 47, May, 1957, pp. 161–174.

same forces as those that create peacetime cyclical movements. Or, to put the matter another way, business cycles result from the way the economy typically reacts to a variety of possible initiating forces. Our concept of business cycles rests not on the nature of the initiating causes (which may or may not be inherent in the operation of the economic system) but on the nature of the economy's responses to these stimuli. It is these responses, arising out of the characteristics of a private-enterprise economy discussed in Chapter 8, that magnify the initial stimuli into the cumulative movements we call business cycles.

Strictly speaking, *business cycles* is a generic term referring to a certain type of fluctuation in *total* economic activity. The cycles we mark off in general business activity will be reflected to varying degrees in the movement of individual series representing the different types of activity that make up the total business situation. The cycles in these different series have a basic similarity, but there will also be significant differences in the cyclical behavior of different series. Thus, the cyclical fluctuations in steel production or in the output of durable goods generally have a much wider amplitude than do the cyclical movements in the relatively stable industries producing nondurable consumers' goods. The cycles in different economic series vary not only in amplitude but also in timing; they do not all reach their peaks and troughs at the same time. As we shall see in greater detail in the next chapter, these differences reflect the varying responses of the different parts of the economy to common cyclical forces. In addition, some sectors of the economy may experience cycles that are peculiar to those areas and do not show up in other lines of activity at all.

BUSINESS CYCLES AS DEVIATIONS FROM "NORMAL" OR "EQUILIBRIUM"

No completely satisfactory technique exists for separating cyclical (plus irregular) fluctuations from the other types of economic change. We have discussed the reasons for this at various points in the preceding sections. The basic difficulty lies in the interdependence of cyclical and secular movements. Cyclical forces do influence trends, and, in turn, the kind of trend we choose to eliminate helps to determine the shape of the resulting cyclical fluctuations. This problem is particularly difficult if intermediate trends exist in our

data. Judgments as to the nature of these intermediate secular movements must of necessity be arbitrary to some extent; and when we eliminate these trends, we run the risk of distorting the shorter-run cyclical movements.

Despite these difficulties, the conventional method of isolating the cyclical-irregular fluctuations in a series may frequently be used. If we have a long enough period to determine a satisfactory approximation of the trend, and if intermediate secular movements do not raise insoluble questions as to what are cyclical and what are secular movements, the method may be used to good advantage to bring the cyclical fluctuations into sharp relief. Put briefly, the technique consists of dividing the seasonally adjusted data by the corresponding computed trend values.[36] Where we deal with a relatively short period, however, it is better merely to eliminate the seasonal variation and not to try to separate the secular and cyclical movements.

Figure 29 provides an example of conventional trend elimination. The upper chart shows the American Telephone and Telegraph Index of Industrial Activity uncorrected for trend. The influence of long-run growth on the index is clear. The lower chart presents the index in form of percentage deviations from a computed trend, so that the series now fluctuates around a horizontal instead of an upward-sloping line.

If we do "eliminate" the seasonal and secular movements, we have left the cyclical-irregular fluctuations expressed as percentages of the combined secular-seasonal movement. It is sometimes said that these residual cyclical movements represent deviations from "normal," or, sometimes, that they represent departures from the equilibrium position portrayed by the secular-seasonal movement that has been eliminated. Actually, there is little justification for such statements. The "base line" from which we measure cyclical fluctuations by conventional methods is not a path of equilibrium,

[36] The details of the method are described in any good text on economic statistics. Though frequent dissatisfaction is expressed with the conventional method of trend-cycle separation, no other convenient method of isolating cyclical-irregular fluctuations is available. Edwin Frickey has devised a method of decomposing time series which reverses the usual procedure and eliminates the cyclical movement in order to determine the trend. The method is laborious, however, and cannot be applied to all series. (See his *Economic Fluctuations in the United States.*) The analytical techniques devised by the National Bureau of Economic Research do not aim at complete separation of trend and cycle.

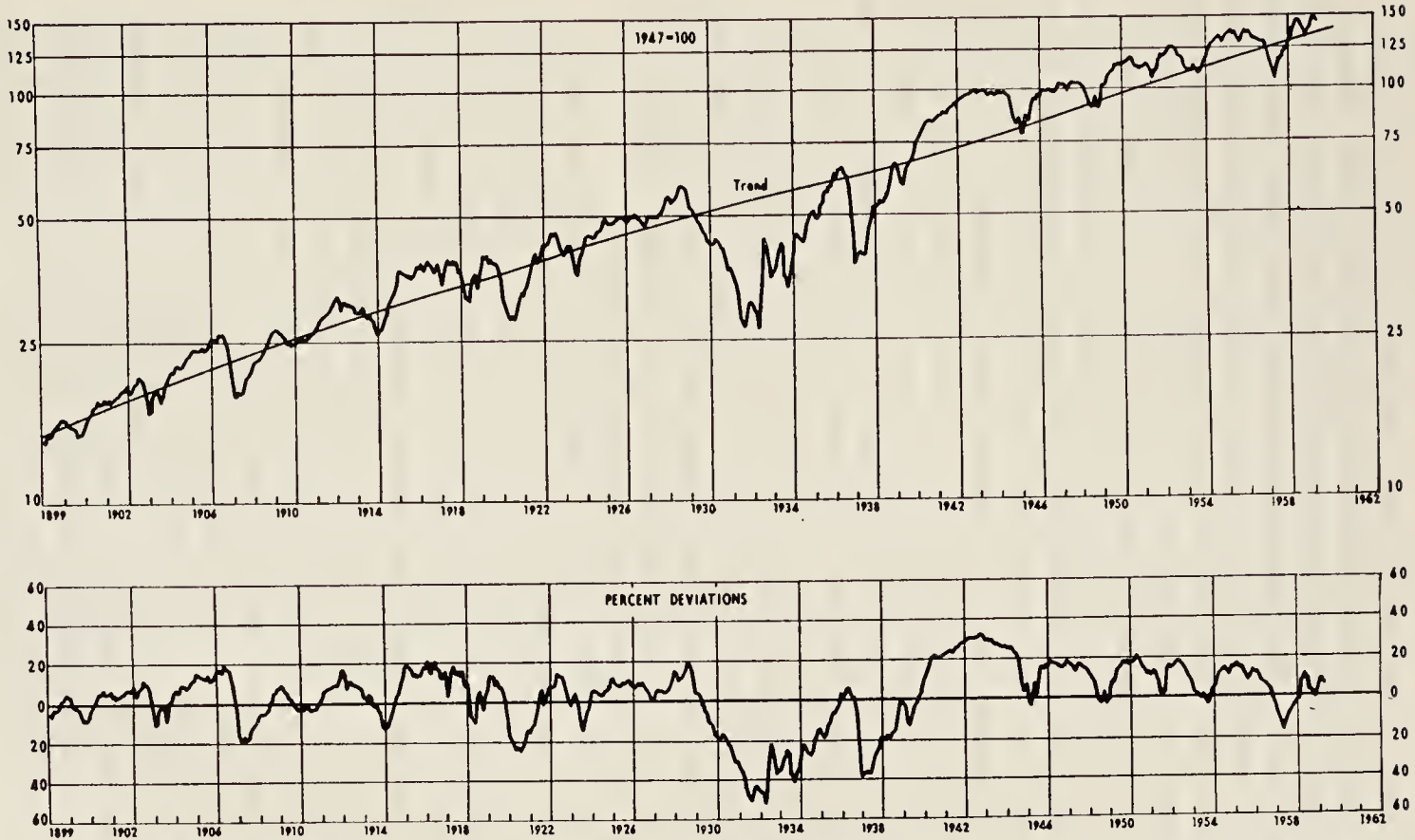


FIGURE 29. Index of Industrial Activity in the United States, 1899–1960, Before and After Correction for Trend.

See p. 174 for a description of this index. Reproduced by permission of American Telephone and Telegraph Company.

and in only a limited sense does it represent a "normal" position.

There is no general agreement as to what we mean by "normal" business conditions. Whatever our concept of "normal," there is little reason to think of the secular trend, from which we measure cyclical deviations, as representing a normal situation. As we have noted, cyclical influences do affect the trend, which represents our judgment as to the nature of the long-run movement resulting from *all* the dynamic forces at work. It is true that trends are computed so that they cut through the short-run fluctuations, and in a rough sense they therefore may be taken as representing a sort of moving average of these fluctuations. In this limited and purely statistical sense, there may be some value in looking on the trend (adjusted for seasonal variation) as a sort of normal. It is an average position, neither markedly good nor markedly bad, which allows for secular and seasonal influences. But the trend is *not* the product of any peculiar set of normal or equilibrium conditions; it does not tell us what would have happened if cyclical and accidental forces had not been at work. To avoid any misconceptions, therefore, it is better not to think of business cycles as fluctuating around any "normal" level, though we may grant that businessmen do have a rough-and-ready idea of "normal business" with which they compare current cyclical fluctuations.

There is no justification for regarding the secular movement as a path of moving equilibrium, around which cyclical fluctuations take place. An equilibrium position implies a position of rest, a balance of forces, such that no further change would take place in the absence of new disturbances. Does the secular trend trace out such a moving equilibrium position, portraying how the economy would behave in the absence of the forces making for cyclical instability? The answer is clearly in the negative. Since trends result from the action of secular forces in a world that is cyclically unstable, we do not know what the secular movement in particular series would have looked like in the absence of cyclical change. The present structure of the economy is a product of, among other things, past cyclical instability and the expectation that fluctuations will continue in the future. Computed trends do not tell us how economic series would have behaved in the absence of cyclical disturbances. If cyclical change were to cease, a projection of past trends would certainly not tell us at what level total output or the

production of particular industries would come to rest. It is therefore not correct to think of the trend, or of any particular position between cyclical low and high points, as representing an equilibrium position which would prevail in the absence of those forces that create business cycles.

CHAPTER 10

HOW THE ECONOMY BEHAVES DURING BUSINESS CYCLES: THE STATISTICAL RECORD

WE ARE now ready to attack directly the main task of this book —to describe how and why the economy behaves as it does during business cycles. A few words may be in order as to why it has taken us so long to reach this point. In a field as complicated as ours, we need tools of analysis and working hypotheses to apply to an otherwise unmanageable mass of facts. The preceding chapters all help to fill our toolbox. They enable us to ask the right questions, to look for the most fruitful types of information, and to put the pieces together into a revealing picture of how the economy generates booms and depressions.

We shall now try to see just what does happen during business cycles. In this chapter we look at the statistical record, and in Chapter 11 we shall pull the threads together into a generalized description of how the economy expands and contracts.

THE PHASES OF THE CYCLE

It is not easy to mark off with any precision a series of phases through which all business cycles are supposed to pass. The difficulties inherent in applying the usual four-phase scheme to business cycles (i.e., revival, prosperity, recession, depression) are readily apparent if we examine the various business indicators charted throughout this book. The cyclical contours of these series are highly irregular, and in none of the cycles is there a clear dividing

line between revival and prosperity or between recession and depression.[1]

Despite these irregularities in cyclical behavior, one simple way of dividing business cycles into phases suggests itself. We have defined business cycles as recurring alternations of expansion and contraction in aggregate economic activity, the movements in each direction being self-reinforcing. Thus, the entire cumulative movement in each direction makes a logical phase to consider, and this gives us the two stages of expansion and contraction. In addition, we have the turning points, where expansion turns into contraction and vice versa. Actually, it is better to speak of *zones* rather than turning *points*. In the neighborhood of the point at which the turn comes, both before and after the statistically determined peak or trough, critical changes occur which seal the fate of the cumulative movement that is ending and bring on the ensuing contraction or expansion. We shall, then, examine cyclical behavior in terms of these four phases: expansion, the critical zone surrounding the upper turning point, the contraction, and the critical zone in the neighborhood of the lower turning point.

SIGNIFICANT DIFFERENCES AMONG BUSINESS CYCLES

We have already emphasized that every business cycle differs in important respects from all other cycles. More than sheer perversity leads us to develop this theme further here and to wait until later to consider the characteristics which all cycles have in common. After all, observation tells us that business cycles do differ from each other; and, if we probe beneath the surface, we find that the combinations of causes responsible for cyclical turning points have varied also. We must give up the idea immediately that cyclical behavior is sufficiently regular for us to build a single, general explanation that will account for what happens during all business cycles.

[1] Such a dividing line can be constructed by a procedure we rejected in Chapter 9—by drawing in a trend to represent the "equilibrium" or "normal" level. The expansion phase below the trend line would be revival; that above the trend would be prosperity. The downswing could be similarly divided. This is what Schumpeter does in his *Business Cycles* (vol. 1, pp. 145 ff.), though his concept of the equilibrium position cannot be found simply by fitting trends by the usual procedure.

DIFFERENCES IN DURATION

Let us pursue this theme of diversity further by looking at the duration and amplitude of the various cycles through which American economic activity has passed during the last century or so. First, consider the duration of past business cycles in the United States. (See the table on page 251.) Since 1854, the business cycles marked off by the National Bureau have varied in length between 28 months (1919–1921) and 99 months (1870–1879). As noted in Chapter 9, there was a tendency toward uniformity in duration between 1885 and 1914, during which period no cycle was shorter than 35 or longer than 46 months. Even during this period, however, business cycles varied considerably in the length of their expansion and contraction phases.

Considerable variability in duration characterized American business cycles between 1914 and 1945, chiefly because of the influence of two world wars and the long depression of the 1930's. The three cycles after 1945 listed in Table 22 show a rough uniformity in duration, with expansions lasting from 35 to 45 months and with brief contractions ranging from 9 to 13 months. This obviously does not mean, however, that the expansion and contraction phases of all future cycles will fall within these relatively narrow ranges.[2]

VARIATION IN AMPLITUDE

Business cycles differ in amplitude as well as duration, and this brings us to an important matter that was first introduced in Chapter 9—the distinction betwen minor and major cycles. Figure 30, on which are plotted four measures of the physical volume of production and trade for varying periods before World War II, is of interest in this connection. In the case of each series, a three-year moving average has been drawn through the original annual data. The original series emphasize particularly the shorter cycles, usually between three and four years in duration, which were characteristic of the

[2] It is perhaps worth noting that the average duration of the three cycles after World War II listed in Table 3 was considerably greater (50 months) than that of the first three cycles after World War I (35 months). The difference lay in the much longer expansions after World War II. As this book was going to press, it appeared that another cyclical peak might have occurred during 1960. If so, the expansion of 1958–1960 would have been significantly shorter than any of the three preceding postwar expansions.

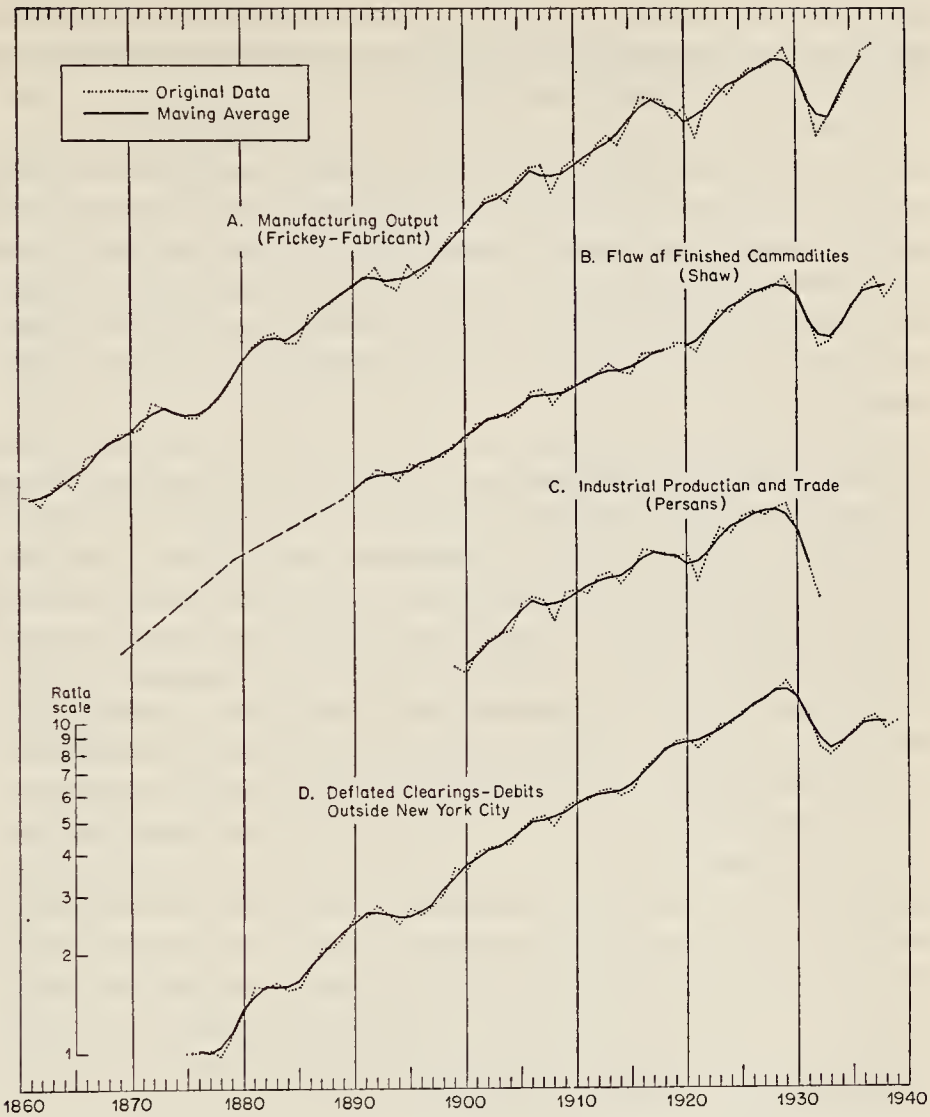


FIGURE 30. Measures of the Physical Volume of Business, with Three-Year Moving Averages.

From *American Economic Review*, vol. 39, May, 1949, suppt., p. 58, by permission.

American economy, at least before 1929.[3] The moving averages seem to trace out an underlying swing, irregular in duration and amplitude, which in Chapter 9 we called the major cycle. Major cycles have been featured by long upswings followed by deep depressions.

[3] Business cycles in the western European countries—for example, England, France, and Germany—have typically been of longer duration. Cf. A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, 1946, pp. 78–79.

Sometimes the major cycles stand out prominently—1878–1885, 1885–1896, and especially 1921–1933. At other times, the pattern of the major cycle is much less clear.

We thus have at least two categories of cyclical fluctuations to explain: the frequent short cycles and the longer major swings. When we get downturns in business, why are they sometimes short and mild and sometimes long and severe? Why do some short cycles have more pronounced expansions than contractions, whereas contraction is the more prominent phase in other minor cycles? The distinction between major and minor cycles should help us to answer these questions.

Figure 30 suggests that major and minor cycles do not always occur together in the same combination. No minor cycles occurred during the major swing of 1878–1885 or during the major downswings of 1873–1878 or 1929–1933. During the major cycle of 1885–1896, however, and again during 1896–1914 and 1921–1929, short cycles of three to four years followed each other with remarkable regularity. Between 1900 and 1914, the minor cycles stand out more prominently than does the attenuated underlying swing. During 1921–1933, minor cycles are superimposed on at least the expansion phase of the major cycle, but the major swing dominates the picture.

The United States has not experienced a major depression since the 1930's. Since World War II, we have had only minor cycles, which have been marked by moderately long periods of expansion and by very brief and mild contractions. It may be that the distinction between major and minor cycles is losing much of its significance in an economy that is now much more stable than it was a generation or more ago and in which the government plays a far more important role than formerly.[4]

DIFFERENCES IN CAUSATION

These differences in cyclical behavior reflect the changing character of the destabilizing forces that create business cycles. Major downswings have usually followed periods of excessive investment in capital goods, substantial security speculation, and weakening of the banking system through the expansion of credit based on inflated

[4] R. C. O. Matthews suggests that what we have called the major cycle is primarily a reflection of the building cycle and that probably only minor cycles and building cycles have to be recognized. *The Business Cycle,* 1959, chap. 12.

security values. The resulting temporary saturation of investment opportunities and weakening of the banking system lead to contractions that are not easy to reverse. On the other hand, minor cycles tend to be associated with less important and more short-run maladjustments that the economy can much more easily overcome.[5]

Major cycles themselves follow different patterns, and the same is true of minor cycles. Thus, the "crises" of 1873, 1893, and 1907 were accompanied by financial panics; but this was not true of the turning points in 1882, 1920, or 1929. In the major downswings of 1882–1885 and 1929–1933, the banking system collapsed in deep depression rather than at the beginning of the downswing. Similar contrasts stand out if we examine the three most important declines occurring since World War I—namely, those of 1920–1921, 1929–1933, and 1937–1938. The character of the preceding boom was different in the three cases; the part played by inventory accumulation, stock-market speculation, and banking developments was different; so, too, was the role played by international trade; and so on. Similar though perhaps less pronounced differences show up when we compare the minor peaks in 1923 and 1926 or those that occurred after World War II.[6]

Thus, the degree and kind of vulnerability of the economy to deflationary forces has been different at different cyclical peaks. And the initial impetus that has uncovered the underlying weaknesses and started the cumulative spiral downward has also differed from cycle to cycle. Similar contrasts stand out if we consider lower turning points, though it is less easy to spot the immediate causes of revivals than of downturns. Depressions develop differently, depending on the character of the preceding expansion and the course of liquidation during the downswing. Sometimes external events clearly enter to help speed a revival. In some cases, the need to replace depleted inventories is enough to stimulate revival; in other cases,

[5] Elsewhere I have suggested that some severe but relatively short depressions might be called "hybrid" contractions, intermediate between the longer and severe major depressions and the mild and brief minor recessions. The recessions of 1907, 1937–1938, and (with some qualifications) 1920–1921 would fall into this intermediate category. See R. A. Gordon, "Investment Behavior and Business Cycles," *Review of Economics and Statistics,* vol. 37, February, 1955, pp. 23–34.

[6] Chapters 14–16 examine the business-cycle history of the period since World War I in considerable detail. For further discussion of the major and minor cycles that have occurred in the United States since the Civil War, see A. H. Hansen, *Business Cycles and National Income,* 1951, chap. 2.

recovery waits on an improvement in long-term investment opportunities. And once revival begins, a host of factors help to determine the character of the expansion—the relative roles of long-term investment and inventory accumulation, which industries lead in the expansion, the extent of speculation in commodities and securities, and so on.

DIFFERENCES IN TIMING

If the economy always reacted in the same way to the same set of initiating forces, we should expect that the timing relationships among important economic series would be the same at successive peaks and troughs. The available evidence indicates that some series do tend fairly regularly to lead at the turning points and others to lag. These leads and lags are a part of the response mechanism through which the economy reacts to initiating forces making for either expansion or contraction, and they also throw light on what the initiating forces are. We shall look into this matter in more detail later in the chapter.

What we want to emphasize here is that timing relationships are not constant from cycle to cycle. When we speak about characteristic leads and lags, we are talking about *averages*. Wherever we find a series that typically leads at one or both turning points, we are almost certain to find that, over a number of cycles, there is a considerable dispersion around the average figure. It is well known, for example, that industrial stock prices tend to lead general business at both the peak and the trough. On the average, over a period of nearly 90 years spanning 21 business cycles, stock prices have begun to decline about three or four months before the cyclical peak in business activity. But this lead has varied from 30 months to a month or two, and in a few cases stock prices actually lagged. Similar or even greater variability shows up in other series also. In this connection, Table 37 on page 516, which summarizes the behavior of a number of "economic indicators" at cyclical turning points, deserves careful study.

Just as a single series will vary its timing over a succession of cycles, so the pattern of timing relationships among any large group of series will differ somewhat from cycle to cycle. The pattern at the peak in 1929 was not precisely the same as that in 1920; the sequence of leads and lags at the low point in 1958 looks different from that in 1949; and so on. Yet there are also regularities in behavior, which re-

flect the tendency of the economy to respond in somewhat the same way to different sorts of destabilizing forces.[7]

We turn now to a study of the features that different cycles do have in common. Having made our point regarding diversity, we are now in a position to see what can be said about the typical behavior of the economy during business cycles. This raises such questions as the following: How do different kinds of economic series typically behave during business cycles? Which regularly show the widest amplitude of movement? Which tend to lead at the turning points? What else of significance can we say about the typical behavior of important economic series?

THE NATIONAL BUREAU'S METHOD OF MEASURING TYPICAL CYCLICAL BEHAVIOR

By far the most exhaustive investigation of typical cyclical patterns has been made by the National Bureau of Economic Research. Before studying the results of the National Bureau's work, however, we must learn how it measures business cycles.[8]

Its first step is to mark off the turning-point dates in general business activity. These reference dates, which we have already presented on page 251, mark off the expansions and contractions in general business. These reference dates are then used to divide every series studied by the Bureau into "reference cycles." Each such reference cycle portrays the behavior of a single series during one particular cycle in general business activity. If the actual troughs and peaks of the series in question differ from the reference dates for general business, it is still the latter that determine the reference cycles into which the series is divided.

This procedure is illustrated in Figure 31. The left-hand panel of this chart shows separately, one above the other, the three "reference cycles" in the Federal Reserve Index of Industrial Production between 1945 and 1958. The original series for industrial production for the postwar years is shown in Figure 40 on page 452. All that we have done is to cut up this series, for the period 1945–1958, into three segments according to the turning-point reference dates shown on

[7] For an interesting case study in this connection, see G. H. Moore, "The 1957–58 Business Contraction: New Model or Old?" *American Economic Review*, vol. 49, May, 1959, pp. 292–308.

[8] For a more detailed description, see Burns and Mitchell, *op. cit.*, chap. 2.

page 251. These dates would be the same no matter what the series studied, and regardless of whether the series in question had its own turning points at these dates or not. Thus, in the first postwar reference cycle in the left-hand panel of Figure 31, we see that industrial

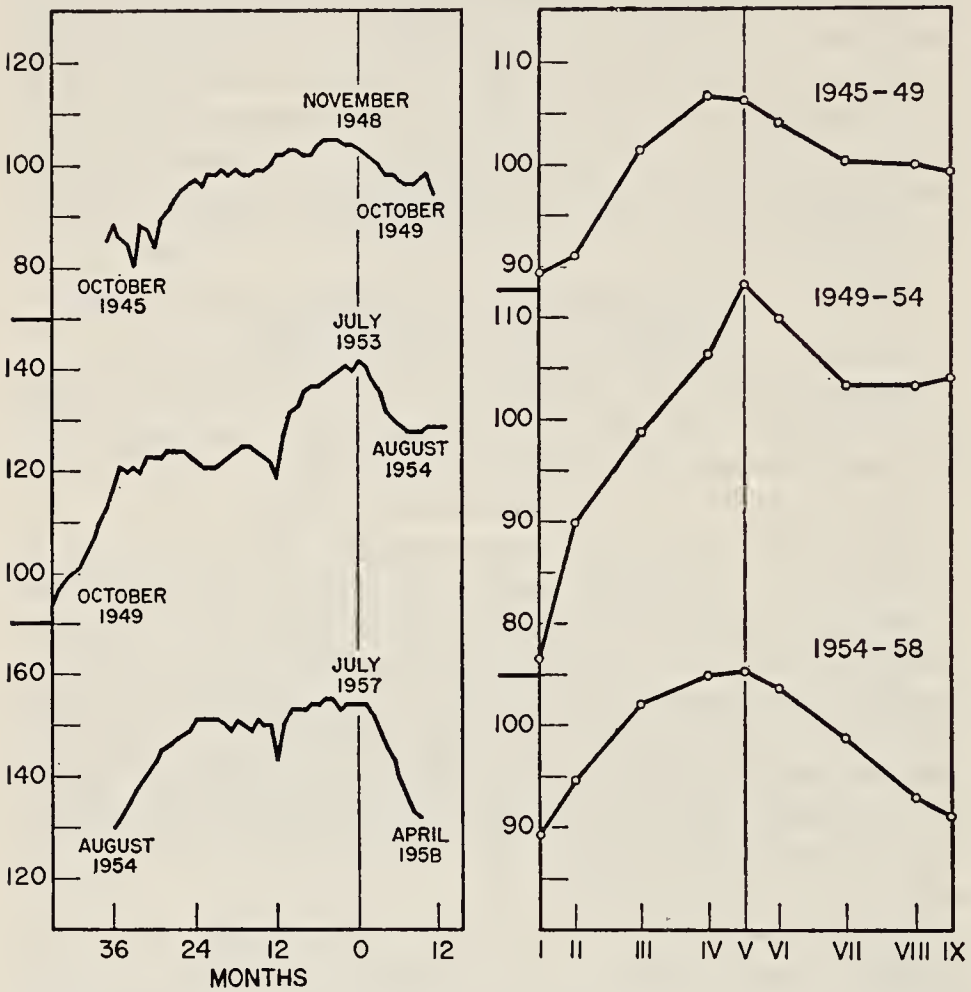


FIGURE 31. Postwar Reference Cycles in Industrial Production, 1945–1958.

The original series is the Federal Reserve Index of Industrial Production. Reference-cycle relatives were computed by the National Bureau of Economic Research.

production continued to decline for several months after the initial reference trough of October, 1945. We also see that this series turned down in 1948 several months before the reference peak—that is, before business generally turned down.

This suggests one of the advantages of this sort of reference-cycle

analysis. By breaking a series up into segments corresponding to the reference-cycle chronology (i.e., according to the turning-point dates for general business), we can see how any particular series behaves when the economy as a whole is expanding or contracting. We can also look to see if the series in question leads or lags at the turning points (i.e., reference dates) in general business.

Having marked off the separate reference cycles for a series in the way just described, the remainder of the National Bureau's procedure consists in good part of converting these reference cycles into a standard form that will facilitate comparisons and then computing a set of average measures for the reference cycles in any particular series.[9]

The first step is to divide the original data (of the series being studied) in each reference cycle by the average for that reference cycle. The resulting "reference-cycle relatives" thus express the data as percentages of the average for the reference cycle in which they fall. Thus, to go back to Figure 31, the average for the index of industrial production in the 1945–1949 reference cycle is 97.1 (with 1947–1949 = 100). All the monthly figures for industrial production in this cycle are divided by this average, which thus becomes the base of 100 for this cycle. The average for the 1949–1954 cycle is 124.7; this is used as the base for this cycle; and so on.

Now we are ready to move to the right-hand panel of Figure 31. Each reference cycle is divided into nine stages. Stage I covers the three months centered on the trough preceding the upswing; stage V covers the the three months centered on the peak; and stage IX covers the three months centered on the trough ending the cycle. These, then, are the turning points. Then the period of expansion is divided into three stages of equal length (stages II–IV), and the period of contraction is similarly divided into thirds (stages VI–VIII). The monthly cycle relatives are then averaged for each stage of each cycle. We now have a series of reference-cycle patterns, each showing

[9] The National Bureau also computes "specific cycles" in a similar way, but in this case the actual turning points in industrial production (or in any other series being studied) would be used to mark off the expansions and contractions. Thus the first postwar *specific* cycle in industrial production would show an initial trough in February, 1946, a peak around August, 1948, and a final trough around July, 1949. (Compare the corresponding reference-cycle dates.) The dates of specific cycles depend on the behavior of the series in question. The dates of a particular reference cycle are the same for all series, since they depend on a single set of turning points in general business activity.

the behavior of some series through the nine stages of each reference cycle covered. Stage I covers the beginning trough in general business; II–IV, the upswing; V, the peak; VI–VIII, the downswing; and IX, the ending trough.[10]

All this is illustrated in the right-hand panel of Figure 31, where we find the computed reference-cycle patterns corresponding to each of the three postwar reference cycles in industrial production. Thus, to take the 1945–1949 cycle pattern, the point plotted at stage I is the average of the monthly data for the three months centering on October, 1945; the value for stage II is an average of the monthly data for the first third of the expansion (November, 1945–October, 1946); the value for stage III is an average for the next third; and so on. (The monthly data that are averaged are the cycle relatives.) The nine-stage patterns for 1949–1954 and 1954–1958 are computed in a similar way.

Thus the reference-cycle patterns give us a standardized and somewhat stylized way of studying the behavior of a series over a succession of business cycles.

We are particularly interested in seeing how industrial production *typically* behaves during business cycles. This suggests the need to average the separate patterns for the individual reference cycles. This we can do by averaging all the stage I's, then all the stage II's, and so on. We thus get an average figure for each of the nine stages for all the cycles covered (for three postwar cycles, in this case). The final result is shown in Figure 32. The top half of the chart shows the average pattern computed from the three postwar cycle patterns presented in Figure 31. The lower half of the chart presents the average reference-cycle pattern for industrial production for the five cycles that occurred between 1919 and 1938. The prewar average pattern in Figure 32 was computed in exactly the same way as the one for the postwar period, but in this case we used the data and reference-cycle dates for the period 1919–1938.

Let us look carefully at Figure 32, taking first the average pattern

[10] "Specific cycle patterns" are also derived in a similar way, using the actual turning points found in the series itself. The National Bureau's measures include more than the reference and specific cycle patterns discussed in this section. Measures are also computed for each series to show the average duration and amplitude of its cycles, the average lead or lag shown by its cyclical turning points, and the extent to which the cycles in each series conform to those in general business.

for 1945–1958 in the upper part of the chart. As we should expect, industrial production conforms very closely to cycles in general business. It rises during all the stages of general business expansion and falls during all stages of general business contraction. If there is any

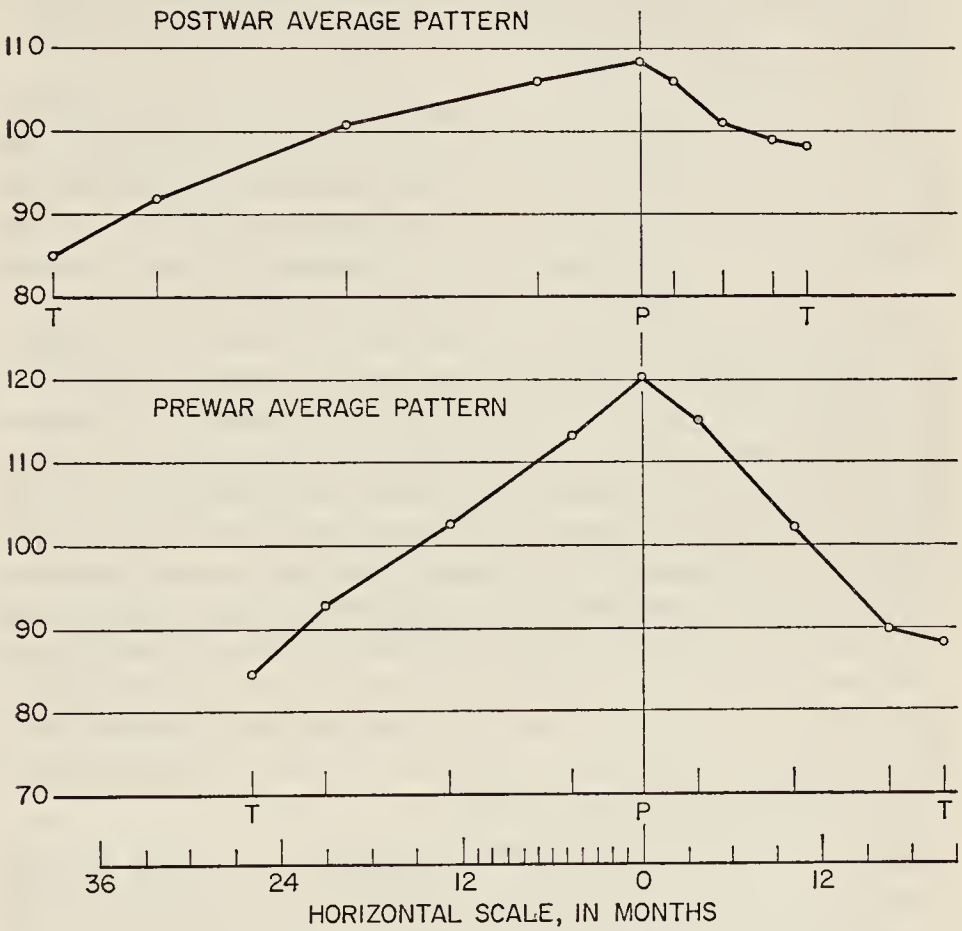


FIGURE 32. Prewar and Postwar Average Reference Cycle Patterns in Industrial Production, 1919–1938 and 1945–1958.

Based on computations of the National Bureau of Economic Research. The original series is the Federal Reserve Index of Industrial Production.

typical lead or lag at either peak or trough, it is too short to reveal itself in this chart. We can note also a tendency for industrial production to flatten out somewhat in the late stages of both expansion and contraction.

One of the most striking features of the average postwar pattern is the mildness and brevity of the contraction. The average expansion

was much longer and had a much wider amplitude than the average contraction.

Let us now refer to the average prewar pattern in the lower part of Figure 32. The greatest difference from the postwar pattern is in the longer duration and much greater amplitude of the average prewar contraction. (This contrast still holds, though in less marked degree, even if we exclude the 1930's.)

We cannot stop to evaluate in detail this particular technique of showing average cyclical behavior.[11] The critical question is: Do these average patterns reveal more than they hide? I think the answer must be in the affirmative. These average patterns do tell us what typically tends to happen during business cycles. At the same time, we must interpret these patterns with caution. They are summaries of average behavior over a number of cycles. We have seen that each business cycle differs in some respects from every other cycle; averages never tell the whole story. In addition, the National Bureau does not recognize the distinction between major and minor cycles. Any tendency for particular series to behave differently during major as distinct from minor cyclical swings, or at major as distinct from minor turning points, is concealed in these average patterns. This is why we stressed the point of diversity among business cycles before taking up their features of similarity. We can certainly use these average measures of behavior, but we can use them more intelligently if we are aware of the differences that lie behind the averages. Keeping these limitations in mind, let us look at the story that these typical patterns have to tell.[12]

TYPICAL CYCLICAL PATTERNS: PRODUCTION AND PRICES

The cyclical patterns presented in the rest of this chapter are the average reference-cycle patterns described in the preceding pages.[13]

[11] There have been several critical reviews of the National Bureau's methods of measuring cyclical behavior. See particularly E. S. Shaw, "Burns and Mitchell on Business Cycles," *Journal of Political Economy*, vol. 55, August, 1947, pp. 281–298; T. C. Koopmans, "Measurement Without Theory," *Review of Economic Statistics*, vol. 29, August, 1947, pp. 161–172; and Edward Ames, "A Theoretical and Statistical Dilemma—The Contributions of Burns, Mitchell, and Frickey to Business Cycle Theory," *Econometrica*, vol. 16, October, 1948, pp. 347–369.

[12] For another set of reference-cycle patterns, all for the period before World War II, see W. C. Mitchell, *What Happens During Business Cycles*, 1951, pp. 32–49.

[13] These patterns result from special computations made by the National Bureau of Economic Research as part of a study of prewar and postwar cycles. All

The charts that follow show these patterns for a considerable number of series. Each chart contains several sections; and each section shows for some important series the average reference-cycle pattern (1) for the first three cycles after World War II and (2) for a number of cycles before World War II (usually covering most of the 1920's and 1930's).[14] (In some cases data are not available for the prewar period or for one or more postwar cycles.) These charts will help us answer two questions in particular: First, how do different series tend to behave, on the average, as general business activity expands and contracts during business cycles? And, second, in what respects is the cyclical behavior of the American economy different now from what it was before World War II?

A few words of explanation may be helpful in interpreting the charts that follow (which, in fact, are very similar to Figure 32, which has already been discussed). The vertical scale, on which we measure the reference-cycle relatives, is the same for all charts (with one or two exceptions), and therefore differences in cyclical amplitude among the different series can be directly compared. The horizontal scale is measured in months. It will be noted that the duration of the average reference expansion after World War II was considerably greater than for the prewar period. On the other hand, the average postwar contraction was much shorter.

PRODUCTION

Let us begin with the behavior of total output. Figure 33a compares the prewar and postwar behavior of the GNP (in current prices). The prewar pattern is represented by a solid line; the postwar pattern, by a dashed line. What stands out particularly is the

---

of the computations except the averaging of the individual cycle patterns were done by the National Bureau. The final average patterns and the charts were prepared by the author. I wish to acknowledge my very great debt to the National Bureau for making these materials available and permitting me to use them here.

[14] Selection of a prewar period for averaging presents some serious problems. I have not included any cycles before 1919 even when the data were available. I have included the cycles of 1927–1933 and 1933–1938 even though to some extent they distort the averages. In examining the charts, the reader should remember that the period 1919–1938 includes one very long and severe depression (1929–1933) and two other quite sharp contractions (1920–1921 and 1937–1938). I have added some patterns for the period 1919–1927 or 1921–1927 in order to exclude the effect of the 1930's. It should be added that not all series go back as far as 1919.

failure of the GNP to decline very much in the postwar recessions. Indeed, there is no decline at all in the last two stages.

We got a somewhat similar contrast when we examined the prewar and postwar patterns in industrial production (Figure 32). Indus-

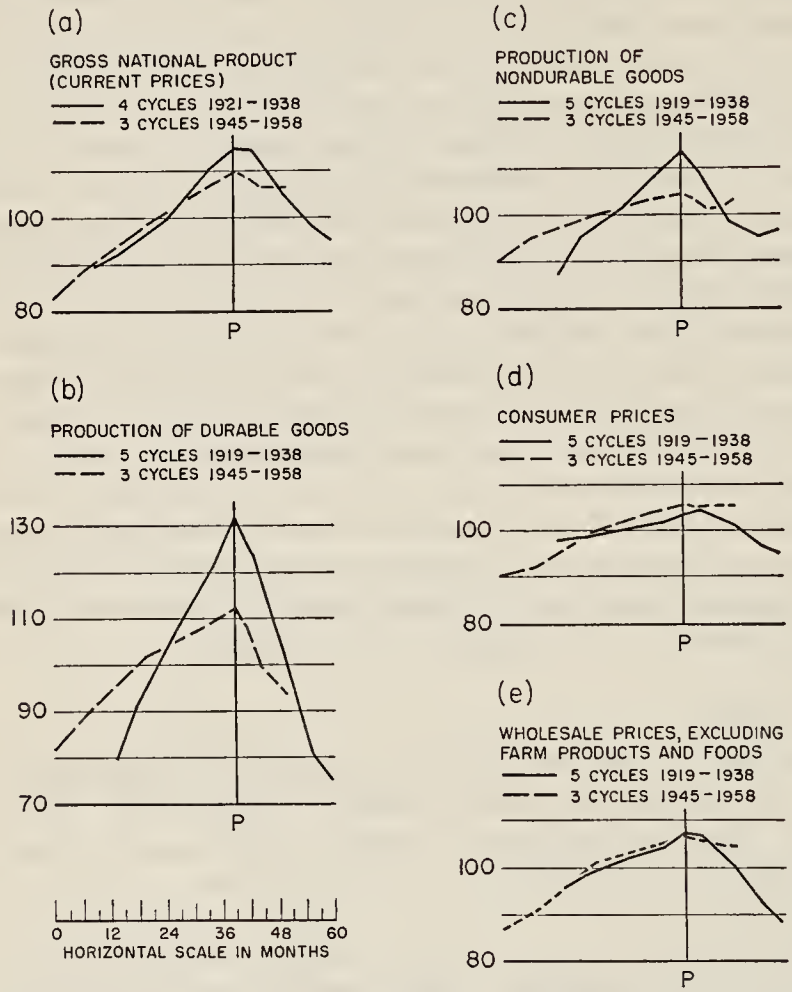


FIGURE 33. Average Reference-Cycle Patterns for Production and Prices.

The source of these data is described in footnote 13.

trial production fluctuates more widely than GNP, but both conform closely to the cycle in general business activity. Figure 32 and Figure 33a tell us that the growth trend of the American economy was more strongly upward during 1945–1958 than during the period between the two world wars, and this showed itself in the postwar

period in relatively long expansions and brief and mild contractions.

As we have already been led to expect, output of producers' and durable goods varies much more widely over the cycle than does the production of consumers' and nondurable goods. Producers' goods (raw materials, semifinished goods, and finished equipment) are bought by businessmen, and their output reflects the wide shifts in business expectations that occur over the cycle. The output of durable goods—consumers' as well as producers'—is particularly sensitive to the business cycle. This is brought out in Figure 33b.

It is clear from Figure 33c that the production of nondurable goods is relatively stable during business contractions. This was the case in prewar cycles; it has been even more true of the postwar period. What is equally striking is the clear-cut tendency for the production of nondurables to turn up before the trough in general business is reached. This lead at the trough existed before World War II and has continued through the first three postwar cycles. No lead or lag is apparent at the peak, however, in either the prewar or postwar patterns.[15] The pattern for consumers' goods production (not shown in the chart) also shows some tendency to lead. There is a marked flattening out in the later stages of contraction, a tendency that is more marked in the postwar than in the prewar pattern. Other evidence available also suggests that, on the average, consumers' and nondurable goods probably do tend to show some increase in output before the decline in general business has fully run its course,[16] and this may be an important factor in helping to bring about general recovery in some cases. We shall consider this matter further in Chapter 11.

PRICES

Two of the more interesting charts in this series are presented in Figures 33d and 33e. It was taken for granted before World War II that prices moved with the business cycle, rising when business was

[15] This does not mean that there may not have been a small average lead or lag in the original monthly data. It means simply that any small average leads or lags that do exist are not large enough to show up in the averaging process used to compute the cycle patterns.

[16] Cf. F. C. Mills, *Price-Quantity Interactions in Business Cycles,* 1946, and N. J. Silberling, *The Dynamics of Business,* 1943, chap. 19, esp. p. 472. See also J. M. Clark, *Strategic Factors in Business Cycles,* 1935, p. 45, for some discussion of the behavior of consumers'-goods production in the business cycle.

expanding and falling when business declined in a recession. Figure 33 suggests that this is an accurate description of the cyclical behavior of prices between the two world wars. This is also the way prices behaved before World War I.

These two charts point up dramatically the radically different behavior of prices since World War II. The contrast is most marked in the case of consumer prices, which, on the average, did not decline at all in the first three postwar recessions.[17] The consumer price index rose in every cyclical expansion, but merely moved horizontally—instead of declining—during recession. This creates what has been referred to as a "ratchet effect," whereby prices move in one direction (upward) but not in the other. The net result, indicated by the sharp upward tilt of the average postwar cycle pattern in Figure 33d, was the inflationary trend in prices during 1945–1958.

Nonfarm wholesale prices have behaved in a similar manner in the postwar period. In this case, there has been a net average decline in recessions, but it amounted to only a small fraction of the average rise during reference expansions. Again, the net result has been a strong inflationary trend.

Normally, we expect wholesale prices to fluctuate more widely over the business cycle than do retail prices. This was clearly the prewar pattern. It has continued to be true, although not quite so noticeably, in the postwar period. The prewar patterns also reveal that consumer prices lagged by a full stage at the downturn. While this does not show up in Figure 33, there was some tendency in the prewar years for wholesale prices to lead at the peak and, to some extent, at the trough. This tendency to lead was most marked in the case of raw material prices.

THE BEHAVIOR OF INVESTMENT

We have already emphasized that private investment plays a key role in making the economy unstable. Investment fluctuates much more widely, relatively, than does consumption (or government spending), and it is not tied to the level of income as closely as are consumers' expenditures. It behooves us, therefore, to pay particular attention to the cyclical behavior of investment.

[17] The consumer price index fell slightly in the 1948–1949 recession, moved more or less horizontally in 1953–1954, and rose in the 1957–1958 recession.

FIXED INVESTMENT

Figure 34 provides us with some illuminating evidence regarding the behavior of investment in prewar and postwar cycles. Consider the behavior of construction and expenditures on producers' durables (i.e., machinery and equipment) in the various sections of this

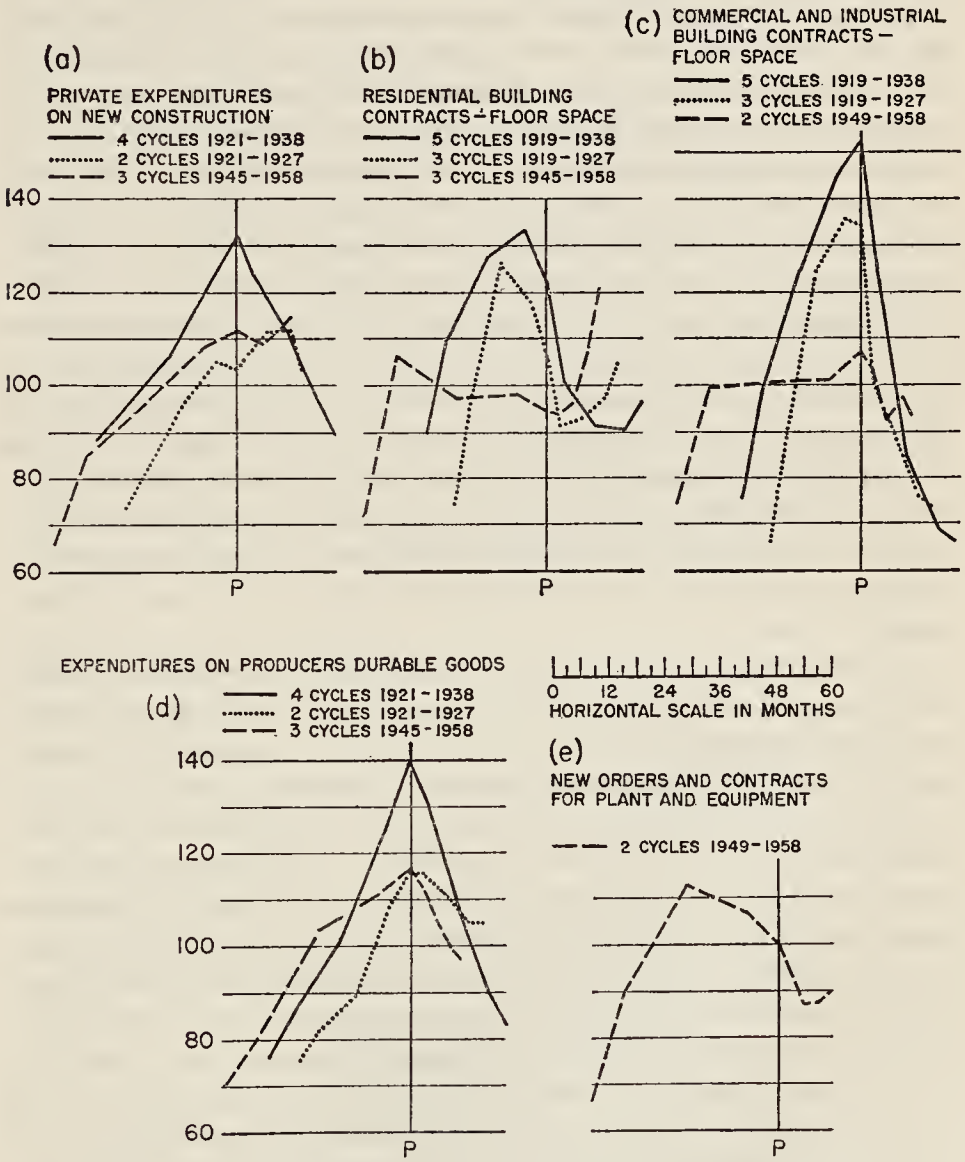


FIGURE 34. Average-Reference Cycle Patterns in Investment.

The source of these data is described in footnote 13.

chart. These are the two main components of gross domestic capital formation.

The difference in the prewar and postwar patterns for construction (Figure 34a) tells us something important about one of the reasons for the greater stability of the American economy since World War II. If we average all the cycles between 1921 and 1938, including the one covering the severe depression of the 1930's, we get the wide amplitude of movement shown by the solid line in Figure 34a. When construction drops as much as it does in this prewar average reference pattern, a serious depression is almost unavoidable.

Now let us look at the other two cycle patterns in Figure 34a. Here we portray the postwar average pattern for new construction (dashed line) and also the average pattern for two minor cycles in the 1920's which ended in mild recessions somewhat similar to those experienced after World War II (dotted line). The postwar pattern shows very little drop during the recession phase and indeed turns up sharply well before the trough in business is reached in stage IX. The average pattern for the two prewar minor cycles is similar in showing only a small decline during the downswing in general business. The timing is quite different, however, there being a substantial lag at the reference peak and no lead at the trough. But the important point is that construction was well maintained in both prewar and postwar *minor* cycles.

The reason for the early upturn in construction in postwar recessions is suggested by Figure 34b. It is residential building that makes for the early revival in construction. Indeed, housebuilding begins to rise so early during the average postwar reference contraction that we almost get the impression of an inverted cycle. This impression is strengthened by the fact that the postwar average pattern for residential building also declines during most of the stages of business expansion. Thus, during the postwar period, housebuilding has been a stabilizing influence in the economy. By declining in the later stages of business expansion, it helps to moderate the boom. By expanding early and vigorously after business turns down, it helps to keep recessions brief and mild. This stabilizing behavior is due to two sets of causes. First, the underlying demand for housing has been strong all through the postwar period. Second, for reasons that we shall explain later, this underlying demand is restrained by a scarcity of mortgage credit during booms and is unleashed by the greater availability of mortgage credit during recessions.

Actually somewhat this same behavior of residential building occurred during the 1920's. The average pattern for 1919–1927 (which excludes the Great Depression of the 1930's) shows a recovery as early as does the postwar record (see the dotted line in Figure 34b), but the preceding decline is much sharper. One thing that is clear from the patterns in Figure 34 is that residential building shows a long lead at both reference peaks and reference troughs. The lead is particularly long in minor cycles.

We need comment only briefly on the cyclical behavior of nonresidential construction in Figure 34c. Note the mildness of the decline in the postwar pattern.[18] Here again we have an indication of favorable investment opportunities that have kept this form of investment high even when business has turned down, which, of course, has helped to keep postwar recessions mild. In contrast, the prewar cycle pattern (including the 1930's) shows nonresidential building rising rapidly throughout reference expansion and falling equally rapidly during reference contraction. If we exclude the 1930's and take only the period 1919–1927 (dotted line), we get somewhat the same pattern, except that the amplitude is somewhat less extreme and a tendency to lead at the peak appears.

Let us now look at the pattern for producers' durables (Figure 34d). There is a marked cyclical pattern in this series for the postwar period, with a significant decline occurring during recessions. Thus, expenditures on producers' durables have been less stable in postwar business-cycle contractions than construction. Actually the postwar pattern has been similar to that of the two minor cycles in the 1920's (although without the lag at the peak shown by the latter), but the amplitude is far less than for the prewar pattern that includes the 1930's.

Except for the leads in construction already described, fixed investment expenditures do not show a clear lead or lag at the turning points.[19] But *new orders* for plant and equipment do lead, as is suggested by the postwar pattern (covering two cycles only) in Figure 34e. Not fully comparable data suggest that similar leads, perhaps not so long, also existed in prewar cycles. Thus, it seems that busi-

[18] This pattern includes only two cycles from 1949 to 1958. The 1945–1949 cycle is excluded because of distortions created by adjustment to a peacetime economy.

[19] As previously noted, expenditures on producers' durables lagged at the peak for the two minor cycles of the 1920's.

ness firms tend to anticipate the turning points in business in the investment orders and contracts they place. Such orders can, of course, decline for quite a while before actual investment expenditures, based on orders already placed, begin to fall.

INVENTORIES

The part of total investment that is most sensitive to business-cycle influences is investment in inventories. The extent of variation in this series—from large positive to large minus figures—is suggested by the data in Figure 35. Changes in inventory accumulation accounted for a larger fraction of the decline in GNP in the recessions after World War II than did any other form of investment. This was generally the case in earlier business contractions, also.[20]

In studying the cyclical behavior of inventories, we need to distinguish between the behavior of total inventories and the cyclical pattern of investment (i.e., net change) in inventories. The total size of inventories, taken in conjunction with the level and direction of change in sales, is of critical importance in determining whether businessmen choose to expand or contract their purchases and payrolls. But it is the amount of *change* in their inventories, not their total inventory holdings, that enters into our data on total investment.

As we should expect, total inventory holdings of business firms tend to move up and down with the business cycle. But this is just the beginning of a fairly complicated story. In manufacturing, for example, which accounts for more than half of all nonfarm stocks of goods, inventories rise and fall with business activity—but with a noticeable lag. Inventories usually do not reach their peak until after business recession begins, and do not begin to rise until after the trough in business has been reached. (See Figure 35, in which the vertical lines mark the reference peaks and troughs in general business.) This lag in inventory holdings also tends to show up in wholesale and retail trade. Interestingly, these lags have become less pro-

[20] On the prewar period, see Moses Abramovitz, *Inventories and Business Cycles,* 1950, pp. 5–8. This is much the most authoritative book on the cyclical behavior of inventories. See also the forthcoming National Bureau study, *Postwar Cycles in Manufacturers' Inventories,* by T. M. Stanback, Jr. The discussion in the text owes much to these two studies. See also G. M. Cobren and M. Liebenberg, "Inventories in Postwar Business Cycles," *Survey of Current Business,* April, 1959, pp. 3–8.

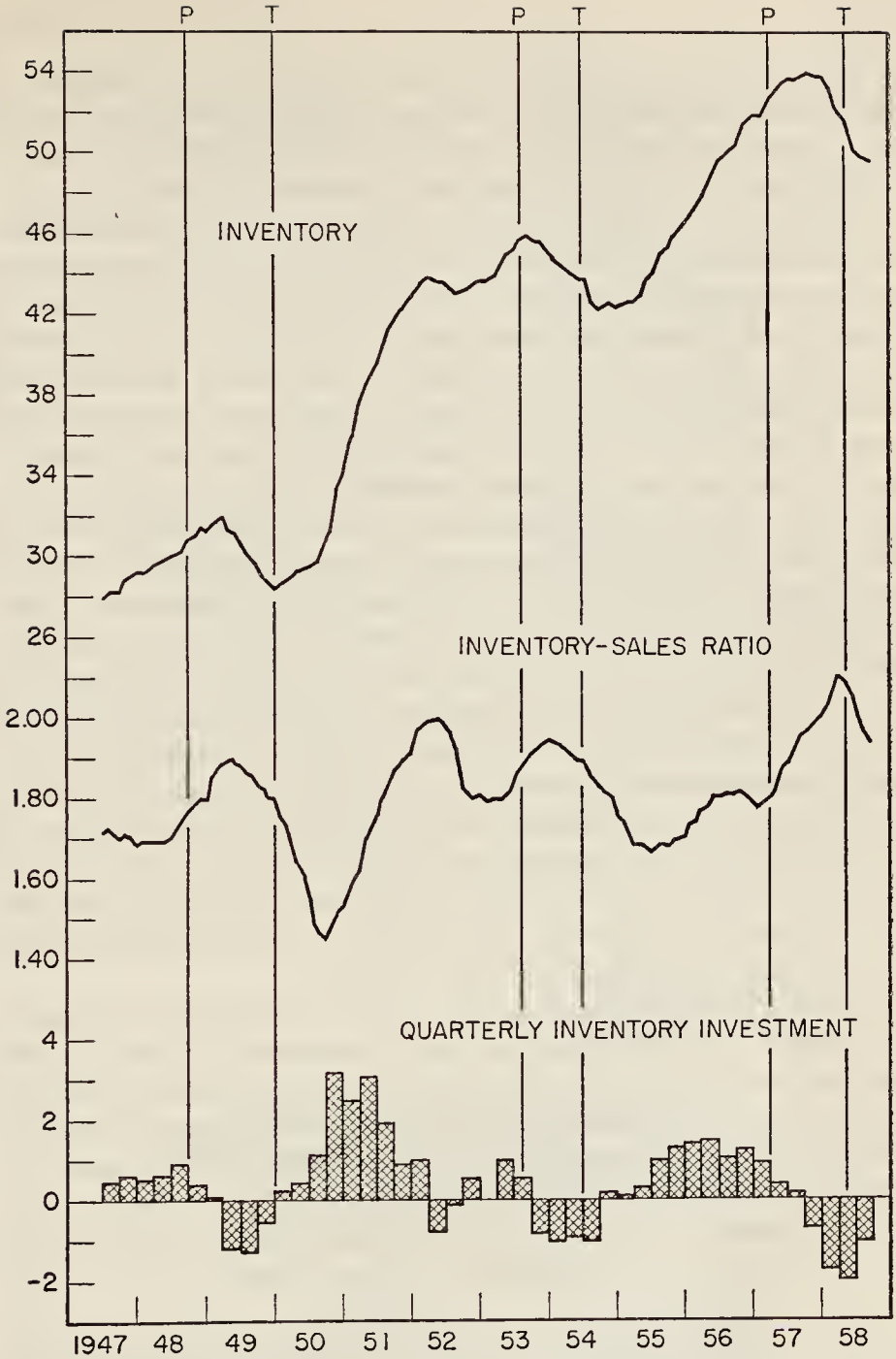


FIGURE 35. Manufacturers' Inventories and Inventory Investment, 1947–1958.

Reproduced by permission, and with some changes, from P. G. Darling, "Manufacturers' Inventory Investment, 1947–1958," *American Economic Review*, vol. 49, December, 1959, p. 954. Vertical lines mark the peaks and troughs in general business. Dollar figures are in billions.

nounced since World War II.[21] This is evident, for example, in Figure 36a. The prewar lag at the peak in department-store stocks is very noticeable, but whatever average lag exists in the postwar period does not show up in the nine-stage pattern in Figure 36.

The lag in total manufacturers' inventories results from the differing behavior of different classes of inventories. In this connection, we need to distinguish between stocks of finished goods, of work-in-process, and of purchased materials. It is finished-goods inventories that tend especially to lag. In the postwar period, stocks of purchased materials have tended to show a lead at reference peaks and some lag at reference troughs. Goods-in-process have tended more or less to coincide with the turns in general business activity.[22]

What do these movements in *total* inventories imply as to the cyclical behavior of current investment in *additional* inventories? If businessmen always sought to maintain total inventories at a constant ratio to sales and output, as the usual formulation of the acceleration principle assumes, then the net change in inventories would vary with the *rate of change* of output.[23] Since the maximum rate of change always occurs before the actual peak or trough in output, this would suggest that current investment in additional inventories should turn up or down before the corresponding turning points in general business. Businessmen should begin to reduce their rate of accumulation of inventories before the peak in business, and they should begin reducing their rate of liquidation of inventories before the contraction has fully run its course.

Actually, considerable evidence exists that before the war total investment in manufacturing and trade inventories reached its peaks and troughs at about the same time as general business activity, not before. This particular timing apparently resulted from a lag in investment in finished-goods inventories and leads in investment in inventories of materials and work-in-process.[24]

There seems to have been some change in timing in the postwar period. Leads have become longer and lags shorter, with the result

[21] For manufacturers' inventories, see Stanback, *op. cit.* Stanback also notes that there is some tendency for manufacturers' inventories to turn earlier at the peak than at the trough.

[22] Cf. Stanback, *op. cit.* For substantially similar findings for the prewar period, see Abramovitz, *op. cit.*

[23] Actually, the inventory-sales ratio varies over the cycle. See Figure 35.

[24] See Abramovitz, *op. cit.*, chap. 20.

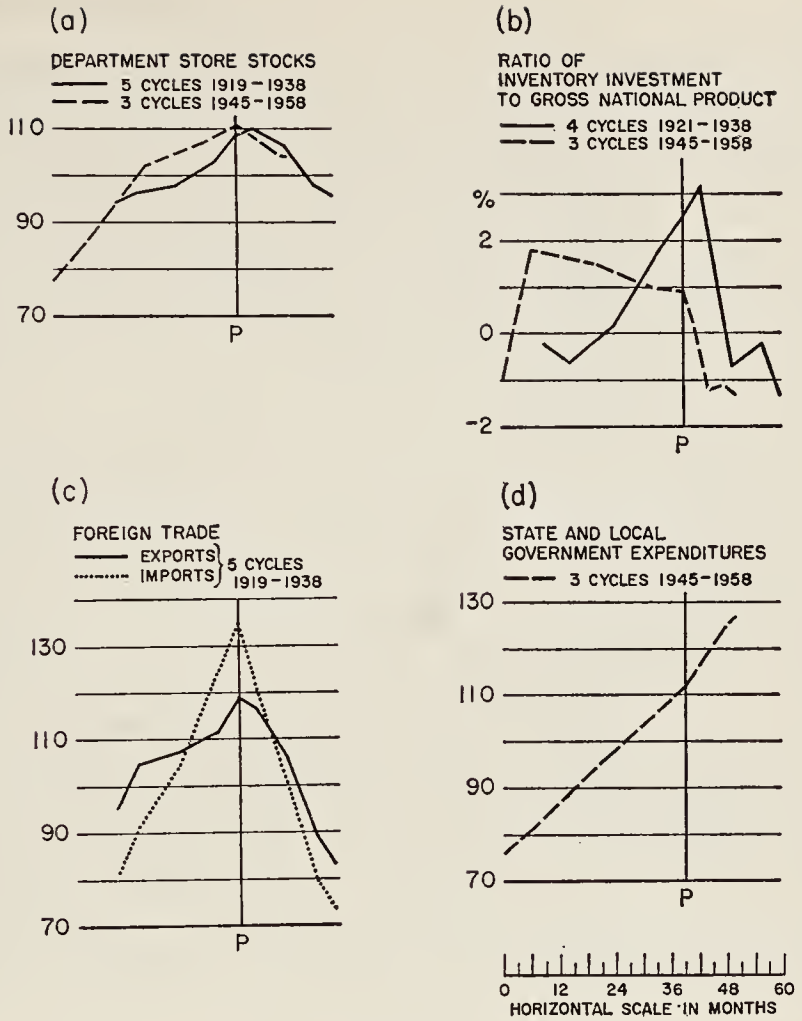


FIGURE 36. Average Reference-Cycle Patterns in Inventories, Foreign Trade, and Government Expenditures.

The source of these data is described in footnote 13.

that total inventory investment in our three postwar cycles shows a marked average lead at the peak, although not at the trough. (See the bottom series in Figure 35, and Figure 36b.) Investment in work-in-process and purchased materials inventories have continued to lead, but the prewar lag in finished-goods inventories seems largely to have disappeared, at least in the postwar cycles thus far observed.[25] Businessmen have been keeping a tighter rein on inventories than before the war, and have been able to reduce them more quickly once sales begin to fall off.[26]

One final question about the postwar behavior of inventories should be raised: Are the cycles in inventory investment wider or narrower than before the war? If we take just the dollar figures, the answer is clear that they are wider. This is not surprising, since prices are much higher and so are total sales and output. One test to correct for this is to divide inventory investment by GNP (both in current prices) and to look at the cyclical behavior of this ratio. This is what we have done in Figure 36b. On this sort of relative basis, fluctuations in inventory investment have, if anything, been somewhat narrower than before the war. They have been of about the same amplitude as during the minor cycles of 1921–1924 and 1924–1927 (not shown on the chart). However, total inventories today are smaller relative to sales than they were in the 1920's and 1930's. Although total inventories are smaller relative to the GNP than before the war, inventory investment still plays about as important a role in minor cycles as it did in the prewar years.

NET EXPORTS AND GOVERNMENT EXPENDITURES

The net export surplus is not a large component of GNP in the United States. Ordinarily, imports are more sensitive to domestic business-cycle forces than are exports. As a result, our export surplus showed a tendency before the war to be reduced during cyclical expansions (as imports expanded more rapidly than exports) and to increase during cyclical contractions. This can be seen in the rela-

[25] Cf. Stanback, *op. cit.*

[26] A look at the behavior of manufacturers' inventory investment in Figure 35 suggests that special factors helped to create the long average lead at the peak shown in Figure 36b. The very early peak reached in 1951 (two years before the 1953 reference peak) was associated with the Korean War. Also, the very long lead before the 1957 reference peak was associated with the particular nature of the 1955–1957 boom. On all this, see Chapter 16.

tionship between the prewar patterns for exports and imports in Figure 36c.[27]

No average pattern is presented for the period after World War II because of the abnormal international conditions that existed during most of this period. The international aspects of the business cycle will be considered further in Chapter 22.

As we shall see in Chapter 16, a major support to the economy since World War II has been the very high level of federal government expenditures and the sharp upward trend in the spending of state and local governments. So far as the federal government is concerned, it is much easier to talk about typical cyclical patterns of receipts and expenditures in the prewar than in the postwar period. Of the three postwar cycles, the first was strongly colored by the sharp drop in military expenditures at the end of World War II, and the second spanned the period of hostilities in Korea. Because of the somewhat erratic character and large size of the changes in postwar defense expenditures, federal government spending has sometimes been a destabilizing force.[28] One illustration is the decline in government expenditures during the 1953–1954 recession.[29]

If we take both total expenditures and receipts, the federal government's budget tends normally to act as a stabilizing factor in business cycles. Tax receipts rise faster than expenditures during business expansion, causing the budgetary surplus to rise (or the deficit to decline), and tax receipts fall during business recessions while expenditures fail to decline or actually rise. This was the situation before World War II, and it is also the case now. This can be seen in the cyclical patterns for the period 1921–1938 and for the 1954–1958 cycle presented in Figure 36e.[30]

[27] See also the painstaking empirical study by Ilse Mintz, *Trade Balances during Business Cycles: U.S. and Britain Since 1880,* National Bureau of Economic Research Occasional Paper 67, 1959. Average reference cycle patterns for the United States trade balance are shown on page 20 of that study. A similar analysis is also provided for Great Britain.

[28] Cf. B. G. Hickman, *Growth and Stability of the Postwar Economy,* 1960, chap. 9.

[29] See p. 489, below.

[30] The data for these particular charts are taken from J. M. Firestone, *Federal Receipts and Expenditures During Business Cycles, 1879–1958,* 1960, pp. 164–165. This source should be consulted for a more detailed analysis of the cyclical behavior of federal receipts and expenditures. It should be noted that these figures are for budget receipts and expenditures and exclude transactions of the social security and other trust accounts.

Figure 36d presents the postwar cyclical pattern for state and local government expenditures. The upward tilt in this pattern is truly striking. State and local expenditures have risen vigorously through every stage of cyclical expansion, and they have continued to rise rapidly during postwar business contractions. Government spending at the state and local level has contributed to the inflation of aggregate demand during boom periods, but it has been an important stabilizing influence during postwar recessions.

CONSUMPTION, INCOME, AND EMPLOYMENT

CONSUMPTION

We already know that consumption fluctuates relatively less widely over the business cycle than does investment. We have seen also that consumers' expenditures are more stable than total national income or GNP and also more stable than disposable income. The percentage of total income consumed tends to rise during business contractions and to fall during business expansions.[31]

We can trace the cyclical behavior of consumers' expenditures in Figure 37a. Clearly, consumers' buying has been a strong stabilizing force in postwar recessions. The postwar pattern does not decline at all after business turns down; it moves horizontally for a while and then starts to rise again before business generally begins to recover. A somewhat similar pattern is evident in the minor recessions of the 1920's (dotted line). Here there is clearer evidence of a lag at the peak; the series then turns down for one stage; and finally there is a clear lead at the trough. If we bring in the 1930's, with the long decline of 1929–1933 and the sharp decline of 1937–1938, we get the usual kind of conforming cyclical pattern—a rise in consumers' expenditures throughout business expansion and a significant decline throughout business contraction.

Figure 37b provides some further evidence regarding the cyclical behavior of consumers' buying. Department store sales, it is clear, are more sensitive to business recessions than are total consumers' expenditures. (The former does not include spending on food, which is very stable, or on services, which have risen through each

[31] In addition to the brief discussion here, the reader is referred to the analysis of the relation between consumers' expenditures and income in Chapter 5.

postwar recession.[32]) The postwar pattern also rises before the final trough in business, but this lead is not evident in the prewar patterns. The prewar patterns, however, show a tendency to lag at the peak. This is particularly the case for the average of the three minor cycles of the 1920's (dotted line).

When we break down consumption into the usual categories, we find the sort of contrast that we should expect. The greatest cyclical fluctuation is in the flow of consumers' durable goods; purchases of perishable and semidurable goods and of services fluctuate much less widely.[33] (As we have already noted, expenditures on services have risen throughout each of the mild postwar contractions.) Purchases of durable goods are postponable, and they are sensitive to changes in income and to changes in consumers' expectations and attitudes. They also frequently involve going into debt. Hence, they are the most volatile part of consumers' expenditures.

INCOME AND EMPLOYMENT

We need refer only briefly to the cyclical behavior of personal income, the prewar and postwar patterns for which are presented in Figure 37i. The stability of personal income in postwar recessions is what we should expect. Note, however, that income was equally stable in the minor recessions of the 1920's (dotted line). There is some evidence that personal income has tended to lag slightly at business cycle peaks but not at the troughs. This lag is too short to show in the patterns in Figure 37.[34]

Several long-run forces have been at work to change the composition of personal income in such a way as to make it more stable in business recessions. Government salaries, which are very stable over the cycle, have become a steadily larger fraction of total payrolls, and the relative importance of transfer payments (which tend to rise in recessions) has also increased. The decline in the importance of

[32] It also does not include the purchase of automobiles, which is sensitive to cyclical declines.

[33] Cf. Simon Kuznets, *National Income: A Summary of Findings,* 1946, p. 107. For detailed information regarding the sensitivity of various types of consumption expenditures to changes in disposable income, see *Survey of Current Business,* September, 1955, May, 1957, and March, 1959.

[34] Cf. Daniel Creamer, *Personal Income During Business Cycles,* 1956, p. 17.

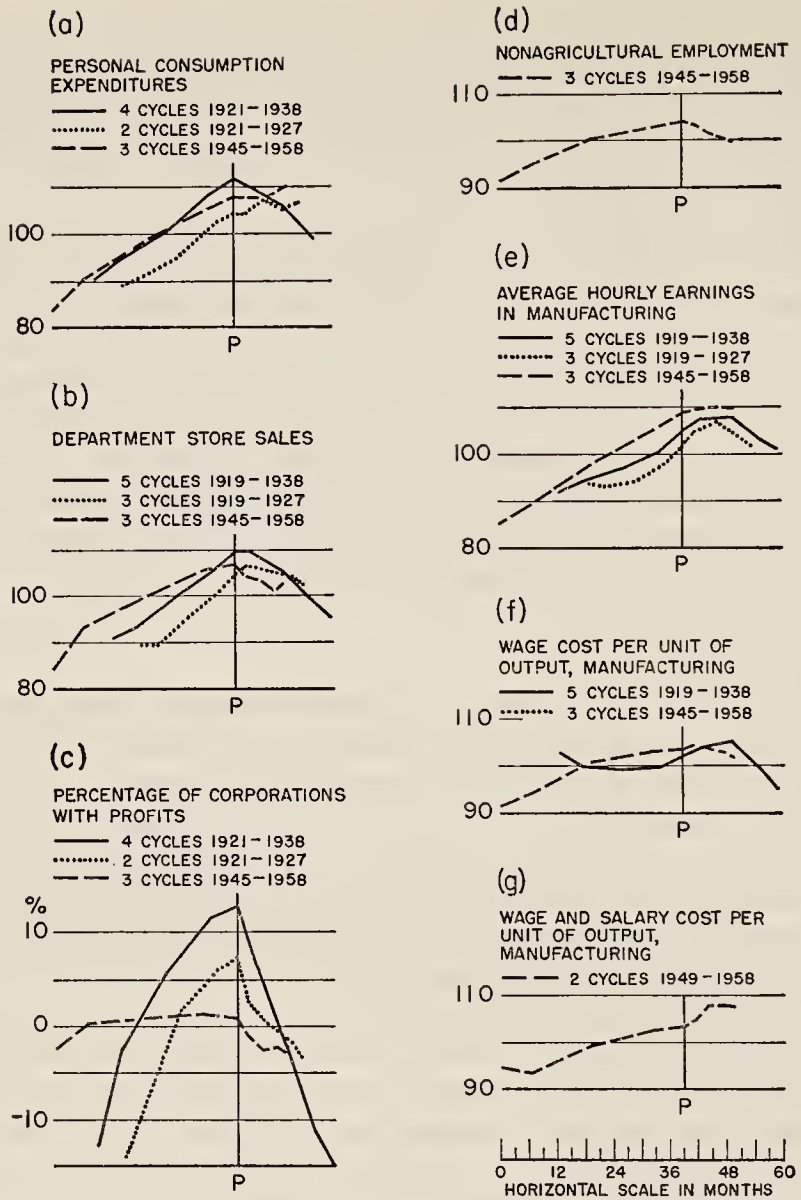
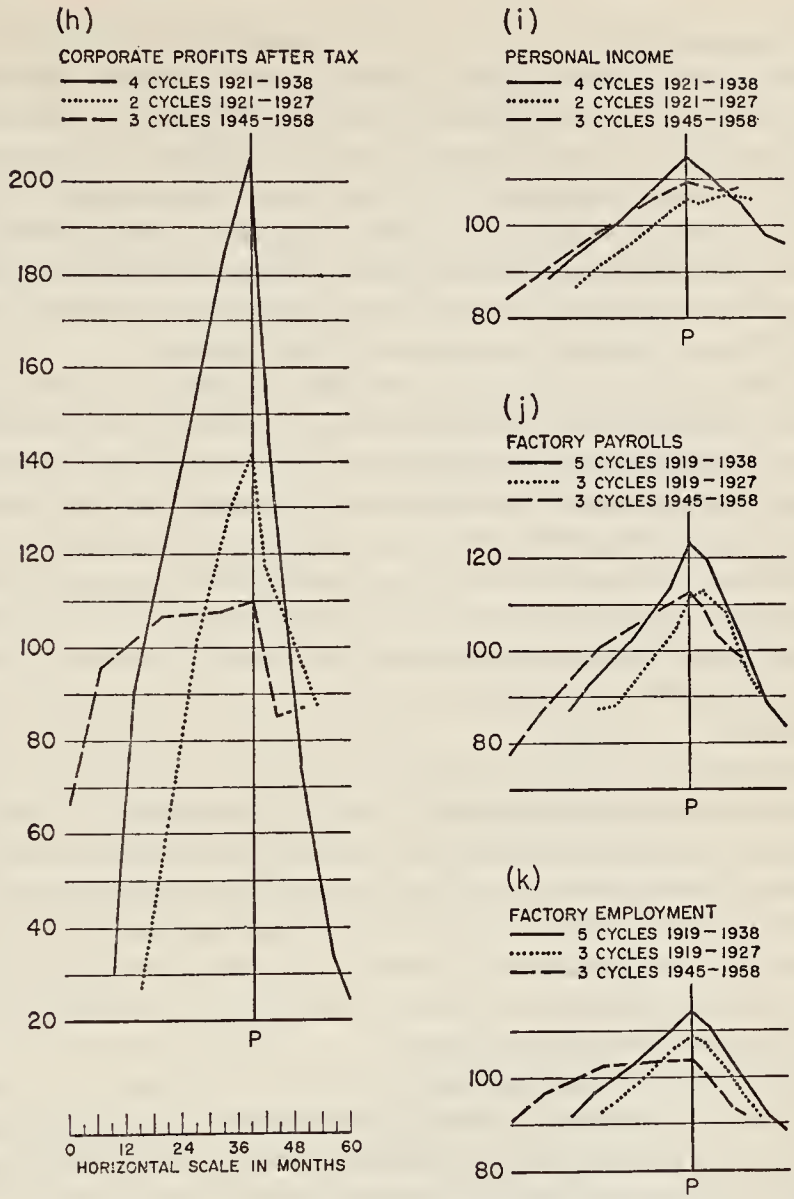


FIGURE 37. Average Reference-Cycle Patterns in Consumption, Employment, Wages and Labor Costs, and Profits.

The source of these data is described in footnote 13.

farm incomes, which are very volatile, has also had a stabilizing effect. These have been only partly offset by the decline in the relative importance of rent and interest (which are comparatively stable over the cycle) and by the increased importance of the durable-goods in-



dustries, in which incomes fluctuate widely.[35] The postwar growth in the relative number of salaried workers in the labor force has also had a stabilizing effect, since salaries fluctuate less over the cycle than do wages.

Of the various shares into which we usually divide the national in-

[35] See Geoffrey Moore's foreword in Creamer, *op. cit.*, pp. xxviii–xxxi. The changes listed in the text refer only to trends in the composition of incomes and do not cover all of the automatic stabilizers referred to on p. 216.

come, business profits show the widest changes over the business cycle. The volatility of corporate profits, particularly in the prewar period, is brought out in striking fashion in Figure 37h. Despite the mildness of the postwar recessions and the failure of personal incomes to decline very much, corporate profits dropped significantly, although not so much as in the minor recessions of the 1920's. The postwar tendency of corporate profits to flatten out in the late stages of expansion (and contraction) is probably of considerable significance. This is a point to which we shall return.

In contrast to profits, payments for interest and rent show a high degree of stability over the cycle. So, on the whole, do dividend payments, which fluctuate much less than do the corporate profits out of which they are paid. On the average, dividends scarcely declined at all in the three postwar recessions. As a result, net corporate savings fluctuate even more widely than profits. They may even become negative in a severe depression, as they did during part of the 1930's. The tendency for dividends to be maintained when corporate profits fall is one of the "automatic stabilizers" to which we have already referred.[36]

The largest share of the national income, of course, goes to labor in the form of wages and salaries. The cyclical behavior of total payrolls is a product of the changes occurring in three variables: employment, average number of hours worked per employee, and average hourly earnings. The typical cyclical patterns for most of these variables in the manufacturing sector are shown in Figure 37. Factory payrolls (Figure 37j) showed considerable variation over the cycle even in the postwar period.

Employment in manufacturing generally fluctuates more widely than it does in retail trade, the service industries, or government. As a result, the cyclical movement in total nonagricultural employment (Figure 37d) is less than it is in factory employment (Figure 37k).

The number of hours worked by the average employee tends to vary over the cycle, and this reinforces the effect on payrolls of the substantial fluctuations in employment.[37] In addition, average hourly earnings tend to respond to some extent to cyclical changes in

[36] The role of this stabilizer in the postwar recessions is treated at some length in Chapters 15 and 16.

[37] Hours worked per week is a sensitive business indicator which tends to lead the turns in general business at both peak and trough. Cf. Gerhard Bry, *The Average Workweek as an Economic Indicator,* National Bureau of Economic Research Occasional Paper 69, 1959.

business (Figure 37e). But the chief cause of the cyclical instability of payrolls is the rise and fall of employment during business cycles. And in this respect, the most unstable sectors of the economy are manufacturing, mining, and construction.[38]

On the whole, both employment and payrolls in manufacturing show some tendency to lag behind production at the turning points, although the lags have not usually been very large. (They are too short to show in Figure 37.) Hourly earnings have a marked tendency to lag, especially at the peak; wage rates do not ordinarily turn down until well after the turning point in general business.[39] (Compare Figure 37e.) As labor has become well organized, the resistance to wage decreases during business contractions has grown stronger, and the postwar pattern for hourly earnings shows no decline in recession. However, this resistance to wage cuts was evident to some extent even before the period of rapid unionization in the 1930's.

Until fairly recently, economists thought that *real* wages tended to move inversely with the business cycle—to decline during business expansions and to rise during business contractions. It was generally assumed that money wages failed to rise as rapidly as retail prices during the upswing or to fall as fast during the downswing. Recent studies indicate, however, that this is much too simple a picture of what happens during the business cycle. Real wages have certainly not behaved in any such simple fashion during American business cycles of the last 40 years. Real factory wages, it is true, have tended to rise during business declines; but they have also, more often than not, risen during business expansions. If we turn to the relation between factory wages and *wholesale* prices, we find that, during peacetime cycles, hourly earnings regularly have failed to fall as much as prices during business declines but sometimes (though not always) have risen as much as or more than wholesale prices of manufactured goods during business expansions.[40] As far as manufacturers are concerned, it is the relation between wholesale prices and

[38] For a more detailed empirical study of cyclical fluctuations in labor income, see Creamer, *op. cit.*, chap. 5.

[39] See Daniel Creamer, *Behavior of Wage Rates During Business Cycles*, National Bureau of Economic Research Occasional Paper 34, 1950.

[40] Based on a comparison of hourly earnings in manufacturing and wholesale prices of manufactured goods. Compare Creamer, *ibid.*, p. 37. The best study thus far done of the relation between money wages and wholesale prices is Sho-Chieh Tsiang, *The Variations of Real Wages and Profit Margins in Relation to the Trade Cycle*, 1947.

wage rates that influences profit margins, not that between wages and retail prices.

PRODUCTIVITY, LABOR COST, AND PROFIT MARGINS

Another assumption frequently made by economists is that labor productivity tends to fall during business upswings (as less efficient labor is hired and as firms begin to encounter diminishing returns in expanding their output) and that the reverse happens during downswings. This is not an accurate picture. Actually, output per man-hour in manufacturing has shown a net increase during every cyclical contraction *and* expansion since World War I.[41]

The preceding discussion refers to output per man-hour of so-called production workers. One striking characteristic of the recent postwar period has been the marked rise in the employment of salaried workers, which has been in part a result of new technological and organizational changes.[42] The increase in salaried workers has come particularly in boom periods, when large investment expenditures have been made to take advantage of these new improvements. But the latter part of the expansion is also a time when the rise in output tends to taper off. As a result, the rise in productivity of production and salaried workers combined has been relatively small in the later stages of recent expansions; it has fallen with the decline of output in the early stages of contraction; and it has risen in late contraction and particularly in the early stages of recovery.[43] Thus, a considerable part of the postwar gains in labor productivity have come in the general neighborhood of business-cycle troughs—that is, in late contraction and especially in the early stages of business expansion.[44]

It is the combination of labor productivity and wages paid that

[41] Cf. Thor Hultgren, *Chnages in Labor Cost During Cycles in Production and Business,* National Bureau of Economic Research Occasional Paper 74, 1960, p. 55.

[42] Cf. F. Beatrice Coleman, "Postwar Changes in Manufacturing Payrolls," *Survey of Current Business,* December, 1959, pp. 19–24.

[43] See, in particular, Edwin Kuh, *Profits, Profit Markups, and Productivity,* Study Paper No. 15 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1960), pp. 85–94; also C. L. Schultze, *Recent Inflation in the United States,* Study Paper No. 1, *ibid.*

[44] In addition to the references in the preceding footnote, see Hultgren, *op. cit.,* pp. 51, 56–58, and Solomon Fabricant, *Basic Facts on Productivity Change,* National Bureau of Economic Research Occasional Paper 63, 1959, pp. 15–16.

creates the labor costs the businessman incurs. If we take labor cost per unit of product (for production workers in manufacturing only), we get the results in Figure 37f. Before World War II, wage cost per unit of product fell at the beginning of expansion, then rose in late expansion and early contraction, and finally declined sharply in the last stages of the business downswing. In the postwar pattern, unit labor cost rises throughout the business expansion and then falls slightly, with a lag, during business contraction. In the prewar cycles, increases in hourly earnings were offset by improvements in labor productivity in the first part of business expansion but not in the latter part. In early contraction, productivity declined while wages continued to rise for a while. In late contraction, both declining hourly earnings and improving productivity brought about a sharp decline in labor costs. As can be seen from Figure 37f, the postwar sequence has been somewhat different. The rise in hourly earnings has exceeded the increase in labor productivity throughout the upswing, so that unit labor cost has risen during the entire expansion. During the downswing, unit labor cost turned down a stage earlier than before the war, but the decline was much less because hourly earnings did not fall.

If we include salaried workers (Figure 37g), we find that total labor cost per unit of manufacturing output falls in the first stage of postwar recovery, then rises steadily through the rest of expansion and well into contraction, and then falls very slightly in late recession. The drop in early recovery is what we should expect, since this is the stage in which output is most likely to rise more rapidly than total labor cost.

What is the effect of these related movements in prices, wages, and productivity on business profits? Our information on profit *margins* per unit of product is limited, and what we have is chiefly for the postwar period. For the postwar period, particularly if we exclude the initial postwar inflation of 1946–1948, there is some evidence that total costs per unit have risen faster than prices in the late stages of expansion, with a consequent decline in profit margins, and the latter have fallen further in the early stages of contraction. Toward the end of postwar contractions there has apparently been some improvement in profit margins.[45]

[45] See Kuh, *op. cit.*, pp. 75–82; also National Bureau of Economic Research, *Thirty-ninth Annual Report*, 1959, pp. 46–49, and P. B. Simpson and P. S.

Before the war, total profits conformed perfectly to business cycles, rising throughout the upswing and falling throughout the downswing. There was no tendency for profits to lead at the turning points.[46] Figure 37h, however, suggests a different cyclical behavior in postwar cycles. There is a noticeable tendency for profits to flatten out in the latter part of expansion, and a clear lead is shown at the lower turning point. Figure 37c is also suggestive in this connection. The percentage of companies showing some profits declines before the peak in general business. (A similar lead, however, does not show up at the trough.) Taking this and other evidence into account, it is probably fair to say that something like the cost-price sequence described by W. C. Mitchell has been more characteristic of postwar cycles than of those between the two world wars.[47] There does seem to have been a tendency in the last decade or more for the price-cost margin to narrow in the later stages of expansion, with a consequent deterioration of profit prospects in an increasing number of firms. And before or at the trough in business, price-cost relations become more favorable, and profit prospects improve.

CYCLICAL PATTERNS IN MONETARY AND FINANCIAL SERIES

The analysis of earlier chapters suggested that monetary factors—particularly fluctuations in bank credit and interest rates—may play an important role in business cycles. Figure 38 permits us to look at the cyclical patterns in some of these monetary variables.

MONETARY PATTERNS

Short-term rates of interest (Figure 38a) generally show a marked cyclical pattern. Figure 38 suggests that this has been even more true

---

Anderson, "Liabilities of Business Failures as a Business Indicator," *Review of Economics and Statistics,* vol. 39, May, 1957, pp. 197–199.

[46] Barger's series for profits in manufacturing, after adjustment for inventory profits or losses, lags at three of the four peaks and four of the five troughs occurring between 1921 and 1938. See also Thor Hultgren, *Cyclical Diversities in the Fortunes of Industrial Corporations,* National Bureau of Economic Research Occasional Paper 32, 1950, p. 12. Hultgren finds, however, that the number of companies experiencing falling profits begins to increase before the peak in business activity is reached.

[47] See pp. 344–346, below. There have been some factors at work since World War II that were not important when Mitchell wrote, for example, the prevalence of collective bargaining and the accelerated rise in salary relative to wage costs.

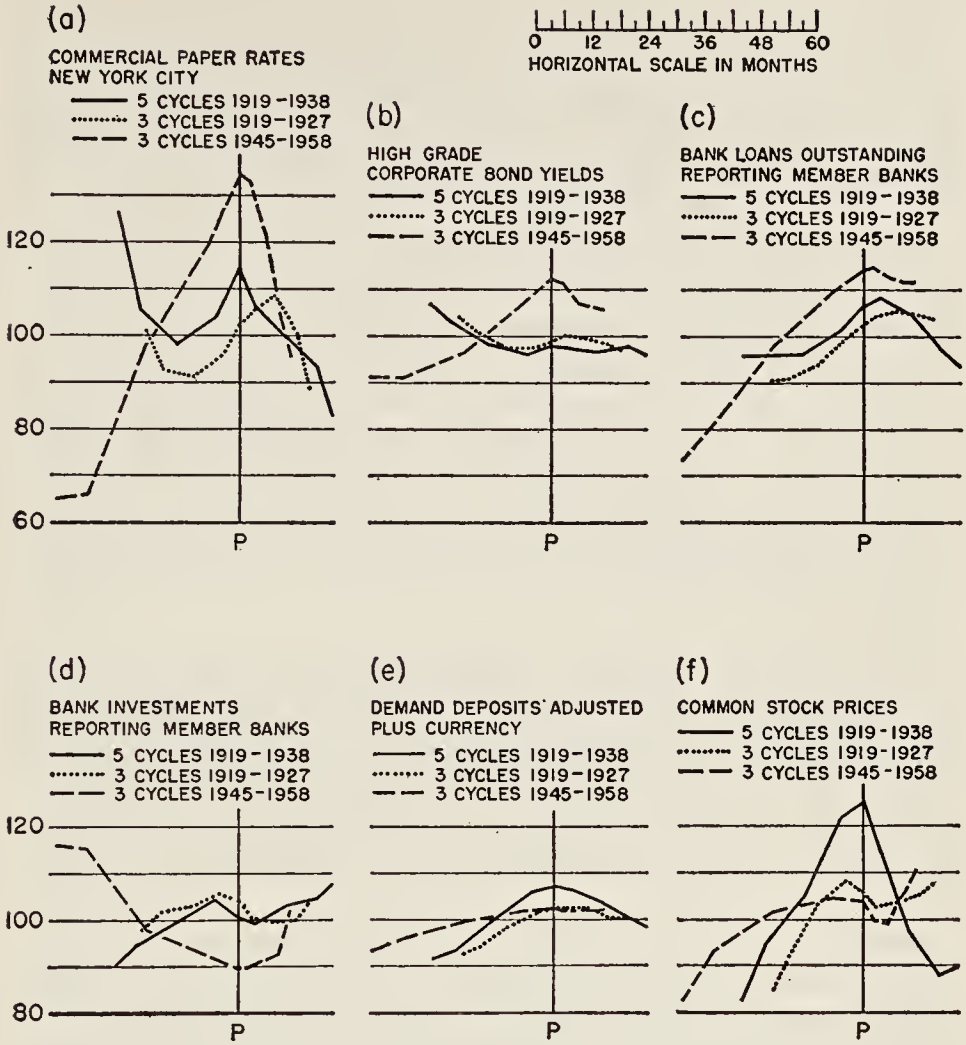


FIGURE 38. Average Reference-Cycle Patterns for Selected Monetary and Financial Series.

The source of these data is described in footnote 13.

in postwar than in prewar cycles. The greater cyclical variability of interest rates in the postwar period is a result of two factors. First, interest rates, on the average, have been lower since World War II than they were before the easy money policy inaugurated in the 1930's. Hence, a 1 percent rise in interest rates represents a larger increase in reference-cycle relatives now than it did in the 1920's or earlier. (These relatives are calculated by dividing the original figures by the average for the cycle.) Second, the absolute changes

themselves have been larger in postwar cycles than in most prewar cycles. Thus, in the postwar period, interest rates, while on the average not so high as they were before the 1930's, have been more sensitive to changing business conditions than in the interwar period. This contrast between prewar and postwar amplitudes is to be found also in the patterns for long-term bond yields (Figure 38b).

Before the war, both short-term and long-term interest rates showed a marked lag at the lower turning point, and, if we take only the period 1919–1927, a lag also at the upper turning point. In general, credit is likely to continue to be tight for a while after business turns down, when deteriorating expectations and declining sales induce a general scramble for liquidity. At the business-cycle trough, banks tend to have excess reserves, and business firms have generally put themselves in a liquid position. As a result, in the first stages of recovery, the demand for loans expands slowly; there is more than an adequate supply of funds; and interest rates may continue to fall for a while.

At least, this was the prewar pattern. These lags do not show up in the postwar patterns in Figure 38, except for the suggestion of a lag in bond yields at the trough and a failure for commercial paper rates to rise much in the first stage of expansion. The tendency for the prewar lags in interest rates to disappear is another reflection of the greater sensitivity of interest rates to which we have already referred.

Some of the reasons for this greater sensitivity is suggested by the other monetary patterns in Figure 38. Despite the vigor of postwar business expansions, the Federal Reserve authorities have kept a fairly tight rein on the money supply (Figure 38e). Expansion of the money supply has been moderate in business upswings and, on the average, has virtually ceased a full stage before the business cycle peak. In contrast, bank loans have expanded more vigorously in postwar than in prewar expansions (Figure 38c). The prewar lag at the peak in bank loans has continued into the postwar period, but the postwar lag is less pronounced.

Banks have not had much in the way of excess reserves since World War II; and, at least since 1951, the Federal Reserve authorities have not permitted a rapid expansion of reserves as loans have expanded rapidly during business upswings. As a result, commercial banks have had to sell securities throughout the expansion phase of the

cycle (Figure 38d). This is in marked contrast to the prewar pattern. During the two decades before World War II, banks could meet the rising demand for loans *and* add to their investments during the first two thirds of the upswing. Only at the end of the expansion and in the first stage of the contraction did they find it necessary to sell securities.

During business-cycle contractions, as the demand for loans falls off and the monetary authorities make credit more readily available, the banks increase their investments—and, in the course of doing so, help to push down interest rates (and raise bond prices). This is very much the postwar pattern. It was also the prewar pattern, except in the first stage of contraction.

What do all of these monetary patterns suggest as to the importance of monetary influences in the business cycle? One inference is that monetary factors may have been, on the average, somewhat more important after World War II than during 1919–1939. However, there are a number of reasons for doubting that monetary factors played a crucial role at either prewar or postwar turning points, although they may well have been contributing factors at particular peaks or troughs. First of all, most bank loans are made at interest rates that fluctuate much less widely than the short-term rates shown in Figure 38a. In addition, there is not much evidence that business generally is much hindered in late expansion by the unavailability of bank credit. And, at the trough, firms usually begin to expand output for reasons other than merely the existence of low interest rates and the greater availability of credit. Nor can most of the fluctuations in long-term investment be explained by swings in bond yields of the magnitude shown in Figure 38b. This is not to deny that monetary influences have some effect, more in some cycles than in others. In the postwar period particularly, for example, the greater availability of loanable funds in recession has been an important reason for the early upturn in residential construction shown in Figure 34b. But the fact that the money supply and interest rates fluctuate with the business cycle does not in itself demonstrate that booms and depressions have a monetary origin.

In this connection, we should be reminded by the equation of exchange that cyclical variations in total spending can go with changes in velocity as well as with fluctuations in the money supply. This, in general, turns out to be the case. Velocity of circulation of the money

supply (both income and transactions velocity) rises during business upswings and tends to fall when business is contracting. Thus, through changes in velocity, total spending can rise and fall even if there are no changes in the money supply.[48] This can happen not only during an individual business cycle but also over longer periods. Thus, the GNP rose much more between 1946 and 1960 than did the supply of money. The result was a substantial net increase in income velocity.[49]

FINANCIAL ACTIVITY

In the area of financial activity, the best known cyclical pattern is that of stock prices (Figure 38f). Many observers have commented on the regularity with which stock prices have tended to lead at the turning points in general business. This has been true at both peaks and troughs. A notable exception, however, was the sequence of events in 1929, when stock prices began their spectacular decline after the peak in business activity had been reached. There is an interesting resemblance between the stock price patterns for the first three cycles after World War I (1919–1927) and after World War II (1945–1958), allowing for the difference in length of the expansion and contraction phases in the two patterns. The general shape and upward tilt of the two patterns is much the same, and so are the leads that show up at the peak and trough.

New security issues, for which we do not present any patterns, go through wide cyclical fluctuations that are related both to the instability of private investment expenditures, on which we have frequently remarked, and to the course of stock and bond prices. The tendency for new stock issues to turn down before general business activity is undoubtedly related to the similar course of stock prices. Bond offerings (excluding refunding issues) show a somewhat different timing. In the late stages of expansion, the rise in bond flotations is likely to be halted by the rise in bond yields (fall in bond

[48] For the behavior of income velocity during reference cycles, see Milton Friedman, *The Demand for Money: Some Theoretical and Empirical Results,* National Bureau of Economic Research Occasional Paper 68, 1959, esp. p. 15. For further discussion of the behavior of velocity as well as numerous references to the literature, see R. T. Selden, "Monetary Velocity in the United States," in Milton Friedman (ed.), *Studies in the Quantity Theory of Money,* 1956.

[49] These developments were also associated with the upward trend in interest rates in the postwar period. This trend is reflected in the upward tilt of the interest-rate cycle patterns in Figure 38. For further discussion of monetary developments in the postwar period, see Chapter 19, below.

prices). Once the peak is passed and credit eases, bond issues may rise again briefly. Thereafter, the contraction in real investment leads to a decline in bond as well as stock offerings, despite a continued decline in bond yields. But as expectations begin eventually to improve, the attractiveness of very low bond yields may lead to some increase in the volume of bond issues in advance of the turning point in general business. If we add new bond and stock issues together, we find that the total usually leads general business at both peak and trough.

The role of the financial markets today is rather different from what it was before World War II. The stock market does not play the role that it once did, and changes in stock prices do not affect business sentiment to the same extent as formerly. Corporations are relying on retained earnings as a source of funds to a greater extent than before; and, when they do go into the market for outside capital, they are more likely to sell bonds than stocks. Further, these bonds may well be sold privately to insurance companies, which, along with pension funds, have come to control a steadily increasing proportion of the total savings of the economy. These and related developments may lead to different cyclical patterns for stock prices and security issues from those previously described. Perhaps more important, even when the patterns do not change significantly, their effect on the functioning of the economy may be quite different from what it once was.

THE NEXT STEP

We have offered brief comments on the significance of various patterns given in the preceding sections. But our aim in this chapter has been primarily to describe what typically seems to happen during business cycles in different parts of the economy, not to evaluate or to set up causal relationships. Thus far we have laid out the pieces of a complicated mosaic without attempting to fit them together. This analogy is a fitting one, for, until we fit the pieces together, we do not get from this array of patterns a clear picture of what happens during business cycles. In particular, we have not yet tried to assess the causal significance of the various types of cyclical behavior or of the relationships among them. This is the job of the next chapter, which—building on the findings of this and earlier chapters—will attempt to describe how the economy generates business cycles.

CHAPTER 11

HOW THE ECONOMY GENERATES BUSINESS CYCLES

WE SHALL now try to explain in some detail how a private-enterprise economy generates business cycles. We can start with several warnings. First, no two business cycles are exactly alike, and hence the general picture we are going to draw will not fit precisely any particular cycle that has occurred. This should not bother us, however. Our job now is not to describe or explain some particular period but to see what causal factors are common to most business cycles that have occurred in the past. We want to construct a general explanation of why business cycles occur in the sort of economy we live in.

This leads to our second warning. So far as we know, Russia is not bothered with what we call business cycles. The analysis of this and the other chapters of this book does not hold for an economy in which all the important economic decisions are made by a central authority. The more we move away from a system of private enterprise toward a centrally planned and controlled economy, the less relevant become the range of considerations that are stressed in this book. We are trying to find out why business cycles have occurred and still occur in the United States. Our conclusions apply to other countries only to the extent that their economies resemble that of this country.

These two warnings suggest a third. This chapter deals almost exclusively with the way in which the internal working of a private-enterprise economy creates the kind of instability we call business cycles. What we shall have to say represents a series of generalizations distilled from past experience. These generalizations therefore assume that the economy will respond to given stimuli in much the

same way that it has in the past. Structural changes in the economic system are always going on, and there is no question that the functioning of the American economy has been significantly affected by the political and economic developments of the last three decades. We are therefore not too sure in what ways the picture drawn in the following pages needs to be revised to explain future economic fluctuations.[1]

These considerations suggest that a few words would be in order concerning the role of external events and influences—so-called "exogenous" forces—in shaping the course of economic fluctuations. The following analysis largely ignores such influences, whereas in fact they always have played some role. Thus, any actual period of the past will show some economic changes that cannot be explained in terms of the self-generating mechanism described in the following pages; we must look also at the effects of government intervention, foreign developments, natural disasters, and the like. In the future, these external forces, particularly those resulting from various types of government action, are likely to play a more important role than they have in past periods (excluding wars) . And, as noted previously, government intervention not only may directly influence the level of economic activity but also may influence it indirectly through changing the way the economy responds to given stimuli. For example, tax legislation can change the way businessmen respond to new investment opportunities; or government action in the field of collective bargaining may change the way wages behave during the business cycle.

MAJOR AND MINOR CYCLES

Let us begin by noting again the distinction between major and minor cycles. The usual sort of cumulative process, in which business expansion (or contraction) feeds on itself but in which elements of vulnerability and some self-reversing forces gradually accumulate, operates during both major and minor cycles. Major upswings are periods of rising or high-level activity during which, even if there is a setback to business expectations, long-term investment opportunities continue favorable and therefore the self-correcting forces inherent in a short recession are sufficient to bring about a new recovery in fairly short order. This was the situation, for exam-

[1] See p. 337 below.

ple, in the mild recessions following World War II, as well as in the minor declines of 1924 and 1927.

In major downswings, on the other hand, long-term investment opportunities become seriously impaired; and hence a downswing, once it develops, is likely to be long and severe. There is no backlog of continuing investment opportunities to hold up the marginal efficiency of capital (to use Keynes' phrase). As a result, either the self-correcting forces inherent in a short recession are insufficient to generate any cumulative expansion at all or, if such an expansion does occur, the ensuing upswing is too weak to generate as high a level of employment as prevailed before the major downswing began.

Business cycles exist because a private-enterprise economy expands and contracts through a cumulative process that eventually breeds a movement in the opposite direction. In minor cycles, the self-reversing, cumulative process operates primarily through the short-period production planning of businessmen and the short-period purchasing plans of consumers. *Short-period* plans and expectations determine the course of the minor cycle. Major cycles, on the other hand, result from the long-term investment planning of businessmen and consumers. In studying minor cycles, we need to concentrate on short-period changes in price-cost relations, on maladjustments in particular industries capable of relatively quick correction, on the behavior of the short-term credit market, and particularly on changes in inventories of manufacturers and retailers. The study of major cycles would center on underlying investment opportunities, on changes in the rate of growth of the economy, on waves of speculative promotion of capital projects, on the effect of monetary and financial developments on the opportunities for profitable investment in long-term projects, and so on.[2]

History tells us that these two sets of cyclical forces are not always combined in the same way. Sometimes the long-term investment expansion proceeds at a moderate pace; financial excesses do not develop; and further investment opportunities arise as fast as old ones are exploited. In this case, the minor cycles are what chiefly meet the

[2] The distinction between major and minor cycles described here is very similar to that outlined by Alvin Hansen in *Fiscal Policy and Business Cycles* (1941), pp. 16–19, and *Business Cycles and National Income* (1951), Chap. 2. Schumpeter similarly distinguishes between "Kitchin" (minor) and "Juglar" (major) cycles. See also R. A. Gordon, "Investment Behavior and Business Cycles," *Review of Economics and Statistics*, vol. 37, February, 1955, pp. 23–34.

eye. The major may be scarcely distinguishable from the underlying trend. This was the case, for example, in the United States and most other countries in the prolonged period of prosperity that followed World War II. In other cases, the underlying investment boom may be so rapid that minor maladjustments do not have a chance to lead to widespread hesitation before the final collapse comes. In this case, only the major cycle would stand out. Most often in the United States, at least before World War II, we find the two sorts of cycles operating together and both distinguishable. Where the major cycles have stood out clearly, they seem to have included from two to four minor cycles.

Major cycles do not always have a chance fully to work themselves out. We may, on occasion, get either *incomplete* or *overlapping* major cycles. An incomplete major cycle would consist of an underlying expansion phase that did not continue long enough to generate a major depression before a new set of investment stimuli took hold and created a new major upswing. This was the case in the United States in the 15 years or so after 1933. One set of major-cycle influences led to a slow and halting expansion from 1933 to 1940 (interrupted by the minor recession of 1938), after which a new set of major-cycle forces geared to military expenditures took hold. There was no intervening major depression.

The concept of overlapping cycles differs only in degree. In this case, a major expansion does continue long enough to generate a decline of some severity. But during this cycle, new investment stimuli may be accumulating and gaining strength, and they may take hold quickly enough to cut short the decline and lead to a prompt and vigorous recovery. Thus, a new major cycle may overlap the old. This was apparently the case at the end of both World War I and World War II. The termination of military expenditures and the collapse of the postwar inflationary boom brought on the sharp decline of 1920–1921. But already the stimulating influence of the automobile, the rapid spread of electric power, and the pent-up demand for housing were so strong (creating new investment opportunities) that the depression was short. In 1945–1946, the sharp drop in military expenditures was succeeded so quickly by an upsurge of private spending that only the mildest sort of recession resulted.[3]

Keeping in mind this distinction between major and minor

[3] These two postwar episodes are discussed in some detail in Chapters 14 and 15.

cycles, let us go on to study in more detail how the economy generates both types of cyclical fluctuations.

WHY AND HOW THE UPSWING BEGINS

Let us assume that a major business contraction has been under way for, say, two years or so. Production has been declining, unemployment has been mounting, bank credit has been contracting, long-term investment is at a very low level, the stock market is in the doldrums and security prices are badly depressed compared to previous boom levels, and extreme pessimism—fed by business losses and numerous bankruptcies—reigns over the business scene.

In the terms used in Chapters 4–6, the marginal efficiency schedule of investment has been shifting downward during the depression; the low level of investment, through the multiplier, has forced down consumption and income; hence the level of aggregate demand is low, with resulting unemployment and excess capacity widespread through the economy.

Now we have to ask: What brings this sorry state of affairs to an end?

In the past, the patient has usually cured himself without much outside help.[4] Every business contraction gradually sets off self-correcting forces which gather strength as the downswing progresses. The economy, through the way it responds to deflationary forces, tends gradually to eliminate many of the maladjustments that brought on the depression. As we should expect, it is the minor-cycle, short-run type of maladjustment that the economy can most easily correct.

LIQUIDATION OF INVENTORIES

This is well illustrated by the behavior of inventories. As the downswing progresses, we find that production falls faster than consumption and retail sales. (See Chapter 10.) Hence, inventories—both in trade and in manufacturing—eventually begin to decline. But this cannot go on forever. The sharper the decline in production and the slower the decline in sales, the sooner must the decline in inventories end. Eventually, firms slacken the pace at which they

[4] This was particularly true before 1929. In 1933, the federal government intervened actively to stimulate recovery, and we may confidently expect that it will do so again in future major depressions. The government has also intervened to some extent in the minor cycles since World War II.

are reducing inventories. The rate of decrease in inventories begins to diminish and eventually comes to an end. Even if total inventories do not increase for a while, the tapering off in the rate of decline is a stimulating factor. A given rate of decline in inventories represents so much negative investment, or disinvestment. Businessmen are producing less than they sell; hence, sales proceeds are in part being hoarded (or used to repay bank loans) instead of being returned to the income stream. If the rate of decline in inventories diminishes, there is less disinvestment and less hoarding than before.[5]

IMPROVED LIQUIDITY OF FIRMS AND BANKS

During business contractions, virtually all firms try to get themselves into a more liquid financial position. They prefer cash to goods, and this is why they try to reduce their inventories. As the downswing goes on, businessmen pay off their bank loans and gradually convert their inventories and accounts receivable into cash. At the same time, by holding back on the replacement of old equipment, they improve their cash position still further. Thus, the decline leads to a general improvement in liquidity; firms find that their cash resources are proving sufficient to weather the storm; gradually the scramble for still more liquidity abates. And to the degree that firms stop trying to acquire more cash and no longer need to repay loans, to that extent they become more willing to spend—particularly in replacing depleted inventories and worn-out equipment.[6]

[5] As we saw in Chapter 10, total inventories lag behind the turning points in general business, but the rate of change (investment) in inventories tends to coincide with or lead the turns in business activity. In this connection, it is important to distinguish between planned and unplanned accumulations of inventories. Total inventories may rise or fall because sales turn out to be less or more than expected and not because businessmen plan it that way. During a downswing, inventories are slow to fall at first, partly because of such unplanned accumulations. Conversely, at the end of a downswing, businessmen may try to increase inventories but may find that their inventories continue to decline for a while because they do not fully anticipate the revival in sales that takes place. In this case, the further unplanned decline in inventories makes businessmen all the more anxious to increase output.

[6] For a similar analysis, see W. C. Mitchell, *What Happens During Business Cycles,* 1951, chap. 7. Cf. also G. Haberler, *Prosperity and Depression,* 4th ed., 1958, pp. 392–393. Mitchell cites the fact that the firms in a sample of large manufacturing corporations studied by the National Bureau tended to hold their largest cash balances at the trough of the cycle.

The improved liquidity of firms is matched by the more liquid position of the commercial banks. As bank loans are paid off and demand deposits decline, the banks begin to accumulate excess reserves—a process that is accelerated by the return of currency from circulation as the public's need for cash declines. In addition, there may be an inflow of gold from abroad. As their reserve position continues to improve, the banks try to find additional earning assets. Since business demand for loans is at a low ebb, the banks purchase securities and invest more in mortgages. Bond prices rise, interest rates decline, and bank credit becomes more readily available.[7] Not only the banks but life insurance companies and other large institutional investors find themselves with excess funds to invest, while the business demand for such funds remains depressed. This causes interest rates to fall further. Short-term rates decline more than long-term rates. Low interest rates induce the larger corporations to refund outstanding securities, since they can pay off old securities bearing high rates of interest with new securities bearing lower rates. Fixed charges are thereby reduced, with a consequent improvement in financial position. The builders and buyers of homes are also important beneficiaries, since mortgage loans are now more readily available and on easier terms.

THE STOCK MARKET

Usually, stock prices begin to rise before the upturn in business. (See Figure 38.) As the downswing progresses, the greater liquidity of the economy and the growing accumulation of idle funds eventually tempt some investors to buy common stocks. The eventual slackening in the pace of the business decline and in the volume of business failures tends to restore confidence in the future of stock prices. The pessimism that had driven stock prices down to bargain levels thus tends to abate somewhat. With this slight improvement in confidence, some funds begin to move into the stock market. The resulting rise in stock prices, in turn, helps to ameliorate the gloom in other parts of the economy.

PRICES AND COSTS

Typically, prices continue to decline throughout the business contraction, although this has been less true since World War II than

[7] See the reference-cycle patterns in Figure 38.

before. During the early part of the downswing, the fact that different kinds of prices decline at different rates adds to the maladjustments from which the economy is suffering. But as the decline proceeds, some self-correcting forces come into operation. Wholesale prices of raw materials and semifinished products fall faster than those of finished goods. This tends to improve profit margins, though at first it may not be enough to offset the increase in unit costs resulting from the spreading of fixed costs over a smaller output. Probably more important is the reciprocal relation between inventory holdings and price changes. In the early stages of the decline, efforts to reduce inventories force some types of prices down rapidly, and the fear of further price declines leads businessmen to cut back on their production and purchases still more. Eventually, as the scramble for liquidity abates somewhat, the pressure on prices tends to ease. Prices continue to decline, perhaps, but at a slower rate. Businessmen are less apprehensive regarding further losses from lower prices, and they may now begin hesitantly to replenish particular inventories that are too low and perhaps to be attracted by bargains in the form of low-priced materials and equipment.

Labor costs also eventually decline during the downswing, though usually not by as much as wholesale prices of manufactured goods. Labor productivity rises, particularly toward the end of the contraction, and some decline in wage rates may take place. In manufacturing, this decline in labor costs is not likely to have much of a stimulating effect until prices stop declining. From then on, profit margins are likely to begin to improve in a number of lines. And in areas such as the building industry, significantly lower material and labor costs may prove to be an important stimulating factor.

CORRECTION OF HORIZONTAL MALADJUSTMENTS

Some business contractions are started or accentuated by a condition of overproduction in particular industries. Businessmen meet such a situation in the obvious way—they try to produce less than they sell, thereby cutting down their swollen inventories, and they call off plans for further expansion of capacity. Sooner or later, depending on the nature of competitive conditions in the industry, they will also reduce prices. If the maladjustment is not serious, a sharp cutback in production, some reduction in prices, and perhaps

the bankruptcy of a few marginal firms may be enough to bring supply back into a profitable relation to demand. In this case, the industry in question may soon expand production again and be in the van of a new revival in business. But if the trouble is more serious—if demand has more or less permanently fallen or if much excess capacity was built during the preceding boom—this industry may continue to stagnate even after other industries begin to revive.

THE STABILIZING BEHAVIOR OF CONSUMPTION

We have really covered this point in our discussion of inventories. Consumption does not fall as much as production and national income during the downswing. Disposable income declines less rapidly than GNP, and consumption falls more slowly still. There are a number of reasons why disposable income falls more slowly than total output: business payments of wages, interest, and dividends decline less than value of output; income taxes fall more than incomes; government transfer payments in the form of unemployment insurance, relief, and farm-price supports actually rise. Also, consumers spend a larger percentage of their disposable incomes in depressions. Personal savings are reduced sharply, and many people find it necessary to dishoard—to spend not only all of their reduced current incomes but also some part or all of their accumulated savings. The resulting floor under consumption gives businessmen the opportunity to carry out the reduction in inventories previously discussed. As we saw in Chapter 10, consumption did not, on the average, fall at all in the first three postwar recessions, and it began to rise before general recovery set in. This was obviously an important factor limiting the severity of postwar recessions.[8]

LONG-TERM INVESTMENT

We began this section by assuming that the economy was experiencing a *major* business decline—that for one reason or another there had been a serious deterioration in long-run investment opportunities in business plant and equipment and perhaps also in housing. If this is the case, the self-correcting forces thus far discussed are not strong enough to lead to a new major expansion.

[8] As noted on p. 191, rapid population growth raises the relation of consumption to income. This was another favorable factor operating on consumption in postwar recessions. Cf. James Duesenberry, *Business Cycles and Economic Growth,* 1958, p. 265, and B. G. Hickman, *Growth and Stability of the Postwar Economy,* 1960, chap. 10.

Without an improvement in long-term investment opportunities, a revival in business would be relatively short-lived. At best, the next boom would end long before full employment was reached. We must therefore consider whether a major business decline is likely to do anything to remove the obstacles to long-term investment that helped to bring on a depression as severe as the one that we are assuming. This depends on what was originally responsible for the decline in long-term investment.

The first possibility to be discussed plays an important role in some theories of the business cycle. If high interest rates and a shortage of loanable funds were responsible for the initial decline in investment at the preceding peak, then the decline in interest rates and the easing of credit which take place on the downswing should create an increasingly favorable environment for renewed investment in long-term projects. And to some extent this is the case. As interest rates decline and as banks, insurance companies, and other large investors bid up bond prices in their attempt to put idle funds to work, firms that have plans for further expansion become more and more tempted to take advantage of the current low interest rates. Reduced building costs and lower equipment prices work in the same direction.

But we must not exaggerate the stimulating effect of these developments. Any businessman would rather borrow at a low than a high rate of interest, other things being the same. But he must first feel confident that he can use the borrowed funds profitably and that he can repay the principal when due. *Low interest rates cannot stimulate investment if the demand for loanable funds, even at these low rates, does not exist.* The essence of the downswing is the collapse in the business community's willingness to invest. In Keynes' terms, the marginal efficiency schedule of capital shifts downward, resulting in a declining volume of investment. For the volume of investment finally to rise again, one of two things must happen: (1) The marginal efficiency schedule stops falling; and, though it does not at first shift upward again, a reduction in interest rates leads to increased investment because the demand for loanable funds is sensitive to reduced interest rates.[9] (2) With or without a

[9] It is possible that, in the later stages of a business contraction, lenders become more willing to take risks and to lend at existing interest rates. This relaxation of stringent credit standards is likely to be more stimulating that a mere reduction in interest rates.

reduction in interest rates, the demand for loanable funds finally begins to shift upward because businessmen for some reason become more willing to invest even at the same interest rates. Of these two possibilities, the latter is the more important. The demand for loanable funds does not seem, for most types of business investment, to be very elastic with respect to interest rates. The really critical issue is: Are there any reasons why the marginal efficiency schedule will stop shifting downward and eventually begin to rise? Luckily, the answer to this question seems usually to be in the affirmative. We now proceed to consider some of the more important possibilities.[10]

One extremely significant factor affecting the length and severity of a depression is the state of the building industry, particularly residential building. If the underlying demand for houses and other buildings is still strong, the reduction in interest rates and building costs may by themselves be enough to lead to an early revival in building activity. This happened in the first three downswings after World War II. With a strong underlying demand for houses and other types of buildings, the decline in interest rates and the greater availability of mortgage credit, after the business decline had been under way for a short while, led to a prompt revival in building activity. This was one of the important reasons for the mildness of these postwar recessions.[11]

If, however, there was serious overbuilding during the preceding boom, with the result that the economy finds itself in the declining phase of a long building cycle, the situation is much more serious. Lower interest rates and building costs are not enough to stimulate much new construction. In this case, the necessary readjustments are much more painful and protracted. We must wait for population growth and migration and for depreciation and obsolescence to eliminate excess capacity and to create again favorable conditions for new building. If, on top of this, the mortgage market has become disorganized because of wholesale foreclosures (resulting from unwise lending during the preceding boom), still further readjustments are necessary. Thus, in the depression of the 1930's the American building industry had to undergo all of these painful readjust-

[10] See also the discussion of the determinants of investment in Chapter 6.

[11] See the discussion of the cycle patterns in construction in Chapter 10 and the detailed treatment of the postwar period in Chapters 15–16.

ments, which were not complete even when business finally began to revive in 1933.

In the field of plant and equipment expenditures, self-correcting forces are also at work during business contractions. For one thing, firms sharply reduce their expenditures on replacement. Since nearly all firms have excess capacity, equipment is not replaced as it wears out. In addition, old equipment continues to be used even after it becomes obsolescent because of new improvements that continue to be introduced. Obviously, this cannot go on forever. Replacement needs accumulate and eventually must be satisfied. When this occurs, gross investment rises.[12]

Not only businessmen but also consumers accumulate replacement needs during the depression. Clothes wear out fairly soon. Eventually, so do more durable goods, of which the most important are automobiles. As replacement needs become more urgent, consumers become more willing to spend accumulated savings or to go into debt. And as old debt incurred during the preceding boom is paid off, more of current income is available for spending, and sellers become more willing to extend new credit. Of course, those who have lost their jobs will not find it easy to get credit, but if the need is urgent enough even unemployed persons may find means (perhaps through borrowing from relatives or friends) to replace essential items that have worn out.

At the same time that replacement needs are accumulating, industries that had expanded too vigorously during the preceding boom are in the course of working off their excess capacity. In new industries in which there is still much room for growth, increasing acceptance of the new product may lead to a quick revival of sales and eventually to the need for further investment. The continual growth of population also helps to absorb excess capacity in both new and old industries.

New industries, which are still in the rapidly expanding phase of their long-run growth curves, probably play an important role in bringing major depressions to an end. The more marked the underlying trend and the less the overcapacity inherited from the preceding boom, the sooner do relatively young and expanding industries find it profitable to increase their expenditures on new plant and

[12] The sooner firms satisfy their depression-born desire for liquidity, the sooner are they likely to begin to meet accumulated replacement needs.

equipment, even if general business conditions remain depressed. This was particularly important in the case of the railroad industry in the latter half of the nineteenth century and of the automobile and electric power industries in the first quarter or so of the present century. New and rapidly growing lines have also played an important role in maintaining investment opportunities since World War II.

THE FINAL IMPETUS TO REVIVAL

All of the factors we have thus far discussed make the economy increasingly susceptible to favorable developments; thus it becomes more and more likely that some stimulus will touch off an expansionary movement that will "catch hold" and develop into a new cyclical upswing. The actual starter of the revival may be one of these self-correcting forces already described, or it may be some other development not yet discussed.

Of the self-correcting forces mentioned, the liquidation of inventories and the failure of consumption to fall as much as production are most likely to be the active forces in initiating recovery after a minor recession.[13] The easing of credit and generally improved liquidity help to make this possible. But such a revival is not likely to carry very far unless conditions are favorable for an expansion of long-term investment. If the economy is in a major depression, the need to replace inventories and other "minor-cycle stimuli" may start the revival, but the factors operating on the level of long-term investment must be sufficiently favorable so that expenditures on construction and producers' durable goods begin to rise promptly, also.

On the side of long-term investment, the most important starters are likely to be the accumulated need to replace worn-out and obsolescent equipment and the absorption of excess capacity in new industries so that further expansion becomes possible. Financial developments may aid in renewing expansion in new industries. The reorganization of weak firms and weak financial institutions helps to restore confidence and to make investors and businessmen again

[13] See p. 306 above; also the evidence in F. C. Mills, *Price-Quantity Interactions in Business Cycles,* 1946, suggesting that consumers'-goods production turns up before general business. See also the reference-cycle patterns for consumers' expenditures in Figure 37.

willing to exploit the opportunities existing in these expanding industries. The early rise in the stock market also generates confidence and makes easier the flotation of new securities.

Sometimes, where the depression occurs during the upper part of a building cycle, residential construction will take the lead on the upturn. In this case, lower interest rates and building costs plus the continued high level of demand for shelter set off the new expansion in construction.[14]

All of the active causes of revival thus far discussed may be called *endogenous* influences, arising from within the ordinary working of the economic system. But "outside" or *exogenous* forces may also be the immediate cause of revival. Whether an exogenous force is strong enough to turn a business contraction into expansion, however, depends on whether the endogenous, self-correcting forces mentioned earlier have done their job well enough so that the economy is ripe for recovery.

These exogenous starters may be of various sorts. An obvious one is a substantial increase in government spending—for example, on armaments or public works. Or the impetus may come from abroad, in the form of an increased foreign demand for a country's exports. Another possible influence, which calls for brief discussion here, is one or more important innovations that have reached the stage of commercial application since the preceding boom.

A number of economists emphasize the role of innovations in setting off cyclical expansions.[15] They do not always make it clear, however, whether they are referring to new or old innovations. If we define a *new* innovation as one that has reached the stage of commercial application since the last cyclical peak, then we may say with some confidence that new innovations are seldom responsible for ending depressions. When a process or product is this new, it is not likely to generate enough investment, compared to that accounted for by older methods, to make much difference in total capital formation. So far as innovations are concerned, it is primarily the "carry-over" effect of older innovations—still new enough to call for a rapid rate of secular expansion but old enough so that this

[14] See p. 276, where we discussed the lead typically shown by building activity at the trough in general business.

[15] Notably J. A. Schumpeter. In this connection, see the survey of business-cycle theories in Chapter 12. Note also what was said in Chapter 8 regarding the impact of technological change on economic stability.

expansion requires a large absolute amount of investment—which helps to initiate a revival in long-term investment after major downswings.

THE RECOVERY AND UPSWING

Thus, for one or more of a number of possible reasons, a new upswing begins. Stock prices have perhaps been rising for several months. Some sensitive wholesale prices have started to increase. Particularly in industries in which inventory policy is sensitive to price expectations, or in which consumption has been well maintained, production increases as firms stop liquidating inventories and perhaps even try to increase them. Thus total output, and with it employment and incomes, begins to rise. In the van of this initial expansion are likely to be some of the consumer's nondurable-goods industries—for example, various food and textile lines.

Equally important, long-term investment begins to increase. Perhaps residential building turned the corner some months back. Equipment manufacturers notice an increase in orders as firms begin to satisfy their accumulated replacement needs. Some relatively young and expanding industries, finding their sales again rising, begin to expand capacity again. Confidence spreads. New security offerings increase. More firms place orders for new plant and equipment, either for replacement or to carry out expansion programs that had been put on the shelf during the preceding decline. Thus, total investment rises and, through the multiplier process, generates additional employment and incomes. Consumers also add their bit to the expansionary forces at work as they rapidly increase their purchases of consumers' durables, especially automobiles. Such purchases are important because they resemble business investment in their effects on employment and incomes. The purchase of automobiles or furniture (and this is of course even more true of houses) ordinarily involves going into debt or the use of past accumulations of cash; the consumer is putting back into the income stream more than he receives as current income.

Now the multiplier takes full hold, raising incomes, employment, and consumption. Also, prices are by now rising generally as businessmen order ahead. Confidence breeds further confidence, and this induces further investment. The familiar cumulative process is in full swing, gathering strength as it proceeds.

At the beginning of the expansion, firms were likely to have financed their initial needs from their own liquid resources. Soon, however, they begin to borrow from the banks. With excess reserves and growing confidence, the banks are increasingly willing to lend. Thus, bank credit expands and helps to finance an excess of *ex ante* investment over saving. As long-term investment expands, so too does the flotation of new securities. At first, chiefly bonds are sold; but later, as interest rates begin to rise and as the public's confidence in stocks continues to improve, the amount of stock issues offered expands rapidly.

We have assumed that a *major* expansion is under way, i.e., that businessmen are able to generate sufficient investment programs (and can find the necessary funds at a reasonable cost) to support a steadily rising level of long-term investment. This means a high level of replacement expenditures, a substantial amount of residential building, rapidly expanding capacity in important new industries, and slower expansion in old industries in the form of additions to and modernization of capacity.

As output expands, firms feel the need for larger inventories. At first, their attempts to increase their stocks may be partly defeated by the rapid rise in sales; but eventually total inventories begin to rise, thereby providing a further stimulus to business generally.

As we already know, the expansion in output is particularly marked in the capital- and durable-goods industries. Investment rises relatively much more rapidly than consumption; and the flow of consumers' durable goods rises much faster relatively than does the output of nondurable goods. We saw in Chapter 6 why this is so. Even though the acceleration principle does not hold in any precise way, the durability of capital goods and of such consumers' goods as automobiles means that the output of these commodities must rise more rapidly than the current flow of services that they are expected to supply.

As the expansion proceeds, it is likely to take a speculative turn here and there. Speculation operates particularly through three channels: commodity prices, prices of existing securities, and promotion of new projects to exploit what seem to be profitable investment opportunities. Speculative increases in commodity prices, expecially of raw materials and farm products, may come quite early in the upswing. As a matter of fact, many textile and food lines

that experience these early increases in raw-material prices may run into a temporary setback to production during the first half of the expansion.[16] Later in the expansion, speculative price increases may become more general, particularly when output has expanded to the point where pressure is put on capacity in various lines and bottlenecks begin to appear. As the expansion proceeds and prices continue to rise, businessmen increase their outstanding purchasing orders and order farther ahead; they attempt to increase inventories; and manufacturers find their backlogs of unfilled orders growing rapidly.

Speculation in stock prices proceeds through the entire business expansion, though the advance in prices is likely to be occasionally interrupted by minor setbacks. The rise in stock prices results from two sets of forces: higher earnings per share as profits rise rapidly, and a steady increase in the rate at which these earnings are capitalized.[17] The latter influence represents growing confidence that profits and stock prices will continue to rise. This growth in confidence spreads to banks and other lenders, who increasingly accept risks, in their loans and security purchases, that they would have refused to take before the expansion began.

The gradual change from conservative optimism into speculative exuberance also appears among promoters and businessmen. New companies are formed, frequently financed on the proverbial shoestring, ranging from neighborhood retail shops to large undertakings involving millions of dollars. Speculative builders push ahead with new housing projects and land subdivisions. Real-estate prices, on both farm and urban properties, rise; and rapidly growing areas may experience a marked real-estate boom. There is a substantial increase in the volume of new mortgages, and some relaxation in the terms on which they are offered. In new industries, where profit prospects are particularly bright, promoters and venturesome businessmen—with their own and borrowed capital—rush in to form new companies, frequently without adequate assessment of market potentialities and without sufficient resources to weather future storms. Capacity in these industries expands rapidly—too rapidly, as later events are likely to prove.

[16] Compare Mills, *op. cit.*, esp. p. 92.

[17] Thus in depression investors and speculators may be willing to pay only $20 per share (or less) for the stock of a company with earnings per share of $2.00, but they may offer $45 per share when earnings rise to $3.00 per share. In this case, the "times-earnings" ratio has advanced from ten to fifteen.

Eventually, if the expansion goes on long enough, full employment is reached. By now security and commodity prices are likely to be rising rapidly. Bottlenecks appear in some industries as shortages of particular types of labor, capacity, or materials begin to develop. The slow rise in retail prices is accelerated. Labor becomes more insistent in its demand for higher wages; and, as full employment is approached, wage rates rise rapidly. If aggregate demand grows still further, price and cost increases will be accelerated; and the continued expansion of demand will dissipate itself in merely higher prices with little further increase in output.[18] If this stage is reached, and sometimes even earlier, symptoms of marked credit stringency will appear. But most peacetime booms end before or about the time full employment is reached, and a full-employment, "demand-pull" type of inflation (such as those immediately after World Wars I and II) does not develop.[19]

WHY PROSPERITY ENDS

It is surprising how quickly a cyclical expansion begins to sow the seeds of future trouble. One evidence of this is the increasing number of series that begin to turn down well before the peak in general business is reached.[20] Increasingly, as the expansion goes on, maladjustments accumulate that make the economy vulnerable to possible future unfavorable developments. If these maladjustments are serious enough, they will eventually bring on a cyclical reversal in business without the help of external disturbances.

MINOR MALADJUSTMENTS

Minor maladjustments, serious enough to give rise to a relatively short cyclical reversal, are particularly likely to occur, and the un-

[18] We can restate this in the terms used in Chapter 4. As the aggregate demand function shifts upward, the aggregate supply curve begins to tilt upward (become steeper) under the pressure of higher wages. Thus costs and prices rise. In addition, output may expand to the point where the aggregate supply function becomes very inelastic. Compare Figures 9 and 10.

[19] The prolonged boom of the 1920's does not fit this description very well. After the 1921 depression, approximately full employment was reached by 1923. From 1923 to 1929 the economy operated at close to full employment, with mild declines in 1924 and 1927. From 1923 to 1929, however, prices tended to sag; wage rates did not rise as much as labor productivity; and in general there was little evidence of labor, material, or capacity shortages. Stock prices, of course, did rise markedly.

[20] See G. H. Moore, ed., *Business Cycle Indicators,* 1960. See also the discussion of "diffusion indices," pp. 519–522, below.

derlying major expansion may be broken by one or two such minor downswings. Prices may rise so rapidly that businessmen soon begin to distrust that the expansion will continue. Price expectations take a turn for the worse. Firms stop ordering ahead and attempt to reduce their commitments and inventories. As a result, production declines; the general decrease in business demand does cause prices to fall; and a cumulative downswing gets under way. Or a tendency for the rise in sales to flatten out, for whatever reason, may reduce the need to expand inventories further, with the result that inventory investment and production begin to decline. Other short-period stresses may play a role. A too rapid rise in the stock market may breed pessimistic expectations. Or the banks, also concerned with increases in prices and inventories, may tighten credit. (They may also be forced or induced to restrict loans by action of the monetary authorities.) Another possible source of disturbance is that one or more important industries may find they are expanding output too rapidly relative to the rest of the economy; they are trying for a larger share of the consumer's dollar than they are able to get. Possibly, also, sharp increases in some costs—of raw materials or of labor—impair profit margins. (This has probably been more true since World War II than before, as we suggested in Chapter 10.)

The maladjustments mentioned in the preceding paragraph have one thing in common. They are all short-term in nature, capable of being corrected by a relatively brief curtailment in output and decline in costs and prices. Long-term investment will also decline as expectations become pessimistic for the time being, but the main decline in capital formation in such minor recessions is in inventory accumulation rather than in plant or equipment. We assume that long-term investment opportunities remain favorable. As a result, the reduction in inventories and orders, accompanied by some liquidation of bank credit and decline in interest rates and stock prices, is sufficient to restore confidence and lead to a resumption of the underlying business expansion. In such minor swings, the decline in inventory investment is likely to be greater than in any other type of capital formation. This is why a number of economists refer to minor cycles as primarily an inventory phenomenon.

We now want to consider the more serious maladjustments that may arise as the major expansion carries the economy up close to full employment. The boom may continue after full employment is

reached, but from then on further increases in output are limited to what can be achieved through improving labor productivity (particularly as new plant and equipment are added), through slow additions to the labor force, and through labor's working overtime. If aggregate demand continues to rise rapidly after full employment is achieved, the main effect will be on prices rather than output.

The same minor maladjustments mentioned before may herald the end of the major expansion. The longer the upswing has gone on without a minor recession, the more serious these maladjustments are likely to become. But here we want to look at the more fundamental strains that are developing—the ones that involve long-term investment. There are several possibilities.

PARTIAL OVERINVESTMENT IN NEW LINES

Perhaps the most important concerns the behavior of important new industries. These may create difficulties in one of two ways. First, they may have been expanding too rapidly, with the result that excess capacity has been built up. Eventually this fact is discovered. Some firms find that there is no market for their new capacity once it is installed; some new firms, speculatively promoted and financed, may fail; inventories may begin to accumulate as the additions to capacity are put into operation. Thus, an industry which has been growing rapidly—and which as a matter of fact still has a considerable amount of growth ahead of it in the long run—may have to mark time for a while to permit demand for its products to catch up with capacity. But if capacity ceases to expand, that means that investment has declined. If the industry or group of industries is important enough, the effects can be serious. The decline in investment means a reduction in incomes and consumption. The decline in orders for capital goods will probably lead to a further decline in investment in the capital-goods and raw-material industries that were supplying the industry in question. If much unwise financing has gone into the expansion of the industry, the solvency of some financial institutions may be put in jeopardy, and the industry's difficulties may be quickly reflected in the financial markets. If the industry is important enough, its difficulties may be sufficiently serious to start a major downswing. The severity of the depression will depend on the size of the industry and the degree of

overcapacity, on how involved the banks and other financial institutions are in its difficulties, and on how vulnerable the economy is in other respects.

There is a second way in which new industries may cause trouble, even if marked excess capacity does not develop.[21] They may reach maturity, so that expansion at the former rate may no longer be possible. And a decline in the rate of expansion means an absolute decline in the volume of investment, unless it is offset by rising replacement needs and by further technological change which calls for new investment. This is the long-run working of the acceleration principle, the importance of which we stressed in Chapter 6.

Thus, the slowing down of expansion in new industries may lead to a decline in investment opportunities. In a growing economy that is undergoing rapid technological change, this kind of difficulty is always occurring. Innovations create new investment opportunities and cause total investment to rise. But once the economy has adapted itself to the new products and new processes in question, investment in other lines must expand promptly if a decline in total investment is to be forestalled. Unfortunately, a private-enterprise economy ordinarily finds it difficult to make such adjustments in time. In the first place, the decline in investment in one part of the economy sets off a chain of deflationary influences that reduces incomes and leads to a deterioration of business expectations in other parts of the economy, with the result that investment in other areas falls rather than rises. Secondly, if investment in other industries is to rise, interest rates must fall; and there must be a general readjustment of prices and costs through the economy so that investment will become more profitable in other lines. Or else investment opportunities must be expanding in some industries for other reasons—for example, because of new technological developments. Our economic system is ordinarily too inflexible, and the factors of production are too immobile, for these price and cost adjustments to take place promptly enough to offset the effects of the

[21] Rapid expansion by a new industry—or the rapid adoption of new techniques by some firms in an old industry—can create difficulties in yet a third way, which is emphasized by Schumpeter. The innovating firms, with their increased output and lower costs, create losses for old firms and cause the latter eventually to curtail output and investment.

decline in investment in the industries that have run into difficulties.[22]

BUILDING ACTIVITY

Another type of investment is likely to experience difficulties in the later stages of some major expansions. We refer to the behavior of building activity, particularly residential and commercial construction. Total private construction of all types amounted to more than half of total private investment during 1955–1960. Of the total value of private building, about two thirds represented residential and commercial building. The rest constituted chiefly industrial and public-utility construction, i.e., construction of factory buildings, power plants, and other structures in various types of industrial operations, in contrast with "commercial building," which includes office buildings, warehouses, garages, stores, etc. We have already talked about industrial building in discussing the behavior of investment in particular industries. As we should expect, industrial construction moves very closely with the major cyclical swings in general business.

It is primarily in residential and commercial construction, particularly the former, that we find the long cycles in building to which we referred in Chapter 9. Special forces operate to create long swings of great amplitude in such building. Houses and apartment and commercial buildings are among the most durable of capital goods, and hence they wear out slowly. If excess capacity develops, building may remain at a very low level for a number of years before a slowly expanding population and deterioration in the existing stock of structures again make building profitable on a large scale. Once a building boom does begin, it may continue for five to ten years or more—which is longer than major cyclical upswings ordi-

[22] In the first 15 years after World War II, the American economy was extremely fortunate in that, as one type of investment started to decline, investment in some other sector expanded so as to provide an offset. This was chiefly because the overall stock of investment opportunities continued to be so large, being continually replenished by population growth and migration, continued technological change, rising defense expenditures, and so on. In addition, no innovation or industry (except possibly federal defense activities) dominated investment behavior in the way the railroads did in an earlier day. Hence, the economy was not so vulnerable to a decline in investment opportunities in any one industry that was approaching maturity.

narily last. But such building booms have in the past always come to an end. Building is a field in which speculative optimism is likely to run rampant; and the fact that considerable time usually elapses between initial planning and final completion means that optimism is particularly likely to lead to mistakes. Even if substantial overbuilding does not occur, building activity must eventually decline for reasons suggested by the acceleration principle. Whether the rise in building activity was initiated by the accumulation of replacement needs, by a spurt in population growth, or by the shift of population to new areas, the increased demand will eventually be satisfied. Since, at least in the United States, building booms are accompanied by speculation in real estate and much unwise lending and borrowing, the demand is likely to be more than satisfied. When this happens, a powerful deflationary force is let loose. One of the most important components of total investment begins to decline.

GENERAL OVERINVESTMENT IN RELATION TO CONSUMPTION

So far we have been considering the possibility of a decline in investment in particular sectors. Let us now see whether "general overinvestment," affecting more than just a few industries, is also a danger with which the economy must reckon. There are two possibilities here, and, paradoxically, they are the reverse of each other. Investment may eventually begin to decline because consumption does not rise rapidly enough. Or investment may eventually fall because saving does not increase fast enough (which is the same as saying that consumption rises too rapidly). This sounds like a case of "heads I win, tails you lose." Let us look into this apparent paradox, which lies behind a good deal of controversy in the literature on business-cycle theory.

Not all investment, of course, is made in response to innovations of relatively recent origin. Apart from replacement, a good deal of investment takes the form of additions to existing types of capacity in old industries. Such investment tends to be of a rather passive character. It takes place in response to an actual or anticipated increase in the demand for the products of the industry in question. Such additions to capacity can be called *induced* investment; it is induced or required by increases in output currently taking place or expected to take place soon. This immediately makes us think of the

acceleration principle. If the rise in total output begins to taper off —i.e., the absolute rate of increase begins to fall—then smaller additions to capacity are required. This is the same thing as saying that total investment begins to decline. Hence we have to ask ourselves: Is there anything in the process of expansion that eventually causes output to expand at a slower rate, and, if this does happen, will total investment necessarily decline? Business-cycle theorists are still vigorously debating this issue.[23]

As far as we can piece together the facts from the way the economy has behaved during recent business cycles, this seems to be the situation. As incomes and employment rise during the boom, the percentage of the GNP that is consumed does apparently tend to decline, and the percentage that is saved tends to rise. The absolute amount of consumption, however, rises throughout the upswing in business. The historical marginal propensity to consume GNP may decline during the latter part of the boom, although it is not clear that this always happens.

This behavior of consumption seems to be due to several factors. First of all, consumers tend to increase their rate of saving as disposable income rises during the upswing of the business cycle. Second, and perhaps more important, the kind of income that rises most rapidly during the upswing is business profits; and a large part of such profits is retained as business savings. Thus, disposable income does not rise as fast as GNP.[24] If GNP were to rise at a constant absolute rate and business savings at an increasing rate, then disposable income and consumption would increase at a decreasing rate.

Yet a third influence may eventually cause the expansion in consumption to level off: the behavior of expenditures on consumers' durable goods and the related behavior of installment credit. In the early and middle stages of the upswing, consumers are likely to be eager to purchase durable goods, expenditures on which were deferred during the preceding depression; and a substantial part of these purchases will be financed by borrowing. The rise in consumers' credit permits retail sales, particularly of durable goods, to rise more rapidly than would have been possible if all purchases had

[23] See the review of recent business-cycle theory in Chapter 13.

[24] Note should also be made of the fact that personal income taxes rise relatively more rapidly than personal incomes.

been for cash. The increase in the amount of such credit (the excess of new debt over repayments) represents consumption not financed out of current income.

As the upswing goes on, forces eventually begin to operate to hold back the rise in expenditures on consumers' durable goods. The more consumers buy, the larger become their stocks of durable goods, and the less strong becomes the incentive to add further to these stocks. Also, for a variety of reasons, extensions of new installment credit are likely eventually to flatten out, while repayments on debt already incurred continue to rise. As a result, the total amount of consumers' credit outstanding finally begins to show a decreasing rate of increase; and this exerts a downward pressure on consumers' expenditures. Borrowing inflates consumer buying only to the extent of the excess of new extensions of credit over repayments. If this difference declines, to that extent a deflationary force begins to operate on consumers' expenditures.[25]

For these and other reasons, the rise in consumption may well taper off after a prolonged upswing.[26] In any event, when full employment is reached, the rate of increase in total output and in consumption must taper off. For one reason or another, then, the rate of increase in total consumption will eventually decline *if* the expansion goes on long enough. But it is also true that business expansions may come to an end, for one reason or another, before the pressures discussed here do begin to slow up the expansion in consumption.

Suppose that, in a particular cyclical expansion, the rate of increase in consumption does begin to taper off.[27] Does it follow that,

[25] On the behavior of consumer credit in the postwar business cycle, see Board of Governors of the Federal Reserve System, *Consumer Installment Credit*, 4 parts, 1957, especially the paper by Don D. Humphrey in part II, vol. 1, and chaps. 11–12 of part I, vol. 1.

[26] If the rate of expansion in total consumption begins to decline, so, too, will the rate of expansion in total output unless investment or government spending expands at an accelerating rate.

[27] A few words of clarification should be added here regarding the relation between a possible decline in the rate of increase per month in consumption and a decline in the marginal propensity to consume GNP. The rate of expansion in consumption may well taper off even if the marginal propensity to consume is unchanged as GNP rises. This would happen if investment increased by a decreasing amount so that, with a constant marginal propensity to consume (i.e., multiplier), total output and consumption also increased at a decreasing absolute rate. If the marginal propensity to consume were to decline, then the expan-

because of the short-run working of the acceleration principle, total investment will then show an actual decline? We provided a basis for an answer to this question in Chapter 6. Even induced investment is not geared so closely to current consumption that it will immediately reflect every change in the rate of expansion in the production of consumers' goods. Producers may build capacity, looking to the long-run future, if they are optimistic; they may refuse to expand capacity, in spite of an increase in demand, if they are pessimistic. They may have excess capacity during most of the upswing and need to make no new net investment until late in the boom, perhaps after the rate of expansion in their output has begun to decline. And other factors—for example, security prices and the receptivity of the capital market to new issues—may cause supposedly induced investment to change in a way different from that called for by the acceleration principle. Finally, we are talking here only about passive, induced investment in consumers'-goods industries. More active or autonomous forms of investment—geared to the expansion of new industries and the introduction of new processes—may expand rapidly enough to offset for quite a while any deflationary force arising from the short-run working of the acceleration principle. On the whole, it seems fair to conclude that the tendency of *total* consumption eventually to rise at a decreasing rate is a *possible* but probably not the usual reason why investment in plant and equipment finally turns down at the end of a major expansion.[28]

GENERAL OVERINVESTMENT IN RELATION TO SAVING

Before 1914, business booms frequently ended in a period of marked monetary and financial stringency, and there has been some evidence of this again in some of the postwar cycles. Rising interest rates and tightening credit eventually made it difficult for business to finance further investment, through either direct loans or new security issues. At the same time, rising costs and prices and shortages of capacity indicated that aggregate demand was outrun-

---

sion in total output and consumption would taper off even if investment continued to increase by a constant amount. See the discussion of the possible interactions between the multiplier and the accelerator in Chapter 6 (pp. 144–148) and Chapter 13 (pp. 370–380).

[28] Such tapering off in consumers' expenditures may, however, lead to a decline in inventory investment.

ning supply. Facts such as these, plus certain theoretical considerations into which we need not enter here, have led some economists to believe that during cyclical expansions investment eventually becomes too large, given people's consumption and saving habits, to be financed by the available supply of saving and new bank credit.[29] There is the possibility that a maladjustment may develop here, although there is little evidence that it has been much of a factor in recent American business cycles. For the sake of completeness, however, let us see briefly how this kind of difficulty might come about.

When investment rises rapidly during a major expansion, *ex ante* saving rises but not so fast. The expansion of bank credit and the increased velocity of money indicate that some of the larger volume of investment is being financed from "inflationary" sources rather than from saving. This cannot continue forever. If nothing happens to reduce the profitability of investment, which continues to exceed the current flow of saving, then the expansion can go on only as long as the banking system creates the additional credit necessary. If such credit expansion continues unabated after full employment is reached, marked inflation will result. However, the banks are certain to call a halt before inflation has gone very far—either because they become alarmed or because they have reached the limit set by their reserves. Then there will be a "shortage of capital," characterized by high interest rates, falling security prices, and refusal of investors to buy all of the new security offerings which are coming on the market. There will be a "real" as well as a monetary shortage of capital. The high level of investment and consumption will result in a scarcity of resources and rising prices and costs. A combination of high prices for capital goods and the greater cost and difficulty in securing funds will then presumably drive investment down. As we said, this might happen if an investment boom went on long enough, and it did apparently happen on some occasions in this country, particularly before World War I. But there is little evidence that it has happened here since then. During the interwar period, and in the limited cyclical experience we have had since World War II, booms have usually ended because investment declined before a marked shortage of capital could develop.

[29] These "shortage-of-capital" theories will be briefly reviewed in Chapter 12.

REPLACEMENT CYCLES AND THE STOCK OF CAPITAL

We saw earlier that one force tending to make for revival after a long downswing is the growing need to replace worn-out capital and consumers' durable goods. Replacement and modernization expenditures are likely to rise rapidly after revival begins, but ultimately this pent-up replacement demand will begin to subside. Thus, an important stimulus to total (gross) investment tends to weaken, and we add another to our growing catalogue of deflationary forces that are likely to be generated by the way the economy behaves during a major cyclical expansion.[30]

Some economists look upon the fact that the country's total stock of capital grows rapidly during cyclical expansions as being in itself a deflationary force. As we shall see in Chapter 13, much of the recent theoretical literature on growth and business cycles is concerned with the fact that current investment increases both the stock of capital and (through the multiplier) the level of income and output. If the capital stock (and therefore productive capacity) increases faster than the level of income, investment will eventually have to fall.

We shall review and criticize this type of theory in Chapter 13. It is certainly true that investment may behave in such a way as to cause capacity to expand faster than aggregate demand, so that investment eventually declines. To the extent that this happens, however, it is only because new investment eventually becomes unprofitable for one of the various reasons suggested in the preceding and following paragraphs. The surface indication provided by the overall

[30] Replacement can conceivably generate cycles in investment that will continue to reproduce themselves. If investment in particular capital goods was heavily concentrated during some short period in the past, then there will be a "bunching" of replacement needs at some future time, depending on the life of the capital goods in question. Then at a still later time this concentration of replacement needs during a short period will be repeated again, and so on. (See, for example, J. Einarsen, "Reinvestment Cycles," *Review of Economic Statistics,* vol. 20, February, 1938, pp. 1–10.) Given the different degrees of durability of various types of capital equipment, the fact that technology creates widely varying rates of obsolescence in different industries, and the effect of general business conditions on the timing of replacement, it is unlikely that replacement affects the business cycle in any more definite or precise fashion than that described in the text. Cf. J. S. Bain, "The Relation of the Economic Life of Equipment to Reinvestment Cycles," *Review of Economic Statistics,* vol. 21, May, 1939, pp. 79–88; also J. R. Hicks, *A Contribution to the Theory of the Trade Cycle,* 1950, p. 42.

ratio of a country's stock of capital to its total output tells us little or nothing about the particular maladjustments affecting the profitability of further investment that develop during the later stages of a cyclical expansion.[31]

THE BEHAVIOR OF SAVING

Let us now summarize the role that saving plays in bringing the upswing to an end. We have seen that saving rises rapidly during the expansion. But so, too, does the volume of investment. Until the peak is reached, investment exceeds saving when both are defined in *ex ante* terms. In what sense, then, can it be said that saving eventually helps to bring on the downswing? The answer is that saving *may* play an active part in ending the boom, and in any event it is certain to play a critically important passive role. We have already described how, on occasion, a rapidly expanding volume of saving *may* tend to hold back the rise in consumption and thus lead to a decline in investment via the acceleration principle. This is the way in which savings may sometimes initiate a decline in investment. But what we have called the passive role of saving is even more important. The volume of saving is closely linked to the size of the GNP; a high level of saving goes with a high level of GNP; and the volume of saving will ordinarily decline only if the GNP falls. Now we come to the most important point. If, at the end of the boom, investment should decline for any reason, the fact that people do not automatically reduce saving and expand consumption correspondingly means that aggregate demand—and therefore production and employment—must decline. Or we can put it this way. Given the way saving behaves, we can never let investment begin to decline without letting a cumulative contraction get under way. The higher the volume of saving is carried by the rise in GNP, the higher is the volume of investment that must be maintained. Yet, as the preceding pages show, fluctuations in the volume of investment are inevitable in a private-enterprise economy. Hence, as we saw in Chapter 8, a "high saving" economy is inherently unstable.

[31] What we have said here regarding the cyclical consequences of an increase in the ratio of the capital stock to output in the short run is not intended to deny that paying attention to the overall relation between capital and output may be useful in studying the determinants of *long-term* growth. See the discussion of the Harrod-Domar growth model in Chapter 6 and of recent cycle-growth models in Chapter 13.

THE UPPER TURNING POINT

We can see that all booms do not end for the same reason. As a result, the sequence of events at the turning point will differ somewhat from cycle to cycle. Ordinarily, the peak of the cycle cannot be clearly distinguished at the time it occurs. There is usually a "turning-point zone" or critical period of six months or more, beginning a few months before and ending a few months after the date that we finally take for the turning point. During this period, output and employment reach a peak and perhaps decline slightly. Stock prices and then commodity prices will have begun to fall early in the critical period. There is a noticeable increase in the liabilities involved in bankruptcies and growing complaints of excessive inventories. Prices of the more speculative commodities may drop sharply, and there are increasing reports that some firms have to curtail output. Firms generally notice that their backlogs of customers' orders are no longer rising and perhaps are beginning to decline. Caution takes the place of buoyant optimism in a number of lines.[32]

If it is only a minor recession that is beginning, the early declines will be most marked in the prices and production of those industries in which firms are most anxious to reduce inventories. The textile and other industries using raw materials that fluctuate widely in price will start to curtail output; so will other industries in which excess capacity or inventories have been accumulating. By the end of the critical period, when it becomes clear that a recession is developing, the drop in security and commodity prices may become quite rapid. The initial declines in output reduce incomes and spending, and thus the contraction spreads. Business optimism turns into caution and even into pessimism regarding the short-term future. Investors become loath to buy new securities, and the banks become less free in their lending. Thus, long-term investment begins to decline, though the decline will not carry very far if only a minor cycle is involved.

Suppose that we are dealing with a major turning point, which ushers in a long and severe decline. The initial symptoms may be the same as for a minor turn. Not till general confidence is first shaken by a drop in the stock market and by evidence of excessive inventories may many firms begin to doubt the wisdom of continuing with

[32] See Table 37 on p. 516 for a list of series that typically lead at the peak.

their long-term investment plans. The tendency for underlying investment opportunities to begin to decline may not have been evident as long as speculative optimism was feeding on itself.

But if investment in some lines had grown particularly excessive, for one or more of the reasons discussed in the preceding section, some types of long-term investment may turn down quite early during the critical period. An outstanding example here is the building industry. As we have seen, it is possible that building activity will not decline very much or for long during a major contraction—in which case, the depression will not be unusually long or severe. But if the underlying building boom does end, its peak may well come before that in general business.[33] The maladjustments discussed earlier lead to a decline in building activity and in real-estate prices and to a rise in mortgage foreclosures. This important decline in investment may be offset for a while by expanding investment elsewhere; but as the decline in building continues, it becomes increasingly difficult to offset the deflationary influence from this source.

Other types of investment may decline early if capacity has been outrunning demand. In this case, the demand for some types of capital goods will begin to decline early in the critical period; the reduction in investment will affect incomes and consumption; security prices will be affected; and so on. In any event, the early decline in stock prices will lead to a decline in new security issues and to downward revisions in investment plans.[34] High interest rates and increasing credit tightness also serve to reduce some forms of investment, especially once expectations begin to weaken for other reasons. The result of these factors is that new orders for capital goods are likely to begin declining early during the critical period, though expenditures on the purchase and installation of capital goods already ordered may continue high for a few months more.[35] Sales of some types of consumers' durable goods—particularly automobiles —may also begin to decline early, for reasons we discussed in the preceding section.

There is no evidence that retail sales or consumption regularly

[33] This seems usually to have been the case. Peaks in long building cycles have usually led by a considerable margin the peaks of the major cycles with which they were associated. Cf. C. D. Long, Jr., *Building Cycles and the Theory of Investment*, 1940, chap. 9.

[34] See p. 296 for mention of the lead in new security issues at the peak.

[35] Cf. p. 277 for mention of the lead in new orders for investment goods.

leads at the peak. As a matter of fact, they are more likely to lag. However, the expansion in retail sales may flatten out noticeably during the critical period. This, together with the initial decline in commodity prices and emerging signs of difficulty in various parts of the economy, may lead retailers to curtail purchases and attempt to reduce inventories even before consumers' buying actually declines.[36]

Cyclical peaks in the past have usually been accompanied by some degree of credit stringency, as reflected in high interest rates, falling bond and stock prices, and a marked decline in the willingness of lenders to continue making loans at the former rate. Though some bank loans are probably called during this critical period, actually the total volume of bank loans outstanding does not decline until the downswing in business is definitely under way. But the volume of *new* lending may begin to decline before or at the peak, and some of the lending that takes place at the peak may be to borrowers already in difficulties in order to help them meet pressing claims and to reduce inventories.[37]

Cyclical peaks are not necessarily accompanied by financial panics or monetary crises. Such crises did accompany the turning points preceding the more severe depressions before 1914. But since the inauguration of the Federal Reserve System, the commercial banks have been able to meet the claims of depositors and the most pressing needs of their customers during whatever credit stringency developed at cyclical peaks. "Panic" or "crisis" situations have occurred twice in the United States since World War I. In 1929, the collapse of stock prices was of panic proportions, but this did not involve a banking crisis. During 1931–1933, repeated waves of bank failures led finally to the closing of all banks in March, 1933. This was a banking crisis, but it occurred at the bottom rather than the peak of a cycle.

THE DOWNSWING

It usually takes a few months after the actual peak in business has been reached before it becomes clear to informed observers that a

[36] Cf. Ruth P. Mack and Victor Zarnowitz, "Cause and Consequence of Changes in Retailers' Buying," *American Economic Review,* vol. 48, March, 1958, pp. 18–49. Their analysis indicates that retailers' orders are likely to lead the final downturn in retail sales.

[37] Banks are likely to start selling bonds during the period of rising interest rates preceding the peak, and this contributes to the tightness of credit. Cf. p. 294, above.

cumulative downswing has got under way. What happens from now on is largely the converse of developments during the upswing.

The virus of deflation spreads rapidly through the economic system. The initial decline in output reduces employment and incomes. Business expectations deteriorate rapidly under the double prod of declining sales and falling prices. For a while inventories continue to rise, in part because firms are unable to adjust their output and outstanding purchasing orders fast enough. This unplanned accumulation of inventories makes businessmen all the more anxious to reduce output. Manufacturers and wholesalers begin to be flooded with cancellations of orders. The decline in output and prices gathers momentum.

In the capital-goods industries, the decline in production and employment is extremely sharp, almost from the beginning of the downswing. Prices of capital goods decline very slowly at first; later the decline in such prices may accelerate. In the case of nondurable and consumers' goods, the fall in prices is more rapid; and, on the whole, the decline in output proceeds more slowly. The physical volume of retail sales and total consumption moves downward also, but more slowly than total GNP. The most pronounced declines in the components of GNP are in the various types of capital formation and in the flow of consumers' durable goods.

The fall in profits is extremely rapid, being most extreme in the durable-goods industries. Total unit costs fall less rapidly than prices. Fixed cost per unit rises because overhead costs are spread over a smaller volume of output, and wage rates are slow to decline. As profits fall, and as some firms with heavy commitments encounter difficulties in meeting their bills, the number of and the volume of liabilities involved in business failures rise rapidly. Business confidence deteriorates further—leading to a further decline in investment, incomes, and consumption, to still lower prices and profits, and thus to a still worse impairment of business confidence.

One reflection of the developing contraction—and a cause of still further deflation—is the scramble for liquidity by businessmen, investors, and banks. If the depression is at all serious, nearly everyone will prefer to hold cash to owning commodities or securities at present prices—because of the fear that prices will fall further and because pressing debts have to be met. Thus, firms tend to hoard some part of their sales proceeds—for example, by not replacing in-

ventories and by not reinvesting depreciation allowances in new equipment. Banks are reluctant to make new loans and ask for the repayment of old loans as they mature. Large investors hold a big part of the current flow of savings in the form of cash or invest it in short-term government securities—both because they lack confidence in present prices of corporation securities and also because firms no longer wish to borrow. Thus, a decline in M and V goes along with the contraction of output as both cause and effect.

It is now time to bring in again our distinction between major and minor cycles. If underlying investment opportunities remain favorable, the contraction will not proceed very far. Many investment projects will continue in the face of the general decline in business, and others will be started again after the initial shock to confidence wears off and after business firms begin to be tempted by lower prices for capital goods and lower interest rates. If the recession is only a minor one, investment and therefore consumption will be fairly well maintained; the general decline in confidence will be moderate; the desire for liquidity will be satisfied relatively easily; and investors will not withhold their funds from the capital markets for very long. In this case, the maladjustments that have to be corrected are not serious, and a sharp curtailment of output for a year or less (or a slower decline for a somewhat longer period) may be sufficient to induce recovery. We saw earlier in this chapter how this would come about.

This is a good point at which to mention again the automatic stabilizers. In a minor recession, much of the moderate decline in GNP will be offset by a sharp drop in profits, while dividend payments remain comparatively stable. In effect, business firms sharply reduce their saving and their tax payments to the government. At the same time, transfer payments increase, particularly to the unemployed; and, to the extent that personal incomes decline, personal income taxes are lower. Thus, the decline in disposable income and in consumers' expenditures turns out to be much less than in GNP, and this permits manufacturers and retailers to bring about a prompt adjustment in their inventories.

The effect of the automatic stabilizers weakens the longer and more severe a depression becomes. If profits fall long enough, dividend payments will finally be reduced; and, the larger the number of firms with zero or negative profits, the smaller will be the continued

decline in corporate income taxes. Also, if the depression lasts long enough, the unemployed begin to exhaust the unemployment benefits to which they are entitled under existing legislation. Thus, while the automatic stabilizers continue to operate in a major downswing and help to make a major contraction less severe than it would otherwise be, the stabilizers do their most effective job in a minor recession in which long-term investment and government expenditures are relatively well maintained. Or, to put it another way, the automatic stabilizers work best if the underlying situation calls for chiefly a decline in inventory investment rather than a protracted fall in plant and equipment expenditures and in residential and commercial building.

If only a minor contraction is involved, this means, among other things, that the underlying demand for housing is still strong. Hence, as interest rates decline and funds become more generally available, prospective builders and purchasers of homes find that mortgage credit is more readily available and on easier terms. As a result, residential building may begin to expand well before the general business contraction ends.

Suppose, however, that the economy is in a major downswing. In the case, confidence in the long-run future deteriorates rapidly as business activity declines. Investment in long-term projects declines to low levels and drags down income and employment as it falls. The character and magnitude of the decline in investment will depend on the nature of the maladjustments that developed during the preceding boom. In this connection, we should distinguish between two general types of maladjustments that may push long-term investment down to a very low level during serious depressions. One centers on the willingness of business to invest—i.e., on the marginal efficiency of capital. The other involves the willingness of lenders to supply funds for investment.

How seriously the incentive to invest is impaired depends on the degree and character of overinvestment during the preceding expansion. We have already discussed the various possibilities in considering why the investment boom comes to an end. The decline in investment will be particularly severe if a long downswing in commercial and residential building is under way, if serious overcapacity in some important industries has to be worked off, and if one

or more large industries that had been expanding rapidly now find that they have reached a position of relative maturity in which future expansion must be at a slower pace than in the past. Here, as in so many other aspects of our subject, we find a vicious circle at work. The more seriously some underlying investment opportunities have been impaired, the greater the decline in investment and incomes, and hence the more seriously impaired investment opportunities seem to be in yet other fields in which the long-run prospects may still be favorable.

Let us now turn to the supply side of the market for investment funds. As the downswing develops, lenders become less and less willing to put their funds into new investment projects. We have already spoken of the general scramble for liquidity and of the loss of confidence in the future of security prices. After a long and speculative investment boom, this range of developments can become a highly important factor accentuating the severity of the depression. Banks and other financial institutions may find that they have acquired securities and made loans that are of dubious value in the harsh light of "the morning after." They may try to dump such securities, thus accelerating the fall in security prices. At the same time, their impaired financial position makes them unwilling to invest in new projects. A "once burned, twice shy" attitude toward new ventures spreads through the investing community. The supply curve of loanable funds shifts to the left and becomes highly inelastic. Regardless of the willingness of borrowers, lenders do not want to lend.

This sort of situation can degenerate into a "secondary deflation" of considerable severity. The fall in security prices and the inability of firms to meet their obligations jeopardize the solvency of some financial institutions. The latter aggravate matters by dumping securities and calling loans. The public loses confidence in the soundness of the banking system and seeks to convert deposits into currency. Some banks fail, causing a still further loss of confidence, and so on. This sort of vicious spiral, leading to eventual collapse of the banking system, was an outstanding feature of the Great Depression of the 1930's. Our banking system today is much sounder than it was then, and deposit insurance should help to maintain the public's confidence in the banks during future downswings. But

it is probably too much to hope that we can completely escape some measure of financial liquidation in future depressions.[38]

Eventually this sort of liquidation comes to an end, with or without help from the government. The weaker firms and banks fail and the stronger ones prove that they can withstand the storm. The scramble for liquidity abates somewhat. Individual and institutional investors find themselves with large amounts of surplus cash. Already short-term interest rates will have fallen considerably, and yields on high-grade bonds will have begun to decline also. As the bond market improves, corporations begin to put out refunding issues to take advantage of lower interest rates.

In the meantime, other corrective forces are at work to abate the pace of the decline and to lead eventually to a revival in business. Even major contractions finally come to an end—though at the bottom of the depression millions may be unemployed and though output and incomes may have sunk to painfully low levels. This is the point in the cycle at which we began our discussion of how the economy generates business cycles. Gradually the inheritance of maladjustments from the preceding boom is eliminated; new stimuli gather strength; and a new upswing finally begins.

RECAPITULATION

The preceding analysis has been highly unsatisfactory in one respect. It has not left us with a simple, clear-cut theory of the business cycle in which we can put the whole blame on one or two obvious causes. Indeed, we have been forced to conclude that there are a wide range of forces that can create cyclical fluctuations and that these influences can be combined differently in different business cycles. It is no wonder that the forecasters have been unable to find a simple, magic formula that will infallibly predict the turning points in business!

Some generalization, however, is possible, and the following series of summary propositions may help the reader to fix more clearly in mind the main conclusions to be drawn from the preceding sections.

[38] Maladjustments in the field of international trade and finance can also add greatly to the severity of a downswing. This was particularly true of the Great Depression of the thirties. Some of the international aspects of business cycles will be discussed in Chapter 22.

1. We need to distinguish between the cumulative processes of expansion and contraction, on the one hand, and the turning points that mark the end of these processes, on the other.

2. Once a turning point occurs, the nature of the cumulative process that follows is basically similar in all cycles. The important distinction here is between major and minor swings. In major cycles, the character of the cumulative process rests on the nature of the expansion and contraction of long-term investment; in minor cycles, the cumulative process rests on the short-period plans and expectations of business firms.

3. Major and minor cycles may occur in various combinations. Each expansion will reflect the particular combination of underlying stimuli that set the cumulative process in motion, and each contraction will differ depending on the particular combination of maladjustments that brought the preceding expansion to an end. Hence, no two cycles will look exactly alike.

4. Each minor or major contraction generates a set of self-correcting forces which gradually brings the downswing to an end. In minor cycles, the most important influences leading to recovery are the liquidation of excess inventories, the correction of horizontal maladjustments (chiefly by working off inventories in those particular industries that had previously been producing too much), and the achievement of a more liquid position by firms and financial institutions. These self-correcting forces are powerfully aided by the automatic stabilizers, which help to support disposable income when the GNP declines. The most important factors likely to lead to a revival of long-term investment, and thus the initiation of a new major cycle, are the accumulated need to replace worn-out and obsolescent capital goods and renewed expansion in new industries that have been growing rapidly. Lower costs and interest rates and the greater availability of mortgage credit sometimes permit building activity to play an important initiating role.[39]

5. Minor recessions are caused by the accumulation of minor maladjustments capable of quick correction, while underlying long-term investment opportunities continue to be favorable. The most important of such minor maladjustments are the decline in new orders and inventory investment that may result from a tapering

[39] These factors may permit residential building to play an initiating role in minor-cycle revivals also.

off in retail sales, a reversal of price expectations (leading to attempts to liquidate inventories), and overproduction in particular industries (horizontal maladjustments). In some cases, tightening of credit may play an important role. The most important causes of major downswings, which involve a serious impairment of long-term investment opportunities, are likely to be the following: reduced investment opportunities in new industries resulting from either the development of excess capacity or the maturing of these industries, the ending of a boom in residential and commercial building, and the satisfaction of accumulated needs for replacement and modernization. On occasion, but not necessarily, the leveling off in total consumers' demand may cause long-term investment to fall. On occasion, also, investment booms may be cut short by a scarcity of capital.

6. The cause of depressions can be restated in terms of the saving-investment analysis elaborated in Chapters 4–6. After an expansion has gone on for some time, private investment will inevitably decline for one or more of the reasons discussed in the present chapter. When this happens, the economy has no way of automatically reducing saving and increasing consumption by a corresponding amount. Hence, aggregate demand must decline, causing investment to decrease still further.

7. Financial and monetary developments frequently make major downswings more severe than they need be. The decline in value of capital assets and the impaired solvency of debtors may set off a vicious spiral of credit contraction, bank failures, and a general scramble for liquidity that for a while make investment funds completely unavailable even to borrowers for whom investment opportunities may still be favorable. The seriousness of such secondary deflations depends on the strength of financial institutions, including the support provided by the government, and on the degree of speculation and unwise lending that went on during the preceding boom.

8. We have considered in this chapter how the ordinary working of the economy itself generates business cycles. In addition to the "endogenous" influences that we have emphasized, a wide variety of "external" forces can intervene to cut short business expansions or contractions—for example, unusually good or bad crops, foreign developments, and various types of government interven-

tion. The last is likely to play an important role in future business cycles, particularly in bringing to an end severe business contractions. Given the present virtually universal assumption that the government must accept responsibility for maintaining a reasonably high level of employment, business declines do not have to become very severe before government intervention becomes a political necessity. Cumulative expansions, on the other hand, have usually come to an end before government intervention was necessary to prevent marked inflation. There is at least the possibility, however, that in the future upswings will generate inflationary spirals more easily than in the past. If this turns out to be the case, we may expect to see an increasing amount of government intervention during cyclical expansions also.

9. Since the 1930's, important structural changes have occurred in the American economy (and in the economies of other countries) that have modified various aspects of the cyclical, self-generating mechanism described in this chapter. We have referred to the nature of these changes in earlier chapters. Experience since World War II indicates clearly that, despite these changes, the minor-cycle mechanism has continued to operate in the United States in a manner that is not radically different from that which prevailed in prewar cycles.

What is uncertain is whether conditions have changed so much that major depressions are a virtual impossibility in the future. The answer to this question is probably in several parts. Certainly, the financial reforms of the last generation have significantly reduced the chances of a severe "secondary deflation," which in the past has made major contractions much worse than they otherwise would have been. Secondly, we are virtually certain never again to have a depression as severe as that of the 1930's. Structural reforms, our greater knowledge of how to cope with depressions, and the inevitability of large-scale government intervention if mass unemployment is threatened all make this a fairly safe prediction. What are uncertain are the future course of long-term investment opportunities and the ability of the economy to adjust to such important changes in investment prospects as may occur. In this respect, the situation in the first 15 years after World War II has been highly favorable. After the postwar backlog demands were satisfied, rapid technological change, continued population growth and migration,

rising incomes, changing tastes, the need to economize on labor, and so on, along with a high level of government spending, helped to maintain a large stock of investment opportunities that left no room for a major depression. This may or may not continue to be the case. To cite only one obvious example, a large-scale disarmament effort might well lead to a serious depression before the necessary adjustments could be made. Or a significant slowing down in the rate of population growth might lead to enough of a decline in building and other long-term investment to bring on a much more serious depression than any yet encountered in the postwar period.

Without question, the American economy is much more stable than it was a generation or more ago. But minor cycles are still with us, and the danger of moderately severe major depressions has not been completely eliminated. And all this is aside from the possibility of chronic inflation, about which we shall have more to say in a later chapter. It is clear that we have not yet completely exorcised the curse of economic instability, although it is no longer as serious a problem as it was in our fathers' and grandfathers' times.

CHAPTER 12

THE VARIETY OF BUSINESS-CYCLE THEORIES

TO WHAT extent does our explanation of what happens during business cycles agree with the opinions of other writers on this subject? Many different theories have been offered to explain why business cycles occur, and the informed student in this field should have a working acquaintance with the more important of these explanations. At the same time, we do not want to spend too much time on what others have said. A detailed study of the theoretical literature had best be left to more advanced courses or to courses specifically concerned with the history of economic thought. Hence, we shall confine our systematic review to this and the next chapter.[1]

HOW BUSINESS-CYCLE THEORY HAS DEVELOPED

Systematic study of business cycles is, surprisingly, a relatively recent development. Most of the important contributions in this field have been made in the last 60 years or so. Although booms and depressions occurred all through the nineteenth century, classical economic theory had relatively little to say about the causes of economic fluctuations. Orthodox theory assumed that Say's "Law of Markets" was a valid representation of underlying forces operating in the real world—that supply created its own demand and that therefore unemployment would be speedily eliminated by appropriate adjustments in interest and wage rates and other prices.[2] The

[1] For more detailed reviews of the literature, see: G. Haberler, *Prosperity and Depression,* 4th ed., 1958; Thomas Wilson, *Fluctuations in Income and Employment,* 3rd ed., 1948; A. H. Hansen, *Business Cycles and National Income,* 1951; D. Hamberg, *Business Cycles,* 1951; and Maurice W. Lee, *Economic Fluctuations,* rev. ed., 1959.

[2] J. B. Say was a French economist who wrote in the early decades of the nineteenth century.

classical economists then went on to concentrate their attention on problems of resource allocation and how a given national income would be divided among the factors of production. Until the latter part of the nineteenth century, only a few dissenters, such as Lauderdale, Malthus, Sismondi, and some of the early socialists, challenged the orthodox assumption that a private-enterprise economy tended naturally toward a condition of full employment in the manner described by Say.

At the same time, the fact that crises and depressions did occur could not be completely ignored, and a body of literature about crises and panics began to accumulate—much of it by men who were not professional economists.[3] Not until the latter half of the century, however, did observers generally come to recognize that it was necessary to study the whole business cycle and not merely the crises that occurred at the end of periods of prosperity.[4]

One of the first writers to recognize the need for systematic study and a full explanation of the entire business cycle, including the cumulative processes of expansion and contraction as well as the turning points, was Clement Juglar, the first edition of whose *Des Crises Commerciales* appeared in 1860.[5]

Between 1890 and World War I, important contributions to the theory of business cycles were made by a number of writers—for

[3] For an interesting account of the early literature on crises and cycles in the United States, see H. E. Miller, *Banking Theories in the United States Before 1860,* 1927, chap. 16. See also Hansen, *op. cit.,* chap. 13. The early European literature on crises and cycles is summarized in J. A. Schumpeter, *History of Economic Analysis,* 1954, pp. 738 ff. See also R. G. Link, *English Theories of Economic Fluctuations, 1815–1848,* 1959.

[4] This statement needs some qualification. Miller points out that in the United States after 1825, "Attention began to be given also to the business cycle as a whole, as well as to its climax in the 'revulsion' " (*op. cit.,* p. 193). Several writers, both in England and in the United States, seem to have had fairly well-developed notions of a complete self-generating cycle before 1860. See also Hansen, *op. cit.,* pp. 216–218, and the other references in the preceding footnote.

[5] Juglar was not the first writer to recognize that modern economies were subject to self-generating cycles and not merely to occasional crises, although general recognition by economists of the wavelike character of business fluctuations may perhaps be dated from Juglar's book. In addition, Juglar made a highly important contribution in recognizing the need for systematic study of the various aspects of cyclical fluctuations. Juglar's work can be most readily examined in the second edition of *Des Crises Commerciales et de Leur Retour Périodique en France, en Angleterre et aux États-Unis,* 1889. For brief evaluations, see J. A. Schumpeter, *Business Cycles,* 1939, pp. 139, 162–163, and *History of Economic Analysis,* pp. 1123–1124; also T. W. Hutchison, *A Review of Economic Doctrines, 1870–1929,* 1953, pp. 370–372.

example, M. Tugan-Baranowsky in Russia, Arthur Spiethoff and J. A. Schumpeter in Germany, Knut Wicksell in Sweden, D. H. Robertson and R. G. Hawtrey in England, Albert Aftalion and Jean Lescure in France, and Thorstein Veblen and W. C. Mitchell in the United States. In 1913 Wesley C. Mitchell published his famous *Business Cycles,* which described in detail what seemed to happen during business cycles and offered an eclectic explanation of the various phases of the cycle.[6]

Since then there has been a proliferation of business-cycle theories as well as of empirical studies of cyclical fluctuations. Although between 1900 and the 1930's an increasing number of professional economists concerned themselves with the study of business cycles, business-cycle theory continued to be something outside the framework of general economic theory. The latter continued to operate on the assumption that full employment was the natural, equilibrium position for the economy, whereas business-cycle theorists concerned themselves with why the economy naturally generated fluctuations in employment and output.

J. M. Keynes' *General Theory of Employment, Interest, and Money,* published in 1936, helped to bridge the gap between these two bodies of economic doctrine. By his analysis of the saving-investment process, Keynes convinced most economists that full employment was not necessarily an equilibrium position, and he provided a powerful set of tools for studying why employment might be at a high or low level.[7]

Today these tools are part of the standard equipment of every economist. At the same time, paradoxically, Keynes himself had relatively little to say about the causes of cyclical fluctuations. As one writer has put it: "Keynesian economics, in spite of all that it has done for our understanding of business fluctuations, has beyond all doubt left at least one major thing quite unexplained; and

[6] Special mention should be made of the neglected writer, N. Johannsen, who presented a very "Keynesian" analysis of the saving-investment process around the turn of the century. Johannsen even had a well-developed concept of the multiplier, which he called the "multiplying principle." See his *A Neglected Point in Connection with Crises,* 1908.

[7] Keynes' emphasis on the saving-investment process had been anticipated in part by a number of writers, particularly by Knut Wicksell and his followers in Sweden. Wicksell's *Interest and Prices* (English translation, 1936) was first published in German in 1898. See B. Ohlin, "Some Notes on the Stockholm Theory of Savings and Investment," reprinted in *Readings in Business Cycle Theory,* selected by a committee of the American Economic Association, 1944, pp. 87–130.

that thing is nothing less than the business cycle itself."[8] We indicated in Chapter 4 why this is so. Keynes showed how investment, saving, and consumption interact to determine a given level of employment, but he did not explain why and how these variables continue to change with the passage of time.

Today we can choose from a wide range of business-cycle theories. These explanations differ among themselves particularly as to what causes the upper turning point of the cycle. This problem still remains the chief interest of business-cycle theorists. Less detailed attention is ordinarily paid to the lower turning point. There is considerable agreement as to what happens between the turning points, during the periods of cumulative expansion and contraction, and many writers pay relatively little attention to these phases.

It is not easy to work out a satisfactory classification of current theories of the business cycle. But though it is at least possible to classify *theories,* it is dangerous to classify *theorists.* Few writers today would hold to a single, simple explanation to the exclusion of all other possible hypotheses. When we come to classify theories, rather than writers, we must keep this problem in mind also. Any theory that hopes to explain the entire cyclical process must take note of a large number of factors. One or two may be stressed above the others, particularly those which it is thought explain why the downturn occurs; but this does not mean that other factors are explicitly rejected as having no causal significance during any phase of the cycle.

In Chapter 8 we saw that there are a number of characteristics of a private-enterprise, industrialized economy that make for instability, and we took this into account in developing our own explanation in Chapter 11. In large part, the difference among current explanations of the business cycle rests on which of these factors making for instability is chosen for primary emphasis. Although this is not completely satisfactory, we shall classify current theories of the cycle chiefly in terms of these factors as follows:

1. Theories that emphasize changes in price-cost relations and profit margins and changes in business expectations. We shall refer to these as "business-economy" theories.

[8] J. R. Hicks, *A Contribution to the Theory of the Trade Cycle,* 1950, p. 1.

2. Monetary theories.
3. Theories emphasizing the role of saving and investment, subdivided into:
   a. Shortage-of-capital theories
   b. Investment opportunity theories
   c. Theories emphasizing the dependence of investment on the behavior of final output.
   d. "Oversaving" and "underconsumption" theories.
4. Miscellaneous explanations—particularly agricultural and meteorological theories.

It is evident that our classification runs in terms of what the various explanations say about the upper turning point. We should keep in mind the possibility that two explanations may agree as to the causes of the downswing but may differ as to why a new upswing eventually begins.

Most modern business-cycle theory falls under the third of the main headings listed above, particularly under heading 3c. The emphasis is on the interrelations of investment, saving, income, and output.

As we have seen in earlier chapters, problems of long-term growth have come in for increased attention since World War II—problems of growth and development in both advanced and underdeveloped economies. To an increasing degree, economists are seeking to formulate dynamic theories or models that will tell us something about those forces that help to determine *both* the character of economic fluctuations and the rate of growth that an economy experiences. Almost without exception, these theories concentrate on the relations between investment and saving, on the one hand, and income and output on the other; that is, they fall under heading 3c in our classification. We shall look at these theories in some detail in the next chapter. But first we must look at the earlier types of explanations in our list. Let us begin with the first type listed.

THEORIES EMPHASIZING THE INSTABILITY INHERENT IN A BUSINESS ECONOMY

This type of explanation emphasizes the dependence of production on profit-making and on profit expectations, the dependence of profits on balanced cost-price relations, the role of uncertainty in

planning for the future and the inevitability of mistakes, the interdependence of all parts of the economic system so that maladjustments spread rapidly through the economy, and so on. These theories are by nature eclectic; they do not emphasize one factor to the exclusion of all others. They are also relatively realistic in that their generalizations are in good part drawn from observed behavior of the economy. At the same time, these theories tend to be somewhat superficial because they frequently do not probe deeply enough into the reasons why the variables emphasized change the way they do.

The outstanding exponent of this type of explanation is W. C. Mitchell, whose 1913 volume is a landmark in the literature on business cycles. In a sense, it is unfair to put Mitchell into any single category. His explanation, which he never looked on as anything but tentative, is a highly eclectic one; and he noted many of the factors that other writers have built into separate theories. But in nearly all his writings on the subject we find this same emphasis on the characteristics of a profit-making, money-using economy that tend to make for instability.[9]

According to Mitchell, depressions eventually produce conditions favorable to a new upswing. Costs are reduced, profit margins eventually improve, inventories are low and eventually require restocking, weak firms are weeded out, banks become more willing to lend as excess reserves accumulate, and so on. A business economy thus tends to generate conditions favorable for revival, and external events may intervene also. Once recovery begins, it grows cumulatively. Increased production expands incomes and retail sales; worn-out equipment and depleted inventories are replaced; business optimism generally improves. Prices begin to rise, and the expectation of further price increases stimulates business demand still further. Rising profit margins and increased volume of business cause profits to increase rapidly; improved profits cause business investment to expand; incomes, spending, production, and prices react on one another and rise further in cumulative fashion.

[9] The most complete statement of Mitchell's position is in Part 3 of his 1913 volume, reprinted as *Business Cycles and Their Causes*, 1941. See also his essay in *Readings in Business Cycle Theory*, pp. 43–60, and *Business Cycles: The Problem and Its Setting*, 1927, chap. 2. For detailed evaluation of Mitchell's life and work, including his contributions to the study of business cycles, see A. F. Burns, ed., *Wesley Clair Mitchell: The Economic Scientist*, 1952.

Mitchell finds the cause of the downturn in "a slow accumulation of stresses within the balanced system of business—stresses which ultimately undermine the conditions upon which prosperity rests."[10] Considerable emphasis is placed on the eventual development of unfavorable cost-price relations, which impair profit margins. Overhead costs eventually rise; and, more important, operating costs rise rapidly because of pressure of output on capacity, rising wage rates, lessened labor efficiency, and rising raw material costs. The boom generates other stresses—for example, limited bank reserves, growing tightness in the market for long- and short-term funds, and rising costs of new construction and equipment. Although he did not mention this in his 1913 volume, Mitchell later took account of the acceleration principle, which would lead to a decline in new orders for capital goods at the same time that the other stresses mentioned were developing.

Eventually, some industries find that their selling prices cannot be raised sufficiently to permit continuation of satisfactory profits.[11] Deteriorating profit prospects, reinforced by the growing monetary stringency, lead to credit contraction in some sectors. New orders decline; output is reduced; creditors refuse new loans and seek repayment of old debts. Liquidation thus begins; and, because the economy is tied together in a network of commercial and financial relationships, it spreads rapidly to other firms. There is a scramble for liquidity; efforts are made to reduce inventories and new orders are curtailed sharply; price declines feed on themselves and affect more and more commodities; production, employment, and incomes decline; and the cumulative process of contraction gets fully under way. The decline goes on until the economy builds up corrective forces strong enough to start a new recovery.

Today all of this sounds highly familiar; and for years, until Keynes shifted economists' attention to the aggregative variables of income, investment, saving, and consumption, the account summarized above provided the basis for most textbook discussions of what happened during business cycles. Mitchell's emphasis differed markedly from that made familiar by the Keynesian analysis. He did not believe that the key to business fluctuations could be found

[10] In *Readings in Business Cycle Theory*, p. 50.

[11] Mitchell considers various reasons why selling prices do not keep ahead of costs indefinitely, but his analysis in this respect is not altogether satisfactory.

in the relations among a few aggregative variables. Business cycles resulted from the action on business profits of the whole complex of trading and financial relationships that tie the economic system together.[12]

In at least two respects, developments since World War II have tended to confirm the importance of Mitchell's approach to the study of business cycles. As we saw in Chapter 10, there has been more of a tendency for rising costs to encroach on profit margins in postwar cyclical expansions than was the case in prewar cycles.

Secondly, the story told by "diffusion indices" tends to confirm the emphasis Mitchell placed on the gradual accumulation of stresses in particular sectors of the economy. The downturn at the end of a boom does not come all at once in all industries. It begins in particular sectors of the economy and then gradually spreads. A diffusion index, as will be described in more detail in Chapter 17, records the proportion of a given number of economic series that are expanding in any given month. If we take a large number of different production series (the same would be true if we took a number of series for employment or profits in different firms or industries), we find that an increasing number of these series will begin to turn down as the boom goes on; the number of series that continue to expand begins to decline well before the peak in general business is reached.

Other writers also have emphasized in one way or another how profit-seeking by businessmen in the face of an uncertain future may lead to cyclical fluctuations and particularly to a downturn after a period of prosperity. Two chief possibilities have been stressed by a number of writers—the possibility of horizontal maladjustments (partial overproduction) and that of alternating waves of optimism and pessimism that sweep through the whole business community.

Horizontal maladjustments arise when particular industries find

[12] Mitchell devoted the latter years of his life to directing empirical studies aimed at throwing light on these interrelationships, and the results of this work can be seen in the publications of the National Bureau of Economic Research. See also his incomplete study, published posthumously, entitled *What Happens During Business Cycles,* 1951. For a vigorous defense of this approach to business-cycle problems and a criticism of recent attempts to study business cycles in terms of a few broad aggregates, see A. F. Burns, *Economic Research and the Keynesian Thinking of Our Times,* 26th Annual Report of the National Bureau of Economic Research, 1946, reprinted in A. F. Burns, *The Frontiers of Economic Knowledge,* 1954.

that, because of an increase in supply or a decline in demand, current prices cannot be maintained.[13] Production is cut back, orders to other manufacturers are reduced, employment is curtailed, and the disturbance spreads. On the reasonable assumption that prices and costs are not perfectly flexible and that the factors of production are not perfectly mobile, such a partial disturbance, if important enough, can lead to a general decline in business activity.

Partial overproduction of this sort can arise, among other reasons, because of "the competitive illusion." It takes time to meet an increase in demand. During the expansion, all firms see that demand is increasing; each seeks to satisfy as large a part of it as possible without knowing fully what its competitors are doing. For a while, particularly while new capacity is being built, it is not apparent that all of these production plans, when they eventually materialize as finished output, will yield a supply of goods in excess of demand at profitable prices. Eventually overproduction does develop in some industries; firms cannot sell at profitable prices what they expected; they curtail output; and the disturbance spreads. Presently these maladjustments are corrected, and a new recovery begins, although this type of theory does not pay much attention to the lower turning point.[14]

Explanations of this sort emphasize, as did Mitchell, the complexities of the modern business economy and concentrate primarily on the errors that are likely to arise from the inability of businessmen to foresee the future. The element of uncertainty and the aberrations of business psychology have been examined in detail by a number of writers, and some have argued that here is an important cause of possible disturbance that may operate directly on the entire economy, not merely through the development of horizontal maladjustments in certain industries. Thus, we may have a "psychological theory" of the business cycle.

A. C. Pigou, the well-known English economist, particularly emphasized this type of explanation.[15] When an expansion begins for

[13] Mitchell emphasized that maladjustments resulting from reduced profit margins occurred first in particular industries. In this sense, he was also concerned with horizontal maladjustments.

[14] For further discussion and detailed references, see Haberler, *op. cit.*, pp. 110–112.

[15] See his *Industrial Fluctuations*, 1927, esp. chaps. 6–7. This type of theory is briefly discussed in Haberler, *op. cit.*, chap. 6. One of the first writers to emphasize the psychological element as a cause of business fluctuations was John Mills

any reason, errors of optimism are inevitably generated, not merely in a few industries but through the entire economy. Business optimism improves not only in proportion to the improvement in the underlying "real" factors but more rapidly than this. In planning ahead, businessmen overestimate future demand and underestimate the future rise in costs. Such optimistic errors are especially likely in investment projects that will yield consumers' goods only in the relatively distant future.[16] For a while, this growing optimism feeds on itself. Finally, after a "period of gestation," an increased supply of goods comes on the market. Businessmen find that their rosy anticipations were unwarranted. A revulsion sets in, and the psychological pendulum now swings to the opposite extreme. Disappointment now breeds errors of pessimism; the pessimism feeds on itself and generates a cumulative downswing. Eventually, businessmen find that their pessimism has gone too far; some profitable opportunities remain and others eventually develop; the pessimism gradually dissipates; and a new psychological cycle begins. This type of explanation can be combined with other theories if one believes that other, nonpsychological factors may play a role, especially at the turning points.[17]

In recent years, economists have been placing an increasing emphasis on the element of expectations in all branches of economic theory. This has been true in the field of business cycles also. We have commented on the fact that Keynes did not formulate a fully developed theory of business cycles. When he did address himself specifically to the problem of cyclical fluctuations, it was the dependence of investment incentives on "the uncontrollable and disobedient psychology of the business world" that he chose to emphasize.[18] The business cycle results primarily from fluctuations in the volume of investment. These fluctuations in turn result from fluctuations in the marginal efficiency of capital—i.e., from changes

(not to be confused with John Stuart Mill) in a paper delivered in 1867. For a summary of his views, see Hansen, *op. cit.*, pp. 267–270.

[16] This point has been emphasized by Albert Aftalion. See "The Theory of Economic Cycles Based on the Capitalistic Technique of Production," *Review of Economic Statistics*, vol. 9, October, 1927, pp. 165–170. This article is a brief summary of his position, which was developed as early as 1909.

[17] Pigou believed that other factors were also important, but that the psychological factor alone was enough to generate fluctuations in business activity.

[18] *The General Theory of Employment, Interest, and Money*, chap. 22. The quotation is from p. 317.

in the rate of *expected* future return on current investment. Thus the marginal efficiency of capital—which, with the interest rate, determines the volume of investment—rises and falls with the state of business psychology. During booms, businessmen overestimate future returns; for the time being this stimulates investment, increases employment, and seems to justify the optimistic expectations. Eventually, realization dawns that anticipations have been too favorable.[19] The shock of this discovery leads to a collapse of confidence; the marginal efficiency of capital shifts downward suddenly; investment rapidly declines; and a cumulative downswing gets under way. Recovery comes when the accumulation of new investment opportunities—chiefly the need to replace worn-out and obsolete capital equipment and to replenish depleted inventories—reveals to businessmen that their pessimism has gone too far. Essentially, this is a psychological theory that centers on the influence of errors of optimism and pessimism on the volume of investment, to which (given the propensity to consume) Keynes assigns the key role in his theory of employment. As with other psychological theories, this explanation is weak in its consideration of the "real factors" that underlie the changes in business expectations. At best, Keynes offers only a few tentative suggestions as to why the optimism of the boom must eventually end, and his explanation of why the pessimism of the downswing ends is certainly incomplete.

Keynes' followers have, on the whole, not tried to develop further his suggestions for a psychological theory of the cycle. Instead, they have turned to a search for the underlying, "real" factors that bring about cyclical changes in the marginal efficiency of capital. We shall discuss some of their findings in the next chapter.

If we take these "business-economy" theories as a whole, they obviously contribute something to our knowledge of business cycles, although they fall short of telling the whole or perhaps even the most important part of the story. The emphasis placed on cost-price relations and on expectations, as well as the stressing of the interdependence of all parts of the economic system, is obviously essential

[19] At one point Keynes suggests two reasons why businessmen eventually realize that current expectations are too optimistic. The growing stock of newly produced capital goods begins to depress the current rate of profits, and the rising cost of new capital goods makes investment seem less profitable than formerly. *Ibid.*, p. 317.

to an understanding of the cumulative processes of expansion and contraction. Discussion of errors of optimism and pessimism and analysis of the possibility of horizontal maladjustments throw some light on the factors that operate at the turning points. But it is in their analysis of the underlying real factors at work at the turning points that these theories are weakest. This is particularly true with respect to what these explanations have to say about the reasons for wide swings in the volume of investment, without which no explanation of the business cycle can be reasonably complete.

Insofar as this range of theories has something valuable to say about the turning points, it is chiefly with respect to the turning points of the minor rather than the major cycles. As we saw in Chapter 11, the minor cycle results primarily from the reversals of short-term business expectations which are a natural concomitant of the minor maladjustments that are inevitable in a private-enterprise, profit-making economy. In terms of the saving-investment analysis, these minor cycles show themselves particularly in wide swings in investment in *inventories*. To explain the major cycle, we must explain why long-term investment expectations fluctuate as widely as they do. It is interesting to remark that Keynes took note of the existence of minor cycles in the United States, which he ascribed to fluctuations in inventories. These fluctuations occur because businessmen, in planning ahead, "make minor miscalculations." When these mistakes are discovered, production declines for a while to allow for the liquidation of excess inventories.[20] In this connection, it will be recalled that we have experienced only minor cycles since World War II, not only in the United States but also in other countries.

Other writers have also stressed the particular importance of inventory fluctuations in minor cycles—for example, Alvin Hansen, whose theory of major cycles we shall examine later in this chapter. But most writers who have attempted to incorporate inventory movements into their theoretical explanations have been rather vague as to the precise role that inventory behavior plays in the cycle and as to what, precisely, causes inventory investment to fluctuate in the way that it does.[21] The most important exception to

[20] *Ibid.*, p. 332.

[21] For a brief survey of the treatment of inventories in business-cycle literature, see Moses Abramovitz, *Inventories and Business Cycles*, 1950, chap. 1. We owe

this generalization is the theory of inventory cycles with which the name of Lloyd Metzler is associated. Metzler, using Keynesian tools, has developed a theory of minor cycles in terms of inventory fluctuations. However, it bears little relation to Keynes' suggestion. Whereas Keynes ascribed the minor cycle to the occurrence of minor miscalculations by businessmen and thus stressed the psychological factor, Metzler largely ignores the element of changing expectations. His theory runs as follows.[22]

Let an expansion in demand occur for any reason. Since production responds to an increase in demand only with a lag, the initial effect is an unplanned reduction in inventories. Seeking to maintain a constant ratio of inventories to sales, producers attempt to increase inventories. This expands output, income, and demand still further, so that inventories do not rise by as much as planned. But with a marginal propensity to consume of less than one, demand increases more slowly than does output. Hence, inventories begin slowly to accumulate, though less rapidly than businessmen desire. Eventually, producers re-establish the desired ratio of inventories to sales, and no further expansion of inventories is planned. This is equivalent to a decline in (inventory) investment; as a result, output and incomes fall, and the downswing begins. During the decline, the attempts of producers to reduce inventories as sales decline leads to further contraction of output, incomes, and sales. Eventually, however, since consumption does not decline as rapidly as incomes, inventories are reduced to the desired level. Once disinvestment in inventories stops, output rises to satisfy current and

---

much of our knowledge as to how inventories actually do behave in the cycle to this volume. Abramovitz's empirical findings have recently been extended into the postwar period by T. M. Stanback, Jr., in *Postwar Cycles in Manufacturers' Inventories* (National Bureau of Economic Research, in preparation).

[22] For a detailed exposition of Metzler's views see: "The Nature and Stability of Inventory Cycles," *Review of Economic Statistics,* vol. 23, August, 1941, pp. 113–129; "Business Cycles and the Modern Theory of Employment," *American Economic Review,* vol. 36, June, 1946, pp. 278–291; and "Factors Governing the Length of Inventory Cycles," *Review of Economic Statistics,* vol. 29, February, 1947, pp. 1–15. Metzler's theory is an example of the sort of business-cycle model, based on the interaction of the acceleration principle and the multiplier, which we shall discuss in more detail in the next chapter. Metzler's is the only one, however, that concentrates on changes in inventories rather than on the behavior of long-term investment. For an important earlier attempt to incorporate inventories into a cyclical model, see Erik Lundberg, *Studies in the Theory of Economic Expansion,* 1937, reprinted 1954.

expected sales. This increase in production expands income and sales, and a new cycle begins.[23]

The trouble with such a theory is its artificially precise character and its attempt to explain the minor cycle in purely mechanical terms.[24] So far as the minor cycle is concerned, Keynes was probably closer to the truth, if less precise, in his emphasis on "minor miscalculations" and changes in short-term expectations. A wide range of minor maladjustments can lead eventually to a reversal in short-term expectations. Only a minor cycle results if these maladjustments can be substantially corrected through changes in inventories without seriously affecting long-term anticipations.

MONETARY EXPLANATIONS

It is perfectly obvious that there is a monetary side to business fluctuations. Output expands and contracts with changes in total spending, and this spending is done with money. Indeed, as we saw in Chapter 2, we can represent the total of spending on newly produced output during any period by the expression MV, where M stands for the average amount of money (including bank deposits) in circulation and V stands for the income velocity of money. If total spending changes, M or V or both must change. In this sense, all business-cycle explanations have their monetary side; they must assume that either M or V changes to make possible the changes in spending that result from whatever causal factors are taken to be important.

But some writers go much further than this. They argue that the main reason for cyclical changes in total spending comes from the side of money. They explain the turning points in terms of changes in M, which, by influencing spending, cause changes in the level of business activity. In a modern credit economy the chief form of money is bank deposits, and it is the action of the banks

[23] A useful exposition and elaboration of Metzler's theory is provided in two articles by Ragnar Nurkse: "The Cyclical Pattern of Inventory Investment," *Quarterly Journal of Economics,* vol. 66, August, 1952, pp. 397 ff., and "Period Analysis and Inventory Cycles," *Oxford Economic Papers,* vol. 6, September, 1954, pp. 203–225.

[24] For example, he assumes that producers always strive for a constant sales-inventory ratio; he ignores the effect of changing price expectations; he assumes a constant marginal propensity to consume national income; and producers' expectations as to future sales are assumed to be geared in a fixed way to the past behavior of sales.

in expanding and contracting the money supply that is assumed to be the primary factor responsible for fluctuations in business activity. This is the approach of those who hold to a monetary theory of the business cycle.

Of course, economists have long been interested in the relations between the supply of money and the price level. But not many have developed a purely monetary theory of the business cycle. Since about 1900, study of the business cycle in monetary terms has proceeded chiefly along two lines. In England and the United States, a few writers, notably R. G. Hawtrey, have sought to demonstrate a direct relationship between the supply of money and the level of spending and income. On the continent—partly inspired by the pioneering work of Knut Wicksell in explaining the relations among the money supply, interest rates, the saving-investment process, and prices—there developed an "Austrian school" of economists who, while analyzing the business cycle in terms of the changing relations between savings and investment, found that these relations depended on the credit-creating activities of the banking system. In a later section we shall briefly consider this type of "monetary overinvestment" theory. At this point we want to look briefly at Hawtrey's and other monetary explanations that do not emphasize the saving-investment process.

Like many other economists, Hawtrey emphasizes the circuit flow of income and spending.[25] Unlike most recent writers, however, he believes that changes in the supply of bank credit are primarily responsible for cyclical changes in spending. In depressions, the banks lower interest rates in response to the accumulation of excess reserves. Low interest rates induce businessmen to borrow in order to expand their inventories. It is through inventory accumulation rather than long-term investment that the interest rate has its effect. Hawtrey differs from most cycle theorists in his emphasis on the key role of inventories and in his tendency to minimize the causal significance of changes in long-term investment.

[25] Hawtrey has been a prolific writer. The development of his ideas can be traced in the following books, among others: *Good and Bad Trade,* 1913; *Currency and Credit* 4th ed., 1950; *Trade Depression and the Way Out,* new ed., 1933; *Capital and Employment,* 1937; *A Century of Bank Rate,* 1938; and his essay on "The Trade Cycle," originally published in 1926; reprinted in *Readings in Business Cycle Theory.* For a summary of his views, see R. J. Saulnier, *Contemporary Monetary Theory,* 1938, Part I.

Thus, low interest rates lead business firms to expand inventories. This begins a cumulative expansion of production, incomes, and spending, which is reinforced by further borrowing from the banks. This process is ordinarily brought to an end by the banks. Eventually they find themselves short of reserves and are forced to raise interest rates and restrict credit. This forces businessmen to curtail inventories—and the cumulative sequence of falling production, incomes, and spending begins.

Hawtrey believes that when the Western world was on the gold standard and managed its monetary affairs accordingly—during the decades prior to World War I—this process of monetary expansion and contraction gave to business fluctuations their apparent periodic character. An expansion could continue only so long before the banking system would begin to lose reserves, either because of the rise in currency in circulation or because of gold exports. When the banks finally called a halt to further credit expansion, the drain on their reserves continued for a while. Hence, it became necessary actually to contract, and stabilization at the peak proved impossible. Conversely, reserves flowed into the banks during depressions; this forced down interest rates; and it was not long before business responded to the lower cost of credit and began again to expand.

This periodic quality of business fluctuations ceased with the disappearance of the gold standard. Fluctuations in business, arising chiefly from fluctuations in the supply of bank credit, still occur;[26] but they have largely lost their regular, wavelike character, since modern banking systems no longer are required to expand and contract in response to the inflow and outflow of reserves. Hawtrey also admits that occasionally a "credit deadlock" may occur, when business confidence becomes so seriously impaired that no easing of credit is by itself sufficient to bring about recovery. The Great Depression was one such occasion. However, such episodes do not happen very often.

Among recent American writers, Clark Warburton is the one who has held most uncompromisingly to a strictly monetary theory of business fluctuations. Despite his numerous articles on the subject,

[26] Hawtrey does not deny that nonmonetary factors may initiate a change in aggregate demand, but he argues that a cumulative process will not ensue unless the banking system permits the necessary changes in M or V to take place.

Warburton has never developed his theoretical explanation in any detail.[27] Briefly, he argues that economic stability requires that the money supply must grow at the same rate as the trend in total output (after allowance for any secular change in velocity that may be occurring). Monetary policy, operating on bank reserves, may prevent the supply of money (chiefly bank deposits) from expanding at the necessary rate. If such a monetary deficiency develops, prices fall, profit margins are reduced, inventories may accumulate, investment incentives are impaired, and output and incomes fall. If the money supply is excessive, prices and business activity increase.

Another American economist who has, in recent years, argued that there is a strong causal connection between the supply of money and the level of economic activity is Milton Friedman.[28] He has sought to demonstrate that the economy's demand for money is related in a stable and predictable way not to current income but to "permanent" income—i.e., to what income is expected to be over a number of years. If the money supply rises faster than permanent income, holders of money will try to spend their excess cash balances, and this will drive up money expenditures and incomes. This causes permanent income to rise, which in turn creates an additional demand for cash balances. Thus, the level of income and activity is sensitive to changes in the money supply. To keep the economy growing at a steady rate without serious depressions or marked inflation, the money supply must also grow at a steady

[27] Summaries of his position may be found in the following articles, among others: "The Theory of Turning Points in Business Fluctuations," *Quarterly Journal of Economics,* vol. 64, November, 1950, pp. 525–549; "The Misplaced Emphasis in Contemporary Business-Fluctuation Theory," *Journal of Business of the University of Chicago,* vol. 19, October, 1946, pp. 199–220; "Bank Reserves and Business Fluctuations," *Journal of the American Statistical Association,* vol. 43, December, 1948, pp. 547–558; "Volume of Savings, Quantity of Money, and Business Instability," *Journal of Political Economy,* vol. 55, June, 1947, pp. 222–233; and "How Much Variation in the Quantity of Money Is Needed?", *Southern Economic Journal,* vol. 18, April, 1952, pp. 495–509.

[28] As noted later in the text, Friedman has so far not presented a fully developed theory of business cycles. His views on the relations between the money supply and economic stability can be traced in the following: "The Demand for Money: Some Theoretical and Empirical Results," *Journal of Political Economy,* vol. 67, August, 1959, pp. 327–351, reprinted as National Bureau of Economic Research Occasional Paper 68, 1959; *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee, Part IV (86th Congress, 1st Session, 1959), pp. 605–637.

rate. While Friedman has not expanded his views into a full-fledged theory of business cycles (and in any event would not deny that nonmonetary factors may be of some importance), he argues that "in every deep depression monetary factors play a critical role."[29] In minor cycles, the money supply may be a contributing factor, but he grants that the chief causes of minor cyclical fluctuations may come from the nonmonetary side.

This is not the place for detailed criticism of monetary explanations of the cycle. Obviously, monetary processes cannot be ignored, and we must agree that fluctuations in aggregate demand are fluctuations in money spending. But there is little evidence that monetary causes have been dominant in the cyclical turns of the last 40 years or so. On both a priori and empirical grounds, it is virtually certain that Hawtrey exaggerates the importance of interest-rate changes on the level of business activity. In particular, short-term rates do not seem to play the role in stimulating and retarding inventory accumulation in the way that he suggests. Most important, the theoretical and empirical evidence seems overwhelming that nonmonetary forces, in the various ways described in Chapter 11, can and do lead to cyclical changes in total spending. We are probably safe in concluding that, though changes in the monetary variables (both M and V) are essential parts of the cyclical process, their role today is likely to be a passive one. They must change if spending is to vary, but they ordinarily change in response to the nonmonetary stimuli. This is not to deny that monetary conditions must be favorable if the nonmonetary stimuli are to have their full effect. And on occasion, of course, monetary policy and credit restriction may play an active, initiating role in cyclical developments.

THEORIES EMPHASIZING THE SAVING-INVESTMENT PROCESS

Students of the business cycle have for many years emphasized the crucial role of investment fluctuations in creating fluctuations in business activity. Recognition that the behavior of the saving-invest-

[29] Quoted from *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee, Part IV, p. 619. A similar position is taken by Gottfried Haberler, who believes that modern business-cycle theory "has tended grossly to underestimate the importance of the monetary factors." *Op. cit.,* p. 466.

ment process may contain the main key to an understanding of the causes of cyclical fluctuations goes back long before Keynes. Since about 1900, an increasing number of writers have sought to explain the business cycle in terms of the factors responsible for the extreme instability in the rate of capital investment.[30] For an even longer period, a succession of writers have sought to explain depressions in terms of the effect of saving in increasing productive capacity faster than consumers' buying power. Today, we can count a wide variety of explanations that have in common their concern with some aspect of the process whereby incomes are saved and resources are diverted to the production of capital goods. We shall discuss these theories under several headings in the sections that follow.

Most of the theories that emphasize the role of investment have the following points in common: (1) The impetus to cyclical fluctuations comes from the instability inherent in long-term investment. (2) During the boom the volume of investment rises to a level that it is impossible permanently to maintain. (3) The inevitable decline in the volume of investment sets off the cumulative deflation that results in depression. Where these theories differ is on the reasons for the turning points in investment—as to both why the investment boom begins and why it must end. Our classification in the following sections is based on the reasons offered for the ending of the investment boom. We refer to all of these explanations as "overinvestment theories" because they assume that business booms rest on a position of fundamental disequilibrium in the economic system—i.e., the expanding volume of investment that creates the boom sets to work forces that eventually lead to a decline in investment.

In addition to these various types of overinvestment theories, we must take account of a group of explanations that have come loosely to be called "underconsumption theories." Such explanations run the gamut from obvious absurdity to a high degree of

[30] Perhaps the first writer to evolve a well-developed theory of the cycle based on recurring fluctuations in real investment was the Russian, M. Tugan-Baranowsky, whose main work, later translated into German and French, was first published in Russian in 1894. See the French edition, *Les Crises Industrielles en Angleterre,* 1913. For a summary of his views, see Hansen, *Business-Cycle Theory,* pp. 62–67, and *Business Cycles and National Income,* pp. 277–291. For reference to some earlier writers who discussed the role of capital in business cycles, see F. A. Hayek, *Prices and Production,* rev. ed., 1935, pp. 101–104.

economic sophistication. Insofar as such theories find the reason for "underconsumption" in the behavior of saving, they also are concerned with an aspect of the saving-investment process. In the next chapter we shall look at some of these underconsumption arguments, including those which can properly be called oversaving theories.

THE SHORTAGE-OF-CAPITAL THEORIES

One type of explanation holds that the boom ends because eventually the volume of investment becomes too large relative to the flow of savings; thus a "shortage of capital" develops which forces a decline in investment and in the level of business activity. Shortage-of-capital theories fall into two groups. One finds the reason for the excess of investment during the boom in the operation of the monetary system; the other emphasizes the nonmonetary factors that lead investment to outrun the supply of savings and then to decline.

The monetary type of shortage-of-capital or overinvestment theory is particularly associated with the name of F. A. Hayek, an Austrian economist who eventually moved to the United States.[31] It is also held by several other economists of Austrian origin and has been espoused by a few English and American writers.

The monetary overinvestment theory in the particular form developed by the Austrian school is a highly sophisticated explanation which impresses the reader both by its theoretical subtlety and also by its lack of touch with reality. This type of explanation stems from the work of the Swedish economist, Knut Wicksell, who showed that if the market (long-term) rate of interest falls below the "natural" or "equilibrium" rate,[32] the level of investment will rise above the level of saving, the excess being financed by new bank credit. This creates a cumulative rise of incomes and prices. Actually, Wicksell used this approach to explain changes

[31] See F. A. Hayek, *Monetary Theory and the Trade Cycle* (published in German in 1929 and in English in 1933) and *Prices and Production*, rev. ed., 1935. See also Hayek's essay, "Price Expectations, Monetary Disturbances and Malinvestments," in *Profits, Interest, and Investment*, 1939. This last essay has been reprinted in *Readings in Business Cycle Theory*, pp. 350–365.

[32] The "natural" or "equilibrium" rate of interest can be defined as the rate of interest at which the demand for loanable funds just absorbs the current flow of saving.

in the price level and emphasized nonmonetary factors in talking about fluctuations in the level of business activity.[33]

Hayek and others developed this approach into a full-fledged theory of the cycle. In depression, the market rate of interest eventually falls below the equilibrium rate. As a result, more capitalistic methods of production become profitable, and investment is stimulated. Investment rises above the level of saving, the difference being financed by new bank credit. The expanded money supply enables businessmen to divert resources away from the consumers'-goods industries to those producing capital goods. This additional spending causes money incomes and prices to rise. However, as incomes rise, so does consumption—so that consumers now bid against the capital-goods industries for the available supply of resources. Costs in the capital-goods industries therefore rise. Eventually, the banking system must reduce the rate of credit expansion, at the same time that resources are becoming scarce in the capital-goods industries. The result is that the market rate of interest rises above the equilibrium rate, investment declines, and the downswing begins.

It is important to note that, according to this explanation, it is a *shortage* of saving that brings the boom to an end. The banks encouraged businessmen to expand the volume of investment above that which consumers were willing to finance by voluntarily saving out of their incomes. Credit expansion permits this discrepancy to exist for a time.[34] When the expansion of credit stops, a decline is inevitable unless there is a sudden increase in saving. Since this does not occur, there is a shortage of capital funds; interest rates rise; some existing capital projects cannot be completed; investment de-

[33] For an exposition of Wicksell's position, see his *Interest and Prices* and *Lectures in Political Economy,* vol. 2, 1935, esp. pp. 190–214; also "The Enigma of Business Cycles" in International Economic Association, *International Economic Papers,* no. 3, 1953, pp. 58–74. For a recent summary of Wicksell's contributions to economics, see C. G. Uhr, *The Economic Doctrines of Knut Wicksell,* 1960.

[34] The expansion of credit and rise in prices result in "forced saving." The additional purchasing power made available for investment is used to bid resources away from the consumers'-goods industries. Prices of consumers' goods rise, and consumers find that their money incomes buy less goods than formerly. Consumers are forced to do without some things they bought before; in a sense, they have been forced to save—i.e., not to consume. The outstanding example of this type of phenomenon occurs during war inflations. Even if people do not save voluntarily, the rise in prices means that they can consume less, and thus resources are released to the government for war purposes.

clines; and the downswing begins. Thus, in a literal sense, the boom is a phenomenon of *overinvestment*—that is, too much investment in capital goods relative to the supply of savings.

There is no need to criticize this type of theory in detail. Although credit tightness and a shortage of loanable funds played some role in many of the American downturns before 1914, there is little evidence in the business-cycle history of the last 40 years to support this type of capital-shortage theory.[35] The theory is unrealistic to an extreme degree, particularly in its analysis of the investment process. It ignores nearly the whole range of factors that influence business expectations and the evaluation of investment opportunities by businessmen. It grossly exaggerates the importance of the interest rate in influencing the volume of investment. It largely ignores the role of technological change. It assumes that the volume of investment changes only because changes in the interest rate, relative to the equilibrium rate, make it more or less profitable to engage in more capitalistic methods of production; i.e., to change the ratio of capital to output. No attention is paid to the ways in which investment is linked to consumers' demand, to the possibility of investment saturation in particular industries, to the rough working of the acceleration principle, to the various sorts of maladjustments that affect investment in particular lines and spread through the whole economy, and so on.

A few writers have emphasized capital shortage as the primary reason for the downturn without placing the same emphasis on the monetary side as do Hayek and others of the Austrian school. This type of nonmonetary capital-shortage explanation is associated especially with the names of Arthur Spiethoff in Germany and Gustav Cassel in Sweden, both of whom did their most important work before 1930.[36] Both writers, who were strongly influenced by

[35] It is of course possible to argue that, though investment booms must eventually end for other reasons, a shortage of loanable funds may on occasion end the boom before that point is reached. See, for example, D. H. Robertson, *Banking Policy and the Price Level,* rev. ed., 1949, pp. 87–91; Haberler, *op. cit.,* pp. 367 ff.

[36] The best-known statement of Spiethoff's position is his long article, "Krisen," in *Handwörterbuch der Staatswissenschaften,* 4th ed., vol. 6, 1925, pp. 8–91; see especially pp. 70–86. This has now been translated into English in slightly abridged form, with a new preface, as "Business Cycles," in *International Economic Papers,* no. 3, pp. 75–171. A summary of his views will be found in Ha-

Tugan-Baranowsky, emphasized the role of nonmonetary factors in bringing about fluctuations in long-term investment. The boom represents a spurt in the production of capital goods to a level higher than is warranted by the current flow of saving. The enlarged supply of capital goods eventually proves to be too great. This shows itself in a number of ways. Interest rates rise and new funds become scarce, making investment projects more difficult to carry through than formerly. Wages and the prices of materials rise, increasing the cost of producing capital goods and reducing profits (and hence saving). The volume of saving is not large enough to divert from the consumers'-goods industries the labor and other resources needed by businessmen if they are to undertake new investment projects on the scale required by the expanded capacity of the capital-goods industries. In brief, the boom ends because the supply of saving is not large enough to finance the high level of investment that would be required to absorb the enlarged output of the capital-goods industries. There is at one and the same time a shortage of capital and overproduction of capital goods. A cumulative decline now ensues, which feeds on itself until something happens to stimulate investment again. This may be new technological developments or the opening up of new markets (both emphasized by Spiethoff), the need to replace worn-out equipment, the gradual improvement of business confidence, or a fall in interest rates to a point where new borrowing again becomes profitable.

Both Spiethoff and Cassel emphasize that spurts in investment are inevitable in a progressive economy. Inventions, the opening up of new territory, and so on create new investment opportunities; an investment boom follows which outruns the supply of saving; and a shortage of capital brings the boom to an end. Many other writers have emphasized the role of innovations in initiating cyclical booms but find the reasons for the ending of the boom in factors other than a shortage of capital. This is true, for example, of Schumpeter and Hansen, whose views we shall examine in the next section. Thus two writers may both hold to an "innovations" theory

---

berler, *op. cit.*, pp. 72–84; see also Arthur Schweitzer, "Spiethoff's Theory of the Business Cycle," *University of Wyoming Publications*, vol. 8, April, 1941, pp. 1–30. Cassel's views are conveniently summarized in his *The Theory of Social Economy*, 1924, esp. chap. 19.

of the lower turning point but may have different theories to explain the upper turning point.[37]

Before we conclude this section, brief mention should be made of the revised form in which Hayek has presented his theory, making use of a theoretical relationship which he calls the "Ricardo effect."[38] The result is an even more implausible explanation than that contained in his earlier writings. The "Ricardo effect" states that a rise in real wages leads to a substitution of capital for labor in production (and the reverse if wages fall). Hayek's argument is as follows: In depression, prices fall more than wages, thus reducing profits. With wages relatively high, businessmen begin to substitute capital goods for labor. Thus investment begins to rise. This increases incomes and the demand for consumers' goods, whereupon, through the acceleration principle, there is a further increase in investment and a further rise in incomes and consumption. Eventually the rising demand for consumers' goods causes prices to rise. Since wages do not rise as rapidly as prices, profit margins increase. Now, since labor is relatively cheap, businessmen seek to substitute labor for capital. For a while, the stimulating effect of rising consumers' demand, working through the acceleration principle, offsets this tendency, and investment continues to rise. Eventually, however, the shift from capital to labor becomes so strong that investment actually declines. This tendency is aided by the rise that occurs in the cost of capital goods. Thus, investment eventually falls despite the fact that consumers' demand is still rising. Its fall leads to a decline in incomes, consumption, and profits in cumulative fashion, and the downswing goes on until falling prices reduce profits to the point where there is again a shift from labor to capital.

As in his earlier version, Hayek finds the difficulty in too much consumption and too little saving in the boom. As before, also, he looks on capital as being freely substitutable for labor and finds

[37] Spiethoff also recognized that an investment boom might end because investment opportunities had become saturated for the time being. He thought this was more likely in old than in young countries. Hence, to this extent he had two alternative explanations of the upper turning point. Cf. his article, "Business Cycles," p. 158. It should also be noted that, in his new preface to this article, written in 1953, Spiethoff emphasized that his theory applies chiefly to the century ending in 1913. Another writer with a similarly flexible view is William Fellner. See his *Trends and Cycles in Economic Activity*, 1956, part 4.

[38] See *Profits, Interest, and Investment*, pp. 3–71.

the cycle to be the result of the changing ratios of capital to output. These changing ratios are brought about by changes in the profit rate instead of, as in his earlier explanation, by changes in the rate of interest. Neither version is of sufficient practical importance to warrant further discussion here.[39]

INVESTMENT OPPORTUNITIES AND THE ROLE OF INNOVATIONS

The preceding section dealt with a group of theories that find the main reason for the ending of the boom in the fact that investment becomes too large relative to the current flow of saving. In the next chapter we shall consider a range of explanations that take essentially the opposite view—that the boom ends because investment becomes too large relative to consumption or total output. Both of these groups of theories are concerned with the relation between broad aggregates—between total investment and a few other aggregative variables.

Investment is assumed to depend in a fairly direct way on these other variables. Thus, in the capital-shortage theories, investment is made to depend on such variables as the interest rate, the relative prices of capital and consumers' goods, and so on. In the type of "multiplier-accelerator" theory to be discussed in the next chapter, investment is assumed to be "induced" primarily by one factor alone—the amount of change in final output.

In this section we shall consider a type of explanation that looks in a different direction to account for wide cyclical swings in private investment. Emphasis is placed on the opening up and subsequent temporary exhaustion of "investment opportunities." These opportunities for profitable investment depend on more than the growth of total output and only to a very minor extent on the behavior of the interest rate. They are a product particularly of technological change but can also be created by changes in business organization, the opening up of new territory, spurts in population growth, and the like. New investment opportunities do not affect all parts of the economy equally. They lead to the creation of particular new products or industries or to new ways of producing

[39] For further discussion, cf. Wilson, *Fluctuations in Income and Employment*, chap. 6, and Sho-Chieh Tsiang, *The Variations of Real Wages and Profit Margins in Relation to the Trade Cycle*, 1947, chap. 7.

old goods or to new markets. Usually, the creation of something new means also the obsolescence of something old. Most important for the study of growth and cyclical fluctuations, the opening up and exploitation of investment opportunities do not take place smoothly. Economic growth tends to come in spurts, accompanied by booms and depressions.

The writers in this group place primary emphasis on the role of innovations in creating business fluctuations—not only technological change but the opening up of new markets and new sources of raw materials, the introduction of new forms of business organization, and the like. The (major) business cycle results from the irregular way in which the economy adjusts itself to new products and to new ways of doing things. Of course, many writers agree in recognizing the role of innovations as a factor making for instability. What distinguishes the type of explanation now to be considered is (1) the emphasis placed on innovations as a cause of instability and, more important, (2) the reasons offered for the end of the investment boom.

In broad outline, this type of theory runs as follows: After a downswing has gone on for some time, the accumulation of innovations that have reached the stage of commercial application since the preceding boom creates a new set of investment opportunities. As a result, investment expands, and a cumulative upswing begins. The boom ends because these particular investment opportunities eventually become exhausted, at least for the time being. The trouble is not a shortage of loanable funds or a failure of total consumption to rise sufficiently. The cause of the boom was the opportunity to invest in new types of capital equipment—in order to reduce costs, to produce new products, to serve a new market, and the like. Once the necessary investment has been made in these lines, businessmen can satisfy the expected demand for their products with no further investment except for replacement purposes. Investment in these fields would inevitably decline even if businessmen had perfect foresight and no miscalculations of future demand were to occur. The trouble arises because investment during the boom has been concentrated in particular directions, and such a level of investment in these fields cannot be maintained indefinitely —no matter how total consumption and total saving behave. Of course, the situation is nearly always made worse by the fact that

mistakes are made so that the boom does result in some degree of eventual overcapacity in the fields most stimulated by innovations. Hence, the cycle is more violent than would have been the case if the overbuilding had not taken place. But even without such overinvestment (in the literal sense), cycles in investment and in business activity would result from the way in which investment opportunities are created by innovations and then filled by subsequent investment.

The best-known exponent of this type of business-cycle theory is Joseph Schumpeter, whose explanation of business fluctuations is a part of a broad and brilliantly conceived theory of economic development. A similar position is held by A. H. Hansen, who is better known as the leading American disciple of J. M. Keynes, and by the British economist D. H. Robertson.

This explanation of the upper turning point should be distinguished from that offered by theories that emphasize horizontal maladjustments and the "competitive illusion."[40] The latter emphasize that eventually, after a "period of gestation," the new investment pours forth a flood of goods which cannot be sold at profitable prices. In the type of theory now being considered, investment declines because the new investment opportunities created by innovations are satisfied. Investment declines even though prices, costs, and profits are fully in accord with the expectations that led to the new investment. The writers in this group do not ordinarily put much emphasis on the acceleration principle, but obviously it plays a key role in this type of explanation. Innovations create the need for new capacity of particular sorts. Once the desired capacity is built, no further new (net) investment is necessary except as the demand for the products in question expands further.

Schumpeter adds a good deal to the skeleton of an explanation outlined in the preceding paragraphs.[41] He attempted to explain why innovators appear in clusters so that economic development proceeds in spurts rather than continuously.[42] Further, the innovations are financed by bank-credit expansion, which increases purchasing power and prices and helps to create a cumulative expansion

[40] See pp. 346–347, above.

[41] See his *The Theory of Economic Development*, 1934, chap. 6, which was first published in German in 1911, and *Business Cycles*, 1939, esp. chaps. 3–4. For a briefer statement, see "The Analysis of Economic Change," reprinted in *Readings in Business Cycle Theory*, pp. 1–19.

[42] See, for example, *The Theory of Economic Development*, pp. 228–230.

throughout the economy. Overoptimism and speculation add to the boom. The decline finally comes for two main reasons. The opportunity for investment in the new fields begins to decline as capital projects are completed and the new products come on the market. The innovators are thus permitted to pay off their bank borrowings, which leads to a general credit deflation. Secondly, the new products and new processes resulting from the innovations drive down prices and create difficulties for old firms.[43] The latter contract output; some are forced into bankruptcy; and the economy goes through a painful period of readjustment to the altered economic relationships created by the new innovations.

Since the boom is usually characterized by speculative excesses, the decline proceeds further than is necessary to reëstablish equilibrium. A "secondary deflation" is added to the initial decline. Eventually the natural forces of recovery bring about a revival up to the new equilibrium position. At this point, the more able and courageous entrepreneurs begin with a new set of innovations, others follow, and a new boom begins.

The concept of equilibrium plays a prominent part in Schumpeter's discussion, as is seen in the way he classifies the phases of the cycle. He assumes that the investment boom (the prosperity phase) begins from a position of equilibrium. The recession marks the movement back to a new position of equilibrium (with larger output and lower costs and prices than before the boom began). But the secondary deflation drives output and prices below this new equilibrium level (this is the depression phase), and the economy returns to the new equilibrium position during revival. Thus, revival is considered the last and not the first phase of the cycle.

As we saw in Chapter 9 (page 239), Schumpeter believes that business cycles of different durations occur simultaneously and react on each other. All are generated by innovations; the different durations arise from the fact that "the periods of gestation and of absorption of effects by the economic system will not, in general, be equal for all the innovations that are undertaken at any time."[44] As a working rule, he believes that the simultaneous cycles operating during the

[43] Further difficulties are created by the fact that the new innovating concerns have bid up the prices of the factors of production and thus increased costs for the old firms.

[44] *Business Cycles,* vol. 1, pp. 166–167.

last 150 years or more can be divided into 60-year waves (Kondratieffs), 10-year cycles (Juglars), and 40-month cycles (Kitchins). Shortly before his death, Schumpeter changed his position slightly and admitted that the 40-month cycle might be primarily an inventory phenomenon rather than the result of innovations to which the economy adjusts itself relatively quickly.[45]

Hansen's position is in some respects similar to Schumpeter's, but with strong Keynesian overtones.[46] He attributes the minor cycle chiefly to inventory fluctuations. Major cycles result from spurts in long-term investment. The boom begins primarily because of the new investment opportunities opened up by technological progress. The boom continues both because of the upsurge of "autonomous" investment resulting from new investment opportunities and because of the additional investment "induced" by the rise in income that occurs.

We have here a twofold explanation of the spurt in investment during the boom. Autonomous investment rises because of the new investment opportunities created by technological change, the opening up of new markets and new sources of raw materials, and so on. Through the multiplier, this rise brings about a rise in incomes. Now the acceleration principle enters to cause a further rise in (induced) investment. Thus the boom rests on two precarious supports—a rate of autonomous investment that cannot be indefinitely maintained and a level of induced investment that requires the continued maintenance of the past rate of increase in total output. The boom collapses when these two supports give way.[47] Induced investment declines when the rate of increase in output flattens out as full employment is approached. More important, "the available *autonomous*

[45] See his paper, "Historical Approach to the Analysis of Business Cycles," in Universities-National Bureau Committee, *Conference on Business Cycles,* 1951, p. 154.

[46] See his *Fiscal Policy and Business Cycles,* 1941, chaps. 1, 11–12, 14, 16, and his more recent statement in *Business Cycles and National Income,* particularly chaps. 11, 23–24.

[47] In his earlier writings, Hansen emphasized almost exclusively the role of autonomous investment. New investment opportunities led to a spurt in investment which came to an end when these opportunities became exhausted. More recently, in *Business Cycles and National Income,* he has given the acceleration principle a more important place in his theory. But he still places his main emphasis on the spurts in autonomous investment, which he looks on as the primary *initiating* factor in business fluctuations.

investment becomes progressively exhausted the longer the boom lasts." "Thus the boom 'dies a natural death.' Investment has caught up with the requirements of growth and technical progress."[48]

Hansen ties this analysis of investment fluctuations in with the Keynesian emphasis on the consumption function. It is the slope of the consumption function that creates a gap between income and consumption that must be filled by investment. When income is at a high level, savings are large. Hence, investment must be maintained at a high level or a cumulative downswing will begin. Hansen's emphasis on this point has led to his sometimes being called an "underconsumptionist," a position he has vigorously disavowed.[49] He has stated his theory in Keynesian terms, but this does not alter its essential elements. Once investment begins to decline because investment opportunities are temporarily exhausted, it is true that a sudden rise in consumption and decline in saving would permit aggregate demand to be maintained. But the same can be said of any business-cycle theory. If investment declines for any reason, output and employment will fall unless there is a radical upward shift in the consumption function. The important question is: Given the savings habits of the economy, why does investment decline? Hansen's answer is the temporary exhaustion of investment opportunities created chiefly by innovations plus the workings of the acceleration principle—not that total consumption is too low to maintain the level of investment achieved during the boom. He holds to an "investment opportunity" explanation, to which he has added the sort of acceleration-principle theory that we shall discuss in more detail in the next chapter.[50]

In addition to his theory of business cycles, Hansen also has a theory of secular stagnation which he feels fitted the conditions existing in the United States before World War II. During the nineteenth century, three powerful sets of forces operated to maintain the level of investment: rapid population growth, the opening up of new ter-

[48] *Business Cycles and National Income,* p. 496; compare *Fiscal Policy and Business Cycles,* p. 226.

[49] *Fiscal Policy and Business Cycles,* p. 297 n.

[50] Although he puts it in different language and with a different emphasis, Fellner also argues that there must be a steady stream of technological-organizational improvements sufficient to generate the investment needed to absorb the saving made by a growing economy. The argument is put in terms of the need to overcome scarcities. Cf. his *Trends and Cycles in Economic Activity.*

ritories and new sources of raw materials, and technology. Like many others, Hansen thought that the rate of population growth was declining, and it is true that the period of territorial expansion is largely over. Hence, we are left with technology as the main stimulus to investment in the future, and this cannot be expected to generate the level of investment needed for full employment. Hence the need for public investment and for measures that will raise the community's propensity to consume.[51] This might be termed an "underinvestment theory" of the long-run level of employment. It is not a theory of the business cycle, nor was it intended to be one.

Another writer who emphasizes innovations and the unevenness of economic growth in his explanation of business cycles is D. H. Robertson, the noted English economist.[52] Robertson has analyzed a number of different factors that may be responsible for cyclical fluctuations, but he puts considerable emphasis on innovations and on the recurring exhaustion of the investment opportunities thus created. Somewhat similar positions, with various modifications, have been taken by J. M. Clark, Thomas Wilson, and David McCord Wright.[53]

We have already emphasized—as have many other observers—that each business cycle differs in important respects from every other cycle. Hence, no single simple explanation can explain all cycles. However, history suggests that technological change and the irregular way in which investment opportunities have been exploited have played a strategic role in many major cycles, and we have given these factors a prominent place in the theoretical synthesis we developed in Chapter 11.

[51] See *Fiscal Policy and Business Cycles,* esp. chap. 17; also "Economic Progress and Declining Population Growth," reprinted in *Readings in Business Cycle Theory,* pp. 366–384. See also p. 448, below, and the additional references cited there. Keynes also held to the view that a mature economy has difficulty in generating the investment necessary for full employment.

[52] See *A Study of Industrial Fluctuations* (originally published in 1915 and reprinted with a new introduction in 1948), esp. pp. 66–68, 182–183, 239–241; and *Banking Policy and the Price Level,* chap. 2.

[53] See J. M. Clark, *Strategic Factors in Business Cycles,* 1935; Thomas Wilson, "Cyclical and Autonomous Inducements to Invest," *Oxford Economic Papers,* vol. 5, March, 1953, pp. 65–89, and *Fluctuations in Income and Employment,* chap. 7; D. McC. Wright, *The Economics of Disturbance,* 1947, esp. chaps. 5–6. Cf. also R. A. Gordon, "Investment Behavior and Business Cycles," *Review of Economics and Statistics,* vol. 37, February, 1955, pp. 23–34.

CHAPTER 13

THE VARIETY OF BUSINESS-CYCLE THEORIES (*Continued*)

WE SHALL now look at those theories that stress the dependence of investment on the behavior of total income or output—and not on the availability of saving, the movement of interest rates, or the impact of innovations. This is the type of explanation covered under heading 3c in the outline on page 343.

In one or another form, this is the kind of explanation that has dominated the theoretical literature on business cycles and growth in the years since World War II. Theories of this type have a number of characteristics in common. They are Keynesian in the emphasis they place on the interrelations among income, consumption, and investment. They all place primary emphasis on *induced* investment, which takes place in response to actual or anticipated increases in output, rather than on *autonomous* investment, which is brought about by other factors, particularly innovations. They all use some form of period analysis, in which various kinds of lags (for example, between the receipt of income and expenditure on consumption, or between an increase in demand and the resulting investment) play an important role. They involve a high degree of abstraction and are frequently expressed in mathematical form. Hence, they are often referred to as business-cycle or growth "models." They all rest ultimately on the fact that investment affects the level of economic activity by influencing both aggregate demand and aggregate supply. Through the multiplier, rising investment increases aggregate demand. But investment, by creating new capacity, also increases aggregate supply. Economic stability and continued growth depend on an

appropriate balancing of these demand and supply effects of investment.

We have already looked at very simple examples of this type of theory in Chapter 6, and the reader might well want to review that discussion at this point. There we saw how the interaction of the multiplier and accelerator could generate cycles, and on pages 150–153 we examined briefly the Harrod-Domar type of growth model, which also depends on the interaction of the multiplier and accelerator.

RECENT BUSINESS-CYCLE AND GROWTH MODELS

Business-cycle and growth models of the type that we are now discussing are sometimes referred to as capital-stock adjustment theories. Investment occurs because of the need to adjust the stock of capital to the changing level of output; the relation between the level of output and the size of the capital stock is taken to be the most important influence determining the behavior of investment. But investment, in turn, operating through the multiplier, determines the level of income, which, taken in conjunction with the growing capital stock, determines the future level of investment and (through the multiplier) the future level of income—and so on. If investment increases, it not only leads to a rise in income but also brings about an increase in the capital stock and in productive capacity. The important question is whether this process generates a steadily rising level of output and investment, or whether it will lead to fluctuations.[1]

All theories of this type use the multiplier relationship to show how a changing level of investment causes changes in the level of income or aggregate demand. The main differences among them come from the way they explain the capital-stock adjustment process itself —that is, how investment reacts to changes in income. On this score, these theories divide roughly into two broad groups: (1) those that explain the behavior of investment by the short-run working of the acceleration principle, in much the way illustrated by the multiplier-accelerator model presented on page 146, and (2) those that use a

[1] For a more extended discussion of this question, see D. Hamberg, *Economic Growth and Instability,* 1956, esp. chap. 2. For a general discussion of the principles underlying the construction of business-cycle models, see R. M. Goodwin's chapter on "Econometries in Business-Cycle Analysis," in A. H. Hansen, *Business Cycles and National Income,* 1951.

more flexible theory of investment behavior, in which investment is not rigidly tied to changes in income or output. This second group usually makes the level of investment depend on the level of the capital stock in the recent past and on the recent or current level of output or profits.[2]

These models differ among themselves in other respects besides their explanation of investment behavior. Some emphasize growth and say little about cyclical behavior, while others have the opposite emphasis. There is some variation in the handling of the consumption function. Various sorts of lags may be introduced. A very few attempt to take limited account of monetary factors.[3] But the main difference, as noted, is in the treatment of investment—that is, whether the acceleration principle or a broader theory of investment behavior is used.

These business-cycle and growth models are of relatively recent vintage. Among the pioneers in this field were R. F. Harrod, Erik Lundberg, and Paul Samuelson. In 1936, the year in which Keynes' *General Theory* was published, Harrod explored the business-cycle implications of the interaction of the multiplier and accelerator.[4] He made an assumption, however, that has since been discarded: that the marginal propensity to consume (and therefore the multiplier) declined as the upswing proceeded. Harrod is much better known for his later work on the theory of growth, in which he expressed the conditions for an equilibrium rate of growth in terms of the interaction of the propensity to save (which determines the multiplier) and the accelerator.[5] His growth model, which has come to be known as the Harrod-Domar model, was discussed in Chapter 6 (pages 150–153).[6]

A number of dynamic "model sequences" involving the interaction of income, consumption, saving, investment, and other variables

[2] This type of investment equation was discussed briefly in Chapter 6. See p. 131. We noted there that the acceleration principle is a special case of this more flexible formulation.

[3] One of the few to do this is Hyman Minsky. See his "Monetary Systems and Accelerator Models," *American Economic Review*, vol. 47, December, 1957, pp. 859–883. Another is James Tobin, in "A Dynamic Aggregative Model," *Journal of Political Economy*, vol. 63, April, 1955, pp. 103–115. Cf. also the summary of Hicks' model, p. 375, below.

[4] R. F. Harrod, *The Trade Cycle*, 1936.

[5] R. F. Harrod, *Towards a Dynamic Economics*, 1948.

[6] The reason for the inclusion of Domar's name is explained in the pages cited.

were presented by Erik Lundberg in 1937 in a book that is not so well known by economists as it should be.[7] Widespread interest in multiplier-accelerator models largely dates from 1939, when Paul Samuelson explored the dynamic properties of models of this type in some detail.[8] Since then, a large number of models have been developed. As we suggested earlier, they fall broadly into two groups: those that emphasize the dependence of investment on the accelerator and those that incorporate a broader and more flexible theory of investment. Either type may put its primary emphasis either on cyclical fluctuations or on long-term growth. The recent tendency has been to try to develop models that tell us something about both the causes of the business cycle and the factors that determine the rate of long-term growth.

ACCELERATOR MODELS

Probably the best known of the cycle-growth models that rest directly on the acceleration principle is that of the British economist, J. R. Hicks.[9] Hick's model shows clearly the influence of the earlier work done by Harrod, Samuelson, and some of the other model builders. It rests directly on the simple interaction of the multiplier and the acceleration principle, modified to allow for the effect of various types of lags and for the fact that the economy experiences not only cyclical movements but also secular growth. Hicks, like Hansen, assumes the existence of both induced and autonomous investment; but, unlike the latter, he finds the primary reason for the cycle in the effect of the acceleration principle on induced investment. His model is different from those of his predecessors in several respects. He assumes that there is an upward secular trend in output and that

[7] *Studies in the Theory of Economic Expansion* (reprinted 1954). Another early "model builder" was Michal Kalecki.

[8] "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Statistics,* vol. 21, May, 1939, pp. 75–78; reprinted in American Economic Association, *Readings in Business Cycle Theory,* 1944, pp. 261–269.

[9] J. R. Hicks, *A Contribution to the Theory of the Trade Cycle,* 1950. For an interesting critique of Hicks' theory and an attempt to generalize it to allow for more than the interaction of an unchanging accelerator and multiplier, see S. S. Alexander, "Issues of Business Cycle Theory Raised by Mr. Hicks," *American Economic Review,* vol. 41, December, 1951, pp. 861–878. Among other critical evaluations, see James Duesenberry, "Hicks on the Trade Cycle," *Quarterly Journal of Economics,* vol. 64, August, 1950, pp. 464–476; and Arthur F. Burns, *The Frontiers of Economic Knowledge,* 1954, pp. 236–267.

there is an equilibrium rate of growth that would call forth just the investment needed to absorb the savings generated by a steadily rising volume of income. The cycle is assumed to fluctuate around this upward trend. Secondly, he makes a painstaking analysis of the sort of lags that should be expected between a change in output and the resulting investment and between income and consumption. He also assumes that autonomous investment rises at a steady rate, whereas earlier model builders had neglected this type of investment.

With these modifications, Hicks shows how the interaction of the multiplier and the acceleration principle will generate a series of cycles around the upward equilibrium trend in output. These cycles may be explosive or damped; that is, they may become more and more violent or their amplitude may diminish until the cycles virtually disappear. What happens depends on the size of the accelerator (the amount of investment required per unit increase in output) and on the propensity to consume.[10]

Hicks believes that, once started, cycles tend to be "explosive" in an upward direction. What stops the boom is simply a shortage of resources. When this ceiling is reached, output can expand only at the rate permitted by the upward secular trend. The slower rate of expansion, through the acceleration principle, causes a decline in investment, which causes a decline in income, and thus a cumulative downswing begins.

During the downswing, gross induced investment falls to zero (i.e., even replacement expenditures are halted), but it cannot fall lower. Since there is this limit to the reduction in investment, the acceleration principle ceases to function. Further declines in output cannot make gross investment less than zero. Output thus tends to settle down at a level governed by the current amount of autonomous investment. But Hicks assumes that autonomous investment, which is not geared to current output, continues to expand under the influence of continuing technological change. Output therefore will eventually begin to rise again. As soon as this new expansion, together with the failure to replace worn-out equipment, wipes out excess capacity, the acceleration principle again takes hold; and its interaction with the multiplier generates a new cumulative upswing.

Hicks makes some allowance for possible changes in the rate of growth of autonomous investment, but this does not play an impor-

[10] See the discussion of this point, page 147, above.

tant part in his analysis. The cycle is analyzed almost entirely in terms of induced investment. He also considers the possibility that credit tightness might be responsible for the slowing down in the rate of expansion which brings on the initial decline in investment. He concludes that it is ordinarily a "real" rather than a monetary ceiling that stops the upswing, but that the monetary deflation and rise in liquidity preference that occur during the downswing do much to accentuate the severity of the depression.

A more flexible version of the capital-adjustment process, although it also depends ultimately on the accelerator, is provided in the model suggested by Richard Goodwin.[11] As in the strict accelerator-type models, investment generates income through a multiplier, the value of which depends on the marginal propensity to consume. But the behavior of investment is determined somewhat differently than in Hicks' model. Investment in any period will be some fraction of the difference between the capital stock that is desired by businessmen, given the current level of output, and that which actually exists. We can write this as:

where
$$I = a(\bar{K} - K)$$
$$\bar{K} = \beta Y + \phi(t).$$

The first of these two equations states that current investment will be some fraction of the difference between the *desired* stock of capital $(\bar{K})$ and the *actual* stock (K). The second equation says that the desired stock of capital depends on the product of the accelerator and the current level of output (βY), but that it also tends to increase independently through time (this is the meaning of $\phi[t]$), chiefly because of new technological change that increases the need for capital.

In Goodwin's model, once output begins to rise after a depression, the stock of capital desired by businessmen will rise; and hence investment will increase, thus increasing income, output, and the desired capital stock still further. Eventually, investment can rise no further because the capacity of the capital goods industries is limited. Once investment stops increasing (or increases only very slowly as the industries producing capital goods gradually expand their capacity), income will stop rising. This will stabilize the *desired* stock of

[11] See his essay, "A Model of Cyclical Growth," in Erik Lundberg, ed., *The Business Cycle in the Post-War World*, 1955, pp. 203–221.

capital, except as it is further increased through innovations. The *actual* stock of capital keeps on increasing, however, because of the large amount of investment currently being made. Eventually, the actual stock of capital catches up with that which businessmen want to have; investment then declines; this pushes down income, which lowers the desired stock of capital; and a cumulative contraction gets under way.

The contraction ends for several reasons, some of which Goodwin does not formally incorporate into his model. A number of factors put a floor under the drop in income; at the same time, innovations build up a backlog of demand for new capital goods; some types of capital goods may have to be replaced even though much of industry may still be suffering from excess capacity; the liquidation of inventories may come to an end fairly soon; and there may be an autonomous increase in government spending. Hence, eventually, there will be an increase in investment, causing some rise in income; the desired capital will rise because of the rise in both βY and $\phi(t)$ (the latter being innovational investment); and a new upswing will get under way.

SOME NONACCELERATOR MODELS[12]

James Duesenberry has presented a fairly complicated, disaggregated model involving a considerable number of equations and variables; but much of his theoretical analysis is carried out in terms of a fairly simple model, which runs as follows.[13]

As in other models, steady growth depends on the proper balance between growth in income and growth of the capital stock. Income depends on consumption and investment. Investment depends on the level of income and the stock of capital; the higher is income relative to the capital stock, the more profitable is new investment. (Note that this is an economic relationship, not a technological one as in the

[12] Goodwin's model might have been included in this section, since current investment does not depend simply on the accelerator. But the accelerator, together with technological change, does determine the desired stock of capital and thus plays an important role in his model.

[13] See his *Business Cycles and Economic Growth,* 1958, chaps. 9–10. For a rather similar model, see Arthur Smithies, "Economic Fluctuations and Growth," *Econometrica,* vol. 25, January, 1957, pp. 1–52. Smithies attempts to explain both growth and cycles by the endogenous working of his model, whereas Duesenberry, as noted later, relies heavily on exogenous disturbances to explain why the growth path determined by the model is interrupted by depressions.

case of the accelerator. As a result, the capital-output ratio is given by the working of the model; it is not a datum given in advance.) But investment also increases the capital stock. Thus, assuming that the relationships themselves do not change, steady growth requires that investment increase both income and the capital stock at the same rate. What this rate is (and also what ratio of capital to income is consistent with this rate of growth) depends on the assumed relationships in the model, and these could be such that no equilibrium rate of growth is possible. Duesenberry believes that the American economy has required the help of autonomous forces not included in the simple version of his model—particularly population growth and technological change—to achieve the growth rate that has been experienced.

Unlike Hicks, Goodwin, and most other model builders, Duesenberry does not use the endogenous working of his model to explain business cycles. His model is stable; it does not push income rapidly upward until it hits a ceiling and then rebounds sharply, as does Hicks'; and it does not generate contractions so severe that even replacement expenditures fall to zero.[14] Duesenberry relies on a variety of disturbances to explain the booms and depressions of real life —for example, fluctuations in autonomous investment, speculation, monetary disturbances, and so on.[15]

Dynamic models that make investment depend rigidly on the acceleration principle tend to be highly unstable if we assume a realistic value for the accelerator. They also leave out other variables that are obviously important. Duesenberry's model, in a sense, represents a reaction against these limitations. Two other attempts to get away from rigid accelerator models of the Harrod-Domar or Hicks type are represented by the recent work of Nicholas Kaldor and Robert Solow.[16] Their models are quite different in form and have little in common except their concern with the factors determining equilibrium growth. (Neither is concerned directly with the cycle.) While

[14] Ceilings may help to explain short, minor cycles, but not major cycles. *Ibid.*, p. 280.

[15] *Ibid.*, chaps. 11–12. He suggests that the American economy was more stable in the nineteenth century than it was in the 1920's.

[16] Nicholas Kaldor, "A Model of Economic Growth," *Economic Journal*, vol. 67, December, 1957, pp. 591–624; Robert Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, vol. 70, February, 1956, pp. 65–94. For another example, see Tobin, *op. cit.*

extended discussion of these models would get us into more technical detail than is appropriate for this book, each has certain features worth mentioning, because they point up some of the current dissatisfaction with pure multiplier-accelerator models as representations of the main influences operating on long-term growth.[17]

Kaldor, for example, gives profits—both the profit rate on capital and the share of profits in total income—an important role in the determination of the level of investment, and he brings in technological change as well as investment to explain the rate of growth of output. One result is that the capital-output ratio is not a technologically determined datum, as it is in strict accelerator models; it tends toward an equilibrium value determined by the relationships included in the model.[18]

The same is true of Solow's analysis, which goes back to the classical assumption that there exists a "production function" for the economy as a whole which makes the level of output depend on various possible combinations of labor and capital; i.e.,

$$Y = f(L,K).$$

Labor and capital can be combined in various proportions, depending on technological conditions and the relative prices of labor and capital. Solow attempts to show that the possibility of substituting capital for labor (and vice versa) removes the instability inherent in the Harrod-Domar-Hicks type of accelerator model. Here again, the capital-output ratio is not something to be taken as given but as something that results from the operation of market forces as growth occurs.

SOME CRITICAL COMMENTS

While these models and others like them give us some additional understanding of the forces operating to create both growth and fluctuations, none should be taken too seriously as an explanation of what happens during the business cycles of reality. They are probably more useful for an appreciation of some of the important influences operating on the rate of long-term growth (although they have

[17] Robert Eisner has attempted to defend the Harrod-Domar-Hicks type of model against these criticisms in "On Growth Models and the Neo-Classical Resurgence," *Economic Journal*, vol. 68, December, 1958, pp. 707–721.

[18] This also is the case in Duesenberry's model.

serious limitations here also) than they are for an understanding of what causes business cycles.[19]

As explanations of the business cycle, these models are all much too simple. They all emphasize the relations among only a few variables. They deal only with aggregates—with *total* investment and *total* output—and thus ignore the various types of maladjustments that may occur in particular parts of the economic system. Little attention is paid to the disturbing effect of innovations and to the opening up and exhaustion of investment opportunities. The role of population growth tends to be neglected, even in those models that emphasize long-term growth. Most of these models pay little or no attention to monetary factors; the changing relations between costs and prices are largely ignored, and so is the role of expectations. All of these models, to the extent that they are dynamic, assume the existence of various kinds of lags—for example, between consumption and income or between a change in output and investment. But the actual length of these lags is seldom explored, nor is consideration given to the possibility of systematic changes in these lags during the cycle.[20] The relationships that are emphasized are assumed to be constant over the cycle, which is almost never the case. Most of these models are particularly implausible in their explanation of the lower turning point.

In brief, these models are too simple and too unrealistic. At the same time, they are suggestive. They throw the spotlight on certain types of interaction in the economic system that may possibly be of significance in some types of situations. Only an eclectic type of theory can explain the diverse behavior revealed by actual business cycles.[21] These models suggest some things that should be kept in mind as possibilities—particularly the fact that the volume of investment achieved during the boom may imply a rate of expansion in productive capacity that the economy cannot support without occasional interruptions. But we need more than this for a realistic theory of

[19] For one useful and stimulating critique, particularly of the Harrod-Hicks type of model, see Erik Lundberg, "The Stability of Economic Growth," in International Economic Association, *International Economic Papers,* no. 8, 1958, pp. 45–64.

[20] Cf. R. G. D. Allen, "The Structure of Macro-Economic Models," *Economic Journal,* vol. 70, March, 1960, pp. 38–56.

[21] We noted that Duesenberry winds up with an eclectic theory because he falls back on a variety of disturbances to induce fluctuations in his model.

business cycles; and, of course, we utilized much more than this in our theoretical synthesis in Chapter 11. As we shall see at the end of this chapter, these models have not yet been very helpful in empirical business-cycle research despite their attempt to deal with quantitative relationships. The main difficulty is that which we have already noted: the failure to take into account essential features of the complex world of reality.

One of the interesting aspects of these models, particularly those that put their primary emphasis on growth, is that they provide something of a link between modern, neo-Keynesian, "orthodox" theory and some of the main preoccupations of both early English classical theory and Marxian economics—and also with some of the underconsumptionist literature. The link is a common concern with the way the accumulation of capital—i.e., investment—is related to the long-run growth of output and with the conditions under which, in a private-enterprise economy involving saving and investment, growth in output is self-sustaining.[22] In short, there is a common concern with the nature of the capital-adjustment process.[23]

OVERSAVING AND UNDERCONSUMPTION THEORIES

A generation and more ago, when economists and laymen worried much more about the recurrence of severe depressions than about problems of growth and inflation, the instability inherent in capitalism was frequently attributed to "underconsumption." So-called underconsumption theories of the business cycle have a long history, going back at least to Malthus and other early nineteenth-century writers.[24] Most of them have not been stated in very rigorous fashion, and some have been the creations of amateur economists in search of economic panaceas.

These underconsumption theories can be divided into two broad groups. One argues that the investment of saving in new productive facilities causes production periodically to outrun consumers' purchasing power. Several variants of this approach have sufficient merit to warrant our paying some attention to them here. The second type

[22] Thus Joan Robinson calls her study of the theory of growth *The Accumulation of Capital* (1956).

[23] Cf. Hamberg, *op. cit.*, chap. 1; W. J. Baumol, *Economic Dynamics,* 2nd ed., 1959, part I; E. D. Domar, *Essays in the Theory of Economic Growth,* 1957, pp. 109–128.

[24] Cf. A. H. Hansen, *Business Cycles and National Income,* 1951, chap. 14.

of theory is advanced by those "naive" underconsumptionists who believe that, because of monetary and other defects, the economic system is regularly unable to generate sufficient purchasing power to absorb the current flow of goods and services at profitable prices. This latter approach is almost completely without merit, and we shall therefore concentrate our attention on the first group of theories.

The idea that saving leads to overproduction and thus to business depressions has been expounded for more than a century by a number of writers. These savings, it is argued, are invested in new productive facilities, and the additional capacity eventually turns out more goods than people can buy at profitable prices with their existing purchasing power. Thus depressions are due to too much saving or —if we look at it the other way around—to too little consumption. Hence, we may refer to this type of explanation as either an "oversaving" or an "underconsumption" theory. It may also be considered a particular kind of overinvestment theory, since the saving that supposedly causes the trouble does so because it is invested in excessive productive capacity. It can be seen that there is at least a distant relationship between this kind of explanation and that provided by the models discussed in the preceding section. Both emphasize the relations among saving, investment, consumption, and income.

The best-known version of this theory among American writers is that of W. T. Foster and Waddill Catchings, who did much to dramatize what they called "the dilemma of thrift."[25] Without saving, we cannot have an expanding economy. Yet this same saving, once it is productively invested, insures that the supply of consumers' goods will eventually outrun consumers' demand. The economy provides no way whereby consumers may obtain the additional purchasing power needed to buy the larger volume of goods at constant prices.

The trouble arises because corporate and individual saving, if invested in plant and equipment, is "used twice in production but only once in consumption." If income is saved and invested in new facilities, demand is equal to supply only during the initial period while the new capital goods are being made and installed. During this period, the workers producing the capital goods consume what the savers refrained from consuming. As soon as the new facilities are completed, however, the economy runs into trouble. Production is

[25] Their views were set forth in a series of volumes published during the 1920's. See especially *Profits*, 1925, part 5.

larger because of the additional capacity, but there is not a corresponding increase in consumers' money incomes. Either unsold goods must pile up or prices must fall. In either case, producers are discouraged and a business decline begins.

Foster and Catchings' primary thesis does not stand up under searching criticism, and their main conclusions have not been accepted by professional economists.[26] The fact of the matter is that there is no reason why new capacity, financed out of past saving, cannot generate sufficient income to purchase the additional output. The new production involves costs; if these are paid out, purchasing power exists to buy the output at a price equal to cost. They do have one valid point, however, which neither they nor their critics fully considered. The cost directly chargeable to the new capacity is covered by the charge made for depreciation. Until replacement is necessary, this does not result in cost payments that are returned to the income stream. Only in this sense is it true that an automatic deficiency in purchasing power may arise through the process they describe.

Whereas Foster and Catchings argue that *any* saving leads eventually to overproduction, Hobson believes that there is an optimum rate of saving that, if not exceeded, would permit steady expansion.[27] Booms and depressions occur because the unequal distribution of income permits savings to exceed this ideal rate. During periods of prosperity, the rich save a large part of their incomes; these savings are invested; and, as a result, output rises too fast to be absorbed by the incomes available for consumption. The excessive savings increase output and also hold back consumption. During depressions, production and incomes fall so low that the excessive savings are eliminated. Eventually consumption exceeds production, and this stimulates businessmen to expand again.

Hobson's analysis is not at a high theoretical level, and it cannot be said that his argument is convincing. However, his discussion, while anything but precise, bears a marked kinship to the recent, more rigorous analysis—by Harrod, Domar, and the other model

[26] A number of well-known economists have published criticisms of Foster and Catchings. For the detailed references, see Gottfried Haberler, *Prosperity and Depression,* 4th ed., 1958, p. 125.

[27] J. A. Hobson, *The Economics of Unemployment,* rev. ed., 1931, and *The Industrial System,* 1909. For a recent summary and evaluation, see E. E. Nemmers, *Hobson and Underconsumption,* 1956.

builders discussed earlier—of the conditions necessary for an equilibrium rate of growth. Indeed, by taking a few liberties, one can translate Hobson's discussion into a modern-looking growth model.[28]

Another version of the underconsumption theory emphasizes the relative behavior of wages and profits as the cause of excess savings and overproduction.[29] During prosperity, wages lag behind prices, and profits rise much more rapidly than do payrolls. The great increase in profits stimulates new investment, and thus the economy increases its capacity to produce consumers' goods. But the purchasing power of consumers, who are primarily wage earners, does not rise in proportion. Hence prices must eventually fall, and their fall ushers in a depression. This type of explanation has frequently been advanced to account for the boom of the 1920's and some of the main features of the depression that followed.

The socialist literature since Marx does not add much to the various versions of the underconsumption approach already discussed.[30] Marx himself did not present a fully developed theory of business cycles. His references to the subject in *Capital* point more often to other reasons for depressions than to oversaving and overinvestment in relation to consumers' demand. The body of Marxian literature since then has not made underconsumption the central point in its analysis of economic crises.[31] When the underconsumption argument is stressed, it is along lines similar to those already discussed, but done in terms of Marxian definitions and concepts. Capitalists invest "surplus value" in new capacity, causing production to expand beyond the ability of consumers to buy the enlarged output. Capitalism, through the phenomenon of surplus value, thus breeds a rate of "accumulation" that cannot be supported indefinitely. This conclusion is related to a notion we have already encountered among recent non-Marxian writers—that there is an equilibrium rate of growth

[28] See, for example, E. D. Domar's comments in his *Essays in the Theory of Economic Growth*, pp. 103–104, and D. J. Coppock, "A Reconsideration of Hobson's Theory of Unemployment," *The Manchester School*, vol. 21, January, 1953, pp. 1–21.

[29] See, for example, Paul H. Douglas, *Controlling Depressions*, 1935, esp. chap. 2; also his contribution in *Economic Essays in Honor of Wesley Clair Mitchell*, 1935, pp. 105–130. For references to other writers, see Haberler, *op. cit.*, pp. 137–141.

[30] See the useful summary of Marxian cycle theory in P. M. Sweezy, *The Theory of Capitalist Development*, 1942, part 3, and Henry Smith, "Marx and the Trade Cycle," *Review of Economic Studies*, vol. 4, June, 1937, pp. 192–204.

[31] Cf. Sweezy, *op. cit.*, chap. 10, esp. pp. 178–179.

and of investment that would not cause trouble. In the Marxian system, however, the rate of accumulation that inevitably occurs under capitalism is incompatible with the achievement of equilibrium. The rate of profit tends to decline and capitalist crises become steadily more severe.

Although none of the oversaving theories thus far considered is free of serious analytical defects, these theories do point to an important truth. Capacity can become excessive in consumers'-goods industries. If this happens, investment will tend to decline in these lines. It is highly unlikely that capital can always be substituted for labor and that, if the interest rate is low enough, investment can be maintained even in the face of excess capacity. It is entirely possible for there to be a lack of balance between the way consumers divide their incomes between spending and saving and the proportions in which new investment is distributed between the capital-goods and the consumers'-goods industries.[32] Production in the latter *can* rise faster than consumers will (or can) increase their spending. To say this, however, is not to say that underconsumption in a broad overall sense necessarily occurs during every upswing or that the main difficulty may not be usually in other directions. Ordinarily, the initial trouble is not likely to be a lack of balance between *total* consumption and *total* capacity in the consumers'-goods industries. The maladjustments are more likely to be partial in character, affecting initially particular industries or groups of industries. In this connection it may be remarked that virtually all of these underconsumption arguments ignore the role played by innovations—both in initiating spurts in investment and in bringing on cyclical downturns. The temporary saturation of investment opportunities in particular lines, as the economy adjusts itself to innovations, can lead to a decline in total investment even though total consumption is continuing to rise at a constant rate. And innovations can create a demand for new types of productive capacity, even if investment in old lines is excessive.

So far we have been dealing with the "oversaving" type of underconsumption theory, which does have some degree of plausibility and is buttressed by reasonably logical arguments. In addition, we

[32] For a useful analysis of the ways in which "underconsumption" may arise, see H. Neisser, "General Overproduction: A Study of Say's Law of Markets," reprinted in *Readings in Business Cycle Theory*, pp. 385–404.

should at least mention the voluminous body of writings that have been spawned by the "naive" underconsumptionists. These writers belong for the most part to what we may call, unkindly but not inaccurately, the "crackpot fringe" of monetary reformers—who would cure all of the world's economic ills with monetary panaceas and pension schemes of one sort or another. As is to be expected, this type of literature itself goes through cycles: its volume and popular appeal rise during depressions and wane during periods of prosperity.[33]

AGRICULTURE, THE WEATHER, AND BUSINESS CYCLES

Even in a country as industrialized as the United States, the behavior of agricultural production, prices, and incomes has important effects on other types of economic activity. In many countries, agriculture dominates the domestic economy; and the role played by agriculture in the United States 50 or 100 years ago was, of course, much more important than it is now.

In view of this, it is not surprising that a number of writers have looked into the relations between agriculture and the rest of the economy over the course of the cycle. That cyclical fluctuations in industrial output and nonagricultural incomes affect farm prices and farm incomes is obvious. It is also clear that, once a downswing or upswing begins in the industrial sector, the changes induced in the agricultural variables react in turn on industrial behavior. But can more than this be said?

The possible initiating role of agriculture in creating business cycles has been investigated along two lines, both with unsatisfactory results. One hypothesis, which has never attracted many adherents, is that there is a connection between meteorological conditions (for example, cycles in sunspots) and business activity, the connecting link being changes in agricultural production.[34] Other writers have

[33] The "Social Credit" writings of Major C. H. Douglas and his followers are a good example of this sort of naïve and extreme underconsumptionist position. See, for example, C. H. Douglas, *Social Credit,* rev. ed., 1933, and P. Mairet, ed., *The Douglas Manual,* 1934. For further references to the writings of Douglas, his followers, and his critics, see A. H. Hansen, *Full Recovery or Stagnation?,* 1938, chap. 4.

[34] For a useful summary of the early work in this field—with which the names of W. S. Jevons and H. L. Moore are particularly associated—see C. Garcia-Mata and F. I. Shaffner, "Solar and Economic Relationships: A Preliminary Report," *Quarterly Journal of Economics,* vol. 49, November, 1934, pp. 1–51. See also V. P.

investigated in some detail the effect of changes in agricultural output on business activity but have not sought to establish an independent cycle in agricultural production itself. The possible interrelations between changes in farm production and in business activity are numerous and complex, and the results one obtains depend in part on the particular assumptions made. This is a complicated subject, and we cannot go into it in any detail. We shall limit ourselves to the following cursory comments.[35]

It is unlikely that regular cycles in crop production play an important role in business fluctuations. Quoting J. M. Clark, we may add: "This does not mean that agriculture has no effect on the business cycle. . . . It simply means that agriculture is not a regularly acting force, tending typically and regularly to help initiate the recovery, or stimulate the revival, or in any other way to play habitually the same role in at least a predominant number of cycles."[36] This is almost certainly correct for the United States since World War I. Agriculture may have played a more important role than this implies during the nineteenth century, particularly when farm products bulked much larger in American exports than they do now and when agriculture accounted for a much larger share of total economic activity than is now the case.

Insofar as moderate changes in crop production do affect business, the consensus seems to be that an increase in agricultural production usually is a stimulating influence and that a decrease has a depressing effect. But this is not the only possibility. We may get partial overproduction and partial overinvestment in agriculture as well as in industry. We know that agriculture can suffer from "overproduction," and drastic declines in farm prices resulting from a large increase in production can touch off a highly deflationary train of consequences: a cessation of farm investment, a collapse of farm real-estate prices, mortgage foreclosures, bank insolvencies, a general rise in liquidity preference, and so on. As a matter of fact, it has probably

---

Timoshenko, *The Role of Agricultural Fluctuations in the Business Cycle,* 1930. Timoshenko sought to show that there were cycles in agricultural production and that these were related to business cycles, but he did not try to explain why these cycles in farm output occur.

[35] For a more detailed theoretical analysis of the relations between agriculture and the rest of the economy during the cycle, and for references to the more important literature, see Haberler, *op. cit.,* chap. 7.

[36] *Strategic Factors in Business Cycles,* 1935, p. 62.

been through these monetary and financial channels that agriculture has exerted a good part of its effect on business conditions in the United States.

Probably the clearest case in which agriculture plays an important initiating role is that of a primarily agricultural country that exports a large part of its output. In this case, with the whole world to sell to, the country faces a relatively elastic demand for its exports. Hence, an increased volume of production and exports means larger money receipts—i.e., an increase in the value of exports relative to imports. The result is clearly stimulating. The effect will be particularly marked if good crops in the exporting country occur at the same time that crops are poor in the rest of the world.

So far as the United States is concerned, agriculture's impact on business conditions today does not come primarily from weather-induced changes in crop yields. Here, as in so many other areas of economic activity, the government has become the most important exogenous influence at work. In particular, government farm-price supports and farm subsidies affect agricultural incomes, and these changes in incomes affect business activity in the same way as do changes in the incomes of other groups in the economy. At the same time, the support given to farm prices has indirect effects on the way consumers' spending is distributed between farm and nonfarm products.

One other aspect of cycles in agricultural production should be mentioned. A number of individual farm products show cycles of their own, both in production and in prices, because supply responds to price only with a lag. Suppose that low prices result from a large output in a given year. These lead to smaller supplies in subsequent years, with the result that prices rise; higher prices now lead to increased production, and this is turn causes prices to fall; and thus a new cycle begins. The most famous example of this sort of lagged interaction between supply and price is the corn-hog cycle.

Several points should be made about these special commodity cycles. They are essentially endogenous in nature, resulting from the way supply and price interact. When external influences, such as changes in the weather, intervene, these tend to set off new cycles of the sort described. Secondly, these cycles result from the fact that supply reacts to price only after a lag, because of the fixed production period that exists for various types of agricultural commodities. The

theory underlying such production-price cycles has been worked out by a number of writers and is usually referred to as the "cobweb theorem."[37]

APPENDIX: ECONOMETRIC MODEL-BUILDING

We saw earlier in this chapter that an increasing number of economists have presented their theories explicitly in the form of business-cycle models. A few economic variables are chosen as being of primary importance, and it is then shown how fluctuations are generated by the relationships that are assumed to hold among the variables. These models can be expressed in mathematical or literary form. Rigorous demonstration that cyclical variations result from the assumed relationships virtually requires that the latter be stated in mathematical terms.

During the last 25 years or so, important work has been done in the construction of *econometric* models. Econometrics is a special type of economic analysis in which economic theory (formulated in mathematical terms) is combined with statistical measurement of economic phenomena. [38] Econometric business-cycle research consists of the following steps: A system of equations (i.e., a model) is set up to represent a particular theory of how certain variables interact with each other to determine the behavior of the economic system through successive periods of time. By statistical methods these equations are then "fitted" to the actual statistical data to determine the numerical values of the constants in the equations chosen. The model is then tested to see if the results adequately explain what has actually happened. If the equations fit the data well, and the sampling errors involved are small, the investigator concludes that the theoretical relationships embodied in his equations are consistent with the actual behavior of the economy, at least for the period studied. This does not mean that this particular theory has been "proved" to be correct. For there may be other theories and other sets of equations that will fit the data equally well.[39]

[37] See Mordecai Ezekiel, "The Cobweb Theorem," reprinted in *Readings in Business Cycle Theory*, pp. 422–442; Jan Tinbergen and J. J. Polak, *The Dynamics of Business Cycles*, 1950, chap. 14.

[38] Cf. W. Leontief, "Econometrics," in H. S. Ellis, ed., *A Survey of Contemporary Economics*, 1948, p. 388 n. Leontief's paper provides a useful brief survey of developments in this field of economic analysis. For a more extended but still elementary exposition of econometric methods, see J. Tinbergen, *Econometrics*, 1951. At a more technical level, see L. R. Klein, *A Textbook of Econometrics*, 1953.

[39] For more detailed discussion of the nature of the econometric approach to business-cycle analysis, see T. C. Koopmans, "The Econometric Approach to Busi-

The final test of an econometric model lies in its ability to predict. These models are dynamic in the sense that some of the equations connect variables referring to different moments of time. If a model is to hold for the future as well as for the past, we should be able to substitute into the equations past and present values for some of the variables, and the model should then predict for us what the values of some of the other variables will be next month or next year. So far, no econometric model has met this forecasting test with any marked degree of success.

The following are the main features of econometric business-cycle models.[40]

1. The variables involved are broad aggregates, such as total consumption, total investment, total profits, and so on.

2. The models are considered to be complete, but only in a formal, logical or mathematical, sense; that is, there are as many equations as there are unknowns (i.e., variables whose behavior is to be explained).

3. The models are dynamic in that some of the equations connect variables referring to different time periods. Some of the variables may also be expressed as rates of change.

4. Four kinds of equations may be used: (a) *identities* or definitional equations, which are true by definition (thus: price times quantity equals value, saving equals investment [*ex post*], and so on); (b) *institutional rules,* such as equations that describe the reserve requirements of the banking system or the relation between corporate income taxes and corporate incomes; (c) *technological transformation functions,* such as the way output varies with employment; (d) *behavior equations,* representing the way groups of firms or individuals react to given stimuli (for example, the consumption function, showing how consumers respond to changes in income, and demand equations, which relate the quantity bought of a commodity to its price, to the prices of other commodities, to income, and to perhaps other variables).

---

ness Fluctuations," *American Economic Review,* vol. 39, May, 1949, suppt., pp. 64–72; Jan Tinbergen, "Econometric Business Cycle Research," reprinted in *Readings in Business Cycle Theory,* pp. 61–86; L. R. Klein, *Economic Fluctuations in the United States, 1921–1941,* 1950; and, by the same author, "Statistical Testing of Business Cycle Theory: the Econometric Method," in Erik Lundberg, ed., *The Business Cycle in the Post-War World,* pp. 222–245.

[40] This summary is based on that of T. C. Koopmans, *op. cit.,* p. 64. Cf. also C. F. Christ, "Aggregate Economic Models," *American Economic Review,* vol. 46, June, 1956, pp. 385–388.

All econometric models contain both endogenous and exogenous variables. The former are the ones whose behavior is to be explained by the equations chosen. The exogenous variables are taken as given. They represent forces at work which the model does not try to explain. One measure of the usefulness of a model lies in its treatment of exogenous variables. If important economic magnitudes whose behavior we want to explain are treated as exogenous, the model is of limited usefulness. Thus a model that treats investment as exogenous says in effect: Consumption, employment, and the other endogenous variables will change in certain ways as investment varies (in accordance with the equations used), but we cannot say why or when investment changes. To solve for the endogenous variables, we must wait to find out what investment will be.[41]

Some variables must, by the nature of the case, be considered exogenous—in particular, those that are determined by government policy; for example, the supply of money and government spending. Population or the labor force is another type of exogenous variable that, for obvious reasons, we cannot explain entirely in terms of the behavior of a few other economic magnitudes.

A simple example may help to make these concepts clearer. A simplified version of a model that has actually been fitted to American data for the interwar period runs as follows:[42]

$$
\begin{aligned}
&(1) \quad C = a_0 + a_1 W + a_2 P + u_1 \\
&(2) \quad I = b_0 + b_1 P + b_2 P_{-1} + b_3 K_{-1} + u_2 \\
&(3) \quad W = c_0 + c_1 Y + c_2 Y_{-1} + c_3 t + u_3 \\
&(4) \quad C + I + G = Y \\
&(5) \quad P + W = Y \\
&(6) \quad \Delta K = I
\end{aligned}
$$

The a's, b's, and c's in these equations are the constants for which we have to determine numerical values. The capital letters represent the variables, which are tied together in the way described by the equations. All the variables are expressed in "real" terms; i.e., they

[41] At least one recent model of the United States economy displays some interesting, though limited, possibilities, despite the fact that all noninventory investment is taken to be exogenous. See the reference to the model by Duesenberry, Eckstein, and Fromm, p. 393, below.

[42] This model is taken with permission from L. R. Klein, *Economic Fluctuations in the United States, 1921–1941,* Cowles Commission for Research in Economics Monograph No. 11, p. 62, with no change other than the substitution of Latin for Greek letters.

are corrected for price changes. Let us now look at the six equations one by one.

(1) The variables in the first equation are total consumers' expenditures (C), total payrolls (W), and total nonwage income, which for short we can refer to as profits (P). The equation states that consumption varies with changes in total wages and profits; in other words, this is the consumption function. The coefficient a_1 tells us how much consumption will change for every change of a dollar in payrolls, and similarly for a_2. Since we do not expect this equation to hold perfectly for every year, we add the variable u_1 to represent the "disturbance"—that is, the amount by which unmeasured influences cause C to deviate from the calculated value.

(2) The second equation is the investment function. Investment (I) is made to depend on current profits (P), last year's profits (P_{-1}), and the stock of capital at the end of last year (K_{-1}). In all cases, the subscripts give the date for which the variable in question is to be taken. Again we insert a residual variable (u_2) to express the fact that this equation does not hold perfectly.

(3) This equation expresses the demand for labor. The total wage bill (in constant dollars) is made to depend on current output (Y) and last year's output (Y_{-1}). It is also assumed to vary in accordance with a trend factor c_3t; that is, independently of Y and Y_{-1}, we expect W to change by a constant amount each year in response to the growing bargaining strength and productivity of labor. The variable t represents time (in years) measured from some base period. Here also we add the residual variable u_3.

The remaining equations are identities to complete the system.

(4) We are already familiar with this equation, which states that consumption plus investment plus government spending is equal to total output (Y). Since, by definition, this equation must hold perfectly, we do not have to add another u variable to cover other influences at work.

(5) This is the other side of the income equation and is also true by definition. Total income or output (Y) goes to either wages or profits (which are defined to include all nonwage incomes).

(6) This is also a definition. The amount of net investment in any year (I) is equal to the net change in the stock of capital during the year (ΔK).

All of the variables are defined in "real" terms; that is, they are

corrected for price changes. Thus Y is a measure of the real national income, which is equivalent to total output. We are dealing with the net national income, after the deduction of depreciation and business taxes.

Of our six equations, the first three are behavior equations and the last three are identities. In all we have six equations to explain six endogenous variables (C, I, W, P, K, and Y). In addition we have the exogenous variables t and G. We do not attempt to calculate G; we look up the actual data to find out what value for G to use in any particular year.

Having these equations, the econometrician would next apply the appropriate statistical methods in order to ascertain the numerical values of the a's, b's, and c's—that is, the constants in the equations. Since this is not a textbook in statistics or econometrics, this is a good point at which to leave him.[43]

It is clear that this is much too simple a model for useful results. The aggregates are too broad, and obviously important variables are omitted. In particular, we need to break down investment and set up separate equations for inventories, business equipment, industrial and commercial construction, and residential construction. There is no consideration of government taxes; there are no variables for the money supply, holdings of liquid assets, or interest rates. In addition, consumption is related to national income rather than to disposable income; hence, the influence of changes in business savings and personal taxes is concealed.[44]

The author of this model, Lawrence Klein, has also developed much more elaborate models which he has applied to American data. His most recent published model, for the years 1929–1952, contains twenty equations, an equal number of endogenous variables, and an additional number of exogenous variables.[45] Among the

[43] For a technical discussion of estimating methods, see the textbook by Klein previously cited. For two excellent nontechnical discussions, see Christ, *op. cit.*, pp. 397–401, and K. A. Fox, "Econometric Models of the United States," *Journal of Political Economy*, vol. 64, April, 1956, pp. 128–142.

[44] In applying this model to the United States, Klein elaborated it to the extent of dividing W into two variables, W_1 (wages paid by private enterprise) and W_2 (wages paid by government), and adding business taxes as an additional variable. *Economic Fluctuations in the United States, 1921–1941*, pp. 64–66.

[45] L. R. Klein and A. S. Goldberger, *An Econometric Model of the United States, 1929–1952*, 1955. For valuable reviews of this study, see the articles by Christ and Fox previously cited.

20 endogenous variables for which equations are given are, for example, consumption, gross investment, corporate saving, corporate profits, total compensation separately for private and for public employees, gross national product, the wage level, farm income, personal and business liquid assets, and short- and long-term interest rates. Although this may seem to be a very complicated model, it still is too aggregative, particularly in its failure to break down total investment into its significant components.

Sometimes an econometric model may give us useful partial insights even though it treats as exogenous (i.e., does not try to explain the behavior of) some variables that we know to be important sources of cyclical instability. For example, a recent model of the American economy treats total noninventory investment, as well as government spending, as exogenous.[46] Since the behavior of fixed investment is not explained, this cannot be considered a complete business-cycle model. But it serves a useful purpose in describing how the economy responds to possible changes in either private investment or government spending. This is done in this model by a set of equations which relate, respectively, consumers' expenditures to disposable income, disposable income to personal income, and personal income to GNP.[47] An equation for inventory investment is also included. In effect, this is a model to test the stability of the economy's responses to an assumed change in either private (noninventory) investment or government expenditures. What is measured in particular is the strength of the automatic stabilizers, given the relation of consumers' expenditures to disposable income and given the way inventory investment typically responds to changes in conditions during the business cycle. This particular model suggests that, in the absence of a collapse of private long-term investment, the American economy today is quite stable and that the multiplier

[46] James Duesenberry, Otto Eckstein, and Gary Fromm, "A Simulation of the United States Economy in Recession," *Econometrica,* vol. 28, October, 1960, pp. 749–809. The model is intended to apply only to postwar recessions.

[47] A considerable number of equations have to be combined to relate personal income to GNP—for example, equations that attempt to account for the behavior of depreciation, indirect taxes, corporate taxes and retained earnings, and various kinds of transfer payments. As we saw in Chapter 3, these are the items that have to be subtracted or added in the national income accounts as we move from GNP to personal income payments. It should also be noted that the authors intend this to be a model appropriate for recessions only, since some of the equations were fitted only for recession periods.

effects of a moderate deflationary decline in government expenditures or fixed investment are relatively small.[48]

Econometric models are less than 30 years old. The possibilities of constructing and using them to explain the business cycle were first pointed out by Ragnar Frisch, a Norwegian mathematical economist, in 1933.[49] While a number of economists and statisticians have worked with such models since then, econometric business-cycle research in recent years has been associated particularly with the names of Jan Tinbergen of the Netherlands and Lawrence Klein of the United States. After a good deal of earlier pioneering work, Tinbergen published in 1939, under the auspices of the League of Nations, two volumes which have become classics in the field of econometric business-cycle research.[50] The second of these volumes represented an econometric study of business fluctuations in the United States during the period 1919–1932.

Although the logic underlying econometric model-building remains largely as Tinbergen expressed it, a great deal of work has been done since in improving the statistical methods that are used to estimate the coefficients in the equations comprising a business-cycle model. Considerable progress has been made in working out methods for determining simultaneously the coefficients of all the equations in a model and in relating the estimating problems that arise to recent developments in the theory of statistical inference.[51]

[48] It is interesting to note, incidentally, that the Klein-Goldberger model, described on p. 392, also turns out to be very stable, even with the treatment of private investment as an endogenous variable. That is, the sort of economic system described by this model would not go on generating cyclical fluctuations without recurring disturbances or shocks. Irma and Frank L. Adelman, "The Dynamic Properties of the Klein-Goldberger Model," *Econometrica,* vol. 27, October, 1959, pp. 596–625.

[49] "Propagation Problems and Impulse Problems in Dynamic Economics," in *Economic Essays in Honour of Gustav Cassel,* 1933, pp. 171–205. Pioneering work in this area had already been done by Tinbergen.

[50] *Statistical Testing of Business-Cycle Theories: I. A Method and Its Application to Investment Activity* and *II. Business Cycles in the United States of America, 1919–1932.* See also his *Business Cycles in the United Kingdom, 1870–1914,* 2nd ed., 1956.

[51] The pioneering work here was done by the Cowles Commission (now Foundation) for Research in Economics. See especially T. C. Koopmans, ed., *Statistical Inference in Dynamic Economic Models,* 1950; and W. C. Hood and T. C. Koopmans, eds., *Studies in Econometric Method,* 1953. For a good introductory discussion, see E. G. Bennion, "The Cowles Commission's 'Simultaneous Equation Approach,'" *Review of Economics and Statistics,* vol. 34, February, 1952, pp. 49–56.

At the same time, various economists have experimented with actual models applied to specific periods of business-cycle history, not only in the United States but also in some other countries.

Econometric business-cycle models have not, so far, materially increased our ability to explain or predict business fluctuations. This is not surprising, because the difficulties that have to be overcome are well-nigh insuperable, even after the statistical and computing problems are solved.

The basic difficulty is this: An econometric model purports to explain how the economy behaves during a particular period. The equations used, which are selected on the basis of initial theorizing by the investigator, represent a set of hypotheses as to how the economy functions. After the coefficients are computed for these equations, we may find that the model yields accurate estimates for all the important endogenous variables during the period covered. But we still do not know whether we have explained anything. The endogenous variables may also depend on other factors which are not covered in the model, but these additional influences may not have been at work during this period. We have here the same problem that always arises when we apply the methods of multiple correlation analysis to economic time series. The high correlation coefficients that may be obtained do not prove that the independent variables explain or "cause" the behavior of the dependent variable. They may both depend on some other variable which has not been included. Or the correlation may hold only because certain conditions, the nature of which has not been investigated, are true; and these conditions may not continue to hold in the future.

This last point raises the problem of extrapolation and prediction. At best, a model can "explain" only some past period. To apply it to a later period we must assume that the "economic universe" has not changed, that the variables in the model will be related to each other in the future in precisely the same way as in the past. This is always a dangerous assumption to make, and it almost always turns out to be a false one. The danger is the greater, the fewer the variables in our model. The more limited the model, the more likely is it that we have left out something that, though not important in the past, may become important in the future. Even if we have not omitted any important variables, the significant relationships are almost certain to change unpredictably in one way or

another. All sorts of institutional changes may alter the basic relationships that the equations in our model attempt to describe. Consumption may seem to depend only on disposable income during the period covered, and we may find that later it also responds to changes in liquid-asset holdings. During one period, private long-term investment may seem to be highly sensitive to current changes in profits and relatively insensitive to the prices of capital goods or to increases in the stock of capital; at a later period, these additional variables may take on a new importance. These issues are particularly significant if we project our model very far past the period originally covered, or if major structural changes suddenly occur—say, because of a war. Even very short-period projections may give inaccurate predictions. The constants in our equations represent, in a sense, average behavior over the period studied. These averages conceal a good deal of short-period variability.[52] Thus we come back to the point made before. It is humanly impossible to put into a system of equations every single influence that has affected or may affect the variables we want to study. Some influences, indeed, cannot even be expressed in quantitative form.[53]

Despite such difficulties, work in this field is likely to continue unabated. Anything approaching complete success (in the sense of ability to predict) is highly improbable, but the rewards of even partial success make the attempt well worth while. After all, it is only in this direction that economists can aspire to put their field on the same "scientific" plane as that of the natural sciences. Econometric model-building represents a striving for the goal Schumpeter described when he said that "the highest ambition an economist can entertain who believes in the scientific character of economics would be fulfilled as soon as he succeeded in constructing a simple model displaying all the essential features of the economic process by means of a reasonably small number of equations connecting

[52] Current econometric models generally assume that all relationships are linear and that they remain unchanged over the period studied. Both assumptions are obviously risky ones to make. The functions included in a model may change not only in what seems to be a haphazard way but also in some systematic manner over the course of the cycle or from one cycle to the next.

[53] The reader will recall our summary in Chapter 11 of the nature and causes of cyclical fluctuations. Note the variety of possibilities that may occur and the fact that no two cycles reflect exactly the same set of forces. See also the comments by Milton Friedman in Universities—National Bureau Committee for Economic Research, *Conference on Business Cycles*, 1951, pp. 107–114.

a reasonably small number of variables."[54] Although this goal is almost certainly an impossible one, continued striving for it will teach economists a good deal that will help them in using the blunter tools with which most of their work will have to be done.

As econometric models are improved and more detail added, they might also become useful tools of government policy, and indeed they have already been used for this purpose in some countries. Models, either for particular sectors or for the whole economy, can be constructed to point up the influence of variables under the control of the government, and the equations in the model can help government officials to estimate the effects on various sectors of the economy of deliberate changes in the variables that the government can control—for example, through a reduction in taxes or a change in unemployment benefits.[55]

So far we have been considering econometric models that attempt to deal with the economy as a whole by exploring the mutual interrelations among a workably small number of aggregative variables. We may refer to these as comprehensive models. In addition, as earlier chapters have suggested, there has been a good deal of useful econometric business-cycle research that has dealt only with particular sectors of the economy. There have been, for example, many econometric studies of the consumption function, some of which were cited in Chapter 5. There have also been a number of econometric studies either of total investment or of particular kinds of investment—for example, inventory investment, residential construction, and investment in public utilities and in manufacturing.[56]

[54] "The Decade of the Twenties," *American Economic Review,* vol. 36, May, 1946, p. 3.

[55] Cf. Jan Tinbergen, *On the Theory of Economic Policy,* 1952, and *Economic Policy: Principles and Design,* 1956. Econometric models useful for government-policy purposes do not necessarily have to be dynamic; that is, they may not involve any lagged relationships. Even though static, they can help the policy-maker to estimate the effect of an assumed change in one or more variables on certain other variables in a particular year, although the model will provide no information as to what will happen in subsequent years as a result of the internal working of the system. A study of static relationships can provide some guide to policy in a dynamic world, even if it cannot provide an explanation of why the world continues to change.

[56] A few recent examples are J. R. Meyer and Edwin Kuh, *The Investment Decision,* 1957 (on investment in manufacturing); Avram Kisselgoff and Franco Modigliani, "Private Investment in the Electric Power Industry and the Acceleration Principle," *Review of Economics and Statistics,* vol. 39, November, 1957, pp. 363–

And one will find, scattered through the literature, econometric studies of other variables that enter into business-cycle analysis, of which interest rates, various kinds of prices, the demand for consumers' durable goods, foreign trade, and corporate saving might be cited as examples.

Econometric models represent an attempt to apply theory to statistical data and thus, in a sense, supply one sort of bridge between theory and the real world. Another, less rigorous but nonetheless essential, method of bringing theory and fact together is through the blending of economic and historical analysis. In the next three chapters we shall try to use our theoretical tools to interpret the business-cycle history of the United States during the last forty years.

379; P. G. Darling, "Manufacturers' Inventory Investment, 1947–1958," *American Economic Review,* vol. 49, December, 1959, pp. 950–962; and Robert Eisner, "A Distributed Lag Investment Function," *Econometrica,* vol. 28, January, 1960, pp. 1–29 (for large nonfinancial corporations).

CHAPTER 14

THE INTERWAR PERIOD, 1919–1939

THE WAR of 1914–1918 provided a tremendous stimulus to business activity the world over. The United States shared in this expansion even prior to American entry into the conflict in April, 1917. Before that date it was the expanded demand of the European belligerents for American goods that led to a rise in aggregate demand; thereafter it was primarily the large expenditures made by the American government, both directly and through loans to its allies. The pattern was roughly similar to that we experienced in World War II: sharply expanding demand by government and induced expansion in private demand, production rising to capacity, diversion of resources to war production, inflation of bank credit to finance the excess of government expenditures above tax receipts and borrowing from current saving, a sharp rise in prices under the stimulus of the inflationary pressures thus created, and so on.

It is not our intention to describe the boom that accompanied World War I. Our concern in this chapter is with the course of business in the 20 years following, until the calamity of war again engulfed the world. The general contours of the course of economic activity in the United States during this period are summarized in Figure 39. The series presented there clearly define the cyclical pattern it is our job to describe and explain. Concretely, we shall have to consider the following questions:

1. What caused the initial postwar boom of 1919–1920 and the severe but short depression that followed in 1921?

2. How do we explain the major swing in business activity that began in 1921, reached a peak in 1929, and then fell to the catastrophically low levels of 1932–1933? We must account for the pro-

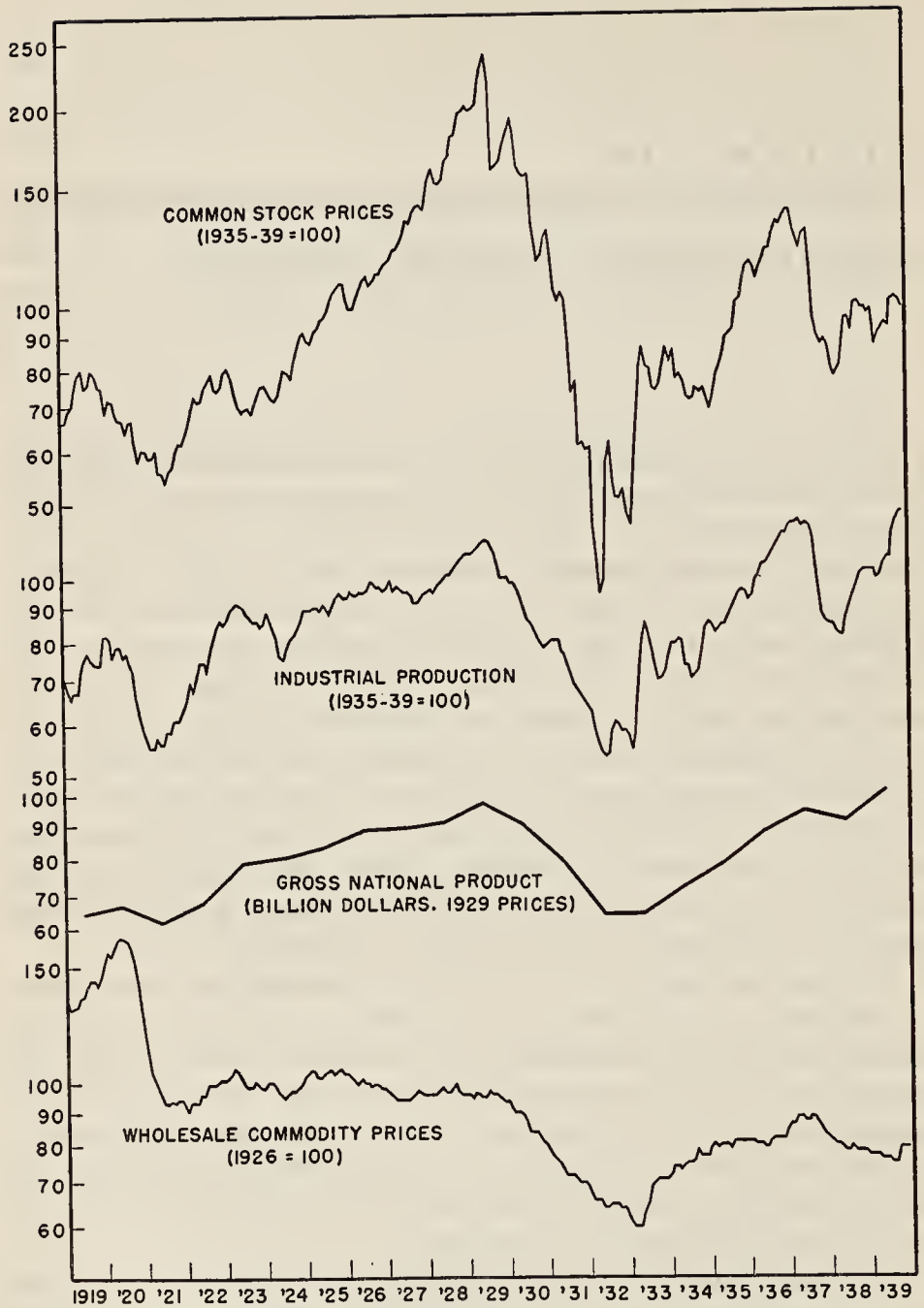


FIGURE 39. Selected Indicators of Business Activity, 1919–1939.

GNP data are Kuznets' estimates in constant prices.

longed period of prosperity during the 1920's and try to ascertain the causes of the downturn in 1929 and of the severity of the downswing which followed.

3. A less important question to be treated briefly is: why the mild recessions in 1924 and 1927?

4. Following 1933, the United States experienced a slow and halting recovery which culminated in the boom of 1936–1937 and the short but moderately severe downswing of 1937–1938. The period 1933–1938, therefore, gives us another complete business cycle to analyze.

5. We then come to the final question with which we must deal. What are the secular implications of the wide and deep valley which characterizes the 1930's on the business charts? Had the United States, after a last spectacular spurt of progress in the 1920's, entered on a period of chronic unemployment and business stagnation, marked by long and severe depressions and weak upswings—a period of Keynesian "underemployment equilibrium"? Despite the levels of output reached in World War II and the years following, this question still haunts many economists, and we shall have to consider the evidence bearing on it.[1]

POSTWAR BOOM AND DEPRESSION, 1919–1921

When World War I ended in November, 1918, the United States was just beginning to achieve "all-out" production of war goods. The economy had been through a strong cumulative expansion of four years' duration. Following the armistice, a minor recession in business occurred—not unlike that which followed the end of World War II. In both cases, resources had to be shifted from war to peacetime production; businessmen were uncertain as to future prospects; and unemployment temporarily increased.[2] But, as in 1945–1946,

[1] The most useful studies of American business cycles during the interwar period are Thomas Wilson, *Fluctuations in Income and Employment*, 3rd ed., 1948, part 2; S. H. Slichter, "The Period 1919–1936 in the United States: Its Significance for Business Cycle Theory," *Review of Economic Statistics*, vol. 19, February, 1937, pp. 1–19; and J. A. Schumpeter, *Business Cycles*, vol. 2, chaps. 14–15. See also R. A. Gordon, "Cyclical Experience in the Interwar Period: The Investment Boom of the 'Twenties," in Universities-National Bureau Committee, *Conference on Business Cycles*, 1951. For the experience of other countries, see particularly W. A. Lewis, *Economic Survey, 1919–1939*, 1949.

[2] The chief difference between the two periods lies in the sphere of prices. Price control in World War I was much less stringent and less extensive than in World

the lull was short-lived. The post-Armistice recession reached a low point in the spring of 1919, and aggregate demand then began to rise. There ensued a speculative boom and sharp reaction, which together mark off the first full cycle of the interwar period. According to the National Bureau of Economic Research, the expansion phase lasted from March, 1919, to January, 1920; the downswing that followed reached a low point in July, 1921.

The causes of the boom of 1919–1920 bear a marked likeness to those responsible for the expansion after World War II. The following were the most important factors operating early in 1919 to generate a rise in total demand:

1. A strong pent-up demand by business for capital goods for replacement and expansion, after a year and a half in which such goods had not been available.
2. Similar pent-up demands by consumers for housing, automobiles, clothing, and other items not freely available during the war.
3. The existence of a large volume of liquid assets (in deposits and government bonds) and a high level of money incomes to make these pent-up demands effective.
4. Maintenance of an abnormally high level of foreign demand for American goods.
5. Continuation of government expenditures on a large scale. Indeed, federal expenditures in the fiscal year ending June 30, 1919, were nearly 50 percent higher than in the preceding 12 months. Because of this, the monetary authorities were compelled to maintain an easy-money policy during most of 1919.

Thus, in the spring of 1919, businessmen saw that aggregate demand was being well maintained and, indeed, rising. Previous expectations that prices would decline, which had led to caution in placing orders, now gave way to optimism. There was a rush to expand output and accumulate inventories; consumers, with their accumulated savings and rising incomes, bought freely; foreign nations competed with Americans for goods, and the value of exports rose above even the wartime peak. Prices increased sharply as money

---

War II, and as a consequence the rise in prices was greater. After the 1918 armistice, businessmen generally expected that the then high level of prices could not be maintained and that prices would fall to "more normal" levels. This led to caution in buying.

demand expanded more rapidly than output. Because of limitations of capacity, labor shortages and strikes, and transportation difficulties, total output expanded only moderately from the post-Armistice low; indeed, for the year 1919 as a whole it was little if any greater than in 1918. This was a "demand-pull" type of inflationary process, with aggregate demand rising faster than aggregate supply.

The rise in prices inspired expectations of further price increases; an outstanding feature of the expansion was the extent of speculative buying, in commodities, securities, and real estate. A good deal of this speculation was done with borrowed money, and loans and deposits of commercial banks rose rapidly. Wages followed prices upward, and there were widespread complaints of labor inefficiency.

By the early months of 1920, the American economy was in a vulnerable position. Business inventories were high; firms and speculators were heavily in debt to the banks, and the banks had rediscounted heavily at the Federal Reserve Banks; consumers were showing resistance to the high levels that retail prices had reached; construction was beginning to decline because of high building costs and the unavailability of mortgage credit; there were increasing fears that Europe could not long continue to finance the abnormally high level of American exports; costs were high, with numerous production bottlenecks.

Two additional factors had also begun to exert a deflationary force. Government spending was steadily decreasing. By the fourth quarter of 1919 the budget had been balanced, and a small excess of receipts over disbursements had emerged. Thus, by the beginning of 1920, government finance had come to exercise a deflationary rather than an inflationary force on the economy.

Monetary forces now began to work in the same direction. Through most of 1919 the Federal Reserve authorities had rediscounted liberally for the banks in order to help the Treasury sell the securities necessary to finance the federal deficit. Toward the end of the year the improvement in federal finances and a deteriorating gold reserve ratio spurred the Reserve authorities to tighten credit. Rediscount rates were raised several times between November, 1919, and May, 1920, and informal pressure was placed on member banks to curb the expansion of credit, particularly for speculative purposes. Parallel with these developments, interest rates rose markedly.

The volume of loans did not decline until well after the turning point had come; but the rate of credit expansion was retarded, and undoubtedly various types of speculative ventures were hampered.

In these circumstances, a sharp downswing was inevitable eventually. A combination of factors seems to have been responsible for the beginning of the downswing early in 1920. In some industries in which raw-material prices had risen most—cotton and woolen textiles, for example—businessmen began to fear price declines and hence curtailed purchases. In textiles, the curtailment began as early as January, 1920, and was accelerated by growing consumers' resistance to high prices, though department store sales did not actually decline until some months later. The existence of heavy inventories permitted retailers, wholesalers, and manufacturers to curtail their purchases sharply once they began to anticipate a decline.

Beginning early in 1920, also, exports showed a tendency to fall, and there were growing indications that foreign countries could not continue much longer the existing level of export demand. This again reacted directly on business expectations and on production. The peak in wholesale prices in a number of countries seems to have come in April; in this country the peak was in May. By the latter month, production had already been declining for several months in various textile lines, in several food processing industries, and in automobile and lumber production; construction had been falling since the beginning of the year; and reduced ordering and even cancellations of orders were beginning to be widespread.

The downswing that now occurred was severe but relatively short. Its outstanding feature was the extreme decline in prices. Wholesale prices declined by about 45 percent, and the drop in some prices, particularly of farm products, was even greater. Eventually steel and other durable goods, for which the demand remained high during the summer, felt the decline also, and the output of these commodities fell sharply. The depression was world-wide, though complicated by financial difficulties and continued efforts toward reconstruction and monetary stabilization in various European countries.

Recovery in the United States was not long delayed. The drastic curtailment in the purchase of raw materials and the decline in production, while consumers maintained their purchases fairly well

with the help of accumulated savings, led to a rapid liquidation of inventories at all stages of production and distribution. Building costs came down promptly, and the volume of construction improved steadily after the close of 1920. Wage decreases were both general and substantial, and there was apparently a widespread improvement in labor efficiency with a consequent decline in production costs. Credit conditions gradually eased as business liquidated its bank borrowings and banks reduced their indebtedness to the Federal Reserve system. By the middle of 1921, the worst of the price decline was over, and some prices were advancing. Exports continued at a high level, though not in so large a volume as in 1920. Gradually confidence returned, and retailers and manufacturers began to increase their inventories and outstanding orders. The process was helped by the fact that the decline in prices had been fairly well balanced; prices of finished products fell nearly as much as those of raw materials; the decline in agricultural prices was not so much greater than that in industrial prices as to lead to serious distortions of the price structure. In this respect 1920–1921 presents a much better picture than 1929–1932.

Most important of all, the long-term investment opportunities that had existed in 1919–1920 were still present in 1921. The need for commercial and residential buildings of all types was very great. In particular, there was a heavy pent-up demand for housing. The automobile industry was still in the stage of rapid growth and so required further expansion in an array of auxiliary industries, in road-building, and in other directions. There was an increasing demand for electric power and electric equipment. Public construction, especially by local government bodies, also provided a strong stimulus. The favorable investment situation is further suggested by the fact that new security issues declined surprisingly little. In short, the downswing in 1920–1921 was associated chiefly with short-run factors—overaccumulation of inventories, speculative excesses, some horizontal maladjustments, tightening credit, and a change in *short-term* expectations. As soon as the short-term factors had been corrected, the favorable long-term investment situation again made itself felt, and recovery set in.

In view of the importance of these short-term factors, there is some reason for thinking of 1919–1921 as an unusually severe minor

cycle.[3] On the whole, however, our concept of overlapping major cycles is a better way of describing what happened.[4] The decline in federal spending and in foreign demand and the collapse of private speculation brought one major cycle to an end. But there were already at work in 1920 powerful expansionary forces which did not reach their full strength until some of the most important distortions created by the boom were eliminated. These forces came into play promptly in 1921 to expand investment and consumption. Thus there was a prompt transition from one major cycle to another. In effect, the war major cycle (1914–1921) overlapped the postwar major cycle that reached its peak in 1929. Short-run and speculative influences, which led to the collapse of short-term expectations in 1920, made the transition adjustment more severe than it otherwise would have been. But the new major-cycle stimuli were there to cut the depression short. In this respect the 1921 depression differed markedly from that of the 1930's or, for example, the protracted depressions of the 1870's and 1890's.[5]

THE PROSPERITY OF THE 1920's

NATURE OF THE EXPANSION IN OUTPUT AND EMPLOYMENT

Business recovered rapidly from the depression of 1921, and by 1923 total output substantially exceeded the peak reached in 1919–1920. From 1923 through 1929, business remained at a high level and tended to increase still further, with minor interruptions in 1924 and 1927. The period culminated with a particularly large increase in industrial output and total GNP in 1929. (See Figure 39 and Table 23.)

The components of the gross national product for the 1920's, as estimated by Simon Kuznets, are given in Table 23. These are not the figures of the Department of Commerce described in Chapter 3. In particular, these figures do not show separately government expenditures, which are concealed in the data on the flow of consumers' goods and on gross capital formation.

The middle years of the 1920's were marked by a steady increase

[3] On p. 263, above, we suggested that unusually severe but short recessions might be referred to as "hybrid" contractions.

[4] Cf. p. 301, above.

[5] Cf. Gordon, *op. cit.* More detailed treatment of the 1919–1921 cycle will also be found in the studies by Wilson and Slichter cited in footnote 1.

TABLE 23. Gross National Product and Its Chief Components, 1919–1929[a]
(Billions of dollars. In 1929 prices.)

	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929
Gross national product	67.8	68.5	65.5	70.4	80.0	81.6	84.3	89.8	90.6	91.9	98.0
Flow to consumers	49.7	51.3	54.1	56.5	61.2	65.3	64.0	68.9	70.7	72.5	76.9
Perishable	19.9	21.0	21.8	22.6	23.5	25.3	25.1	26.3	26.8	26.7	28.0
Semidurable	7.5	6.5	7.8	8.9	9.8	9.0	10.0	10.0	11.2	11.2	11.8
Durable	5.0	4.9	4.0	5.1	6.6	6.9	7.8	8.6	8.2	8.4	8.8
Services	17.3	18.9	20.4	19.9	21.3	24.1	21.2	24.0	24.5	26.2	28.4
Gross capital formation	18.1	17.2	11.4	13.9	18.7	16.2	20.3	20.9	19.9	19.4	21.1
Producers' durables	5.5	5.3	3.6	4.2	5.8	5.4	6.0	6.5	6.1	6.5	7.5
Construction	6.3	5.4	6.3	8.8	9.7	10.8	12.1	12.8	12.7	12.3	11.2
Residential nonfarm	1.5	1.0	2.1	3.6	4.2	5.0	5.4	5.4	5.1	4.7	3.4
Nonresidential	4.8	4.4	4.2	5.2	5.5	5.8	6.7	7.4	7.6	7.6	7.8
Change in inventories	2.8	4.2	0.0	0.3	2.8	−0.9	1.6	1.2	0.4	−0.4	1.7
Foreign investment	3.5	2.3	1.5	0.7	0.5	1.0	0.7	0.4	0.7	1.0	0.8

[a] Revised estimates of Simon Kuznets, as presented in *Supplement to Summary Volume on Capital Formation and Financing. Part A: Annual Estimates, 1919–1953*, mimeographed, National Bureau of Economic Research, n.d. I have used his "Variant I." Kuznets presents estimates of total construction only, but indicates how this total was built up by adding estimates of residential and nonresidential construction. I have reproduced his calculations to obtain estimates of residential construction and subtracted these from his figures for total construction to obtain the nonresidential component.

in GNP, but the rate of expansion was much less than in either 1921–1923 or 1928–1929. More than four-fifths of the increase in GNP between 1919 and 1929 was in the flow of consumers' goods. The expansion in services was particularly marked, in both absolute and relative terms. The increase in output of consumers' durables, dominated by the rapid expansion in the automobile industry, was even greater in percentage terms, though not in absolute amount.

The outstanding fact about the movement of total capital formation in this decade is the high level reached by 1923 and the maintenance of this level for seven years. We have here a prolonged period of high-level investment in producers' durable goods and construction. Inventory accumulation and foreign investment did not play the same important role that they did in 1919–1920. For the period 1923–1929 as a whole, total capital formation averaged about 19.5 billion dollars (in 1929 prices), compared to 11.7 billions in the prewar decade 1904–1913.[6] It is significant that both producers' and consumers' durables formed a larger fraction of the GNP during the 1920's than during any period before World War I.

We thus have a picture of a prolonged investment boom, which supported a steady expansion in incomes and consumers' demand and at the same time provided the enlarged capacity necessary to meet the rising demand for goods and services. An understanding of the business-cycle history of the 1920's and 1930's lies very largely, though not entirely, in the causes and nature of this investment boom.

Table 24 throws further light on the nature of the sustained prosperity of the 1920's. The two columns for each year show, respectively, the total net income originating and the amount of employment in each major industry. Between 1919 and 1929, net income rose by about 22 billions. Manufacturing accounted for only about one sixth of the increase. The service industries and finance each contributed a larger share to the increase in national income than did manufacturing, and the increase in construction and public utilities was also substantial. Manufacturing employed no more persons in 1929 than in 1919; the increase in factory output was achieved entirely through greater productivity per worker, in good part with the help of more and better equipment. The main in-

[6] Figures for 1923–1929 computed from Table 23; for the prewar period, taken from the Kuznets source cited in Table 23.

creases in employment during the decade were in trade, service, finance, and construction, which together accounted for virtually all of the expansion in employment between 1919 and 1929. Expansion in manufacturing output and employment bulk large in the rapid increase between 1921 and 1923 and again in the final upsurge between 1928 and 1929, but the expansion between 1923 and 1928 was primarily in the other sectors of the economy. The expansion in these

TABLE 24. Income and Employment by Major Industries, 1919–1929 (Y = income originating in each industry. E = employees, excluding self-employed. Income in billions of dollars; employees in millions.) [a]

Industry	1919		1921		1923		1928		1929	
	Y	E	Y	E	Y	E	Y	E	Y	E
Manufacturing	16.2	9.9	12.6	7.6	16.8	9.5	17.9	9.2	19.8	9.9
Mining	1.8	1.1	1.7	0.9	2.0	1.1	1.6	1.0	1.8	1.0
Construction	2.0	1.0	2.0	1.1	3.3	1.6	4.0	1.8	4.1	1.8
Transportation and public utilities	6.0	3.3	6.3	3.0	7.1	3.3	8.0	3.2	8.5	3.3
Trade	10.2	3.9	9.5	3.8	10.1	4.4	11.0	4.9	11.4	5.2
Finance	6.8	0.9	7.8	1.0	8.8	1.0	10.9	1.4	10.9	1.5
Service	6.1	3.6	6.7	3.8	8.3	4.4	10.7	5.2	11.3	5.5
Government	3.8 [b]	3.5	6.2	2.8	7.0	2.7	8.3	3.1	8.9	3.2
Miscellaneous	2.2	1.2	2.0	1.1	2.7	1.3	3.7	1.6	3.5	1.7
Agriculture	10.9	2.1	5.5	2.0	6.7	2.0	7.3	2.0	7.7	2.0
Total	65.9	30.4	60.3	27.1	72.9	31.4	83.4	33.4	87.8	35.1

[a] From Simon Kuznets, *National Income and Its Composition, 1919–1938*, I, National Bureau of Economic Research, 1941, pp. 310, 314. The income figures are not adjusted for price changes.

[b] This item, from some points of view, is not fully comparable with the figures for income originating in government for later years. In 1919 the large government deficit not matched by productive capital formation resulted in a large figure for government dissaving, which reduced the figure for income generated by government shown in the table.

fields—trade, service, finance, construction, state and local government, and electric power and communications[7]—together with new products and technological developments in manufacturing, help to explain the expansion of aggregate demand in the 1920's.

The investment boom and the rise in consumption during the 1920's were accompanied by a steady expansion in bank credit, the flotation of an enormous volume of new security issues, and a mounting tide of speculative fever reflected particularly in the pro-

[7] The expansion in electric power and in the telephone industry, particularly the former, was very rapid during the 1920's but is concealed by horizontal or downward movements in railroad and other forms of public transportation.

motion of new enterprises, a boom in real estate, the development of a variety of unsound financial practices, and a violent upsurge in stock prices which culminated in the stock-market crash of 1929. Despite these indications of a rapid rise in total expenditures, commodity prices showed remarkable stability during the 1920's; if anything, wholesale prices tended to decline, particularly from 1925 on (see Figure 39). These and related developments will occupy us at some length in the following pages.

NONMONETARY STIMULI TO INVESTMENT

The main underlying factors responsible for the high level of investment in the 1920's were: (1) pent-up demands for plant and equipment created by the war[8] (2) the direct and indirect effects of the automobile; (3) the rapid expansion of other relatively new industries such as electric power, electrical equipment, radio and telephone, air transportation, motion pictures, and rayon; (4) the rapid pace of technological change, leading to great increases in labor productivity; and (5) the rise to a peak of a long building cycle. Superimposed on these was a wave of optimism that must in part be treated as an independent factor, a fairly high propensity to consume, and an elastic credit supply.[9]

It is impossible to say precisely how important the war was in creating a demand for plant and equipment. It is clear, however, that the pent-up demands that existed after the armistice were not satisfied during 1919–1920, and a substantial amount of investment in the early 1920's must have represented replacement and expansion programs deferred from the war years and investment to capitalize on technological changes occurring during the war.

The most important stimulus to investment and to expansion of total output in the 1920's was the automobile. Like electric power, this was a prewar innovation. But its full impact on the American economy was not felt until the 1920's. Production of motor vehicles had already risen from 485,000 in 1913 to 1,934,000 in 1919. Produc-

[8] This was especially true of the railroads, and it was also probably true of the public utilities. In manufacturing, on the other hand, pent-up replacement demands were probably largely satisfied during 1919–1920. Cf. Gordon, *op. cit.*, p. 188.

[9] For a stimulating and more detailed discussion of many of the topics considered in this subsection, see J. A. Schumpeter, *op. cit.*, vol. 2, pp. 767 ff. See also the more detailed analysis of investment stimuli in the 1920's in Gordon, *op. cit.*

tion jumped to 4,180,000 by 1923 and then rose further to a peak of 5,622,000 in 1929.

The effect of the automobile on aggregate demand came from two sources—the expansion in the *production* of cars and trucks and the enormously increased *use* of motor vehicles. The increase in production created a demand for new plant and equipment in both the automobile industry and the industries that served it: parts and accessories, rubber, steel, plate glass, lead, etc. These auxiliary industries were able, as automobile production grew, to expand their own production and employment and to invest in new plant and equipment.

Even more important was the growing *use* of automobiles. Motor vehicle production nearly trebled between 1919 and 1929, but the increase in registrations—the number of cars and trucks on the road—was even larger. And steadily greater use was made of each vehicle. The result was an enormous expansion in employment in oil refining, filling stations and garages, truck and bus driving, selling of supplies and accessories, and construction and repair of roads. Expansion in these activities meant new investment—in buildings, equipment, and roads. And as the automobile changed methods of living, still further investment was required—in the development of suburban communities, for example.

Another prewar innovation, electric power, was a highly important stimulus to investment. Electric power production more than doubled between 1920 and 1929, and generating capacity increased in proportion. Use of this power in turn required electrical equipment and opened up methods of reducing costs that involved other types of new machinery. Value added by the electrical machinery industry also more than doubled between 1919 and 1929, compared to an increase of about 30 percent for manufacturing as a whole. Along with the growth of electric power production and the use of electrically driven machinery and handling equipment in industry went rapid expansion in the telephone industry (again a prewar innovation), the growth of radio (entirely a postwar development), and the rapid electrification of the home.

Other new industries and products helped to maintain investment and expand production—various chemical products (particularly rayon), oil and rubber products other than gasoline and tires, natural gas, production and distribution of motion pictures, the air-

plane, and so on. Most of these represented prewar innovations that added more to output in the postwar than in the prewar period.

Also highly important in stimulating investment was the increased tempo of technological change in the 1920's. Productivity per man-hour in manufacturing rose some 70 percent between 1919 and 1929.[10] Mass production techniques were extended, greater use was made of automatic and special-purpose machinery, radical improvements occurred in material-handling methods, and so on. These developments made a major contribution to the demand for producers' durable goods. Labor costs fell steadily during the 1920's, as wages failed to rise as rapidly as productivity increased. As a result, stable or falling prices went together with expanded profit margins. The latter bolstered expectations and encouraged further investment; the former led to illusions, in the midst of the speculative boom of the late 1920's, that "conditions were fundamentally sound" because commodity prices were not rising.

As would be expected, corporate profits were high during most of the 1920's, though some industries (e.g., textiles) made a less profitable showing than others.[11] Profit per unit in manufacturing was stable at a high level during 1923–1926, declined in 1927, and rose above the 1923–1926 level in 1929. The rate of profit on invested capital of manufacturing corporations remained at a high level, with no marked trend either upward or downward, between 1923 and 1929.[12]

Table 23 has already indicated how important construction was in maintaining investment in the 1920's. The sources of demand for building are revealed by the data in Table 25. The most important single component of new construction was residential building, which comprised 40 percent or more of the total through 1926, when a decline set in which lasted until 1933. In the nonresidential field, public-utility, government, and "other" construction (i.e., stores, office buildings, etc.) were all more important than strictly industrial (i.e., factory) building. About half the government figure rep-

[10] Based on data in U.S. Bureau of Labor Statistics, *Productivity and Unit Labor Cost in Selected Manufacturing Industries, 1919–1940,* 1942, p. 1; *Handbook of Labor Statistics,* 1947 ed., p. 155.

[11] See *Survey of Current Business,* September, 1944, p. 12.

[12] Compare F. C. Mills, *Economic Tendencies in the United States,* 1932, pp. 492 *ff.*

resented road building. Some of the large volume of building represented accumulated demand from the war years and the needs of an expanding population; part was in response to speculative enthusiasm and the ease with which mortgage credit could be obtained; part was the direct result of the automobile and the changes

TABLE 25. New Construction Activity in the United States, 1919–1939 (In billions of dollars) [a]

Year	Total	Residential (nonfarm)	Private Nonresidential				Government
			Industrial	Farm	Public Utility	Other [b]	
1920	6.7	2.0	1.1	0.6	0.8	0.9	1.4
1921	6.0	2.1	0.6	0.2	0.6	0.9	1.6
—							
1923	9.3	4.4	0.5	0.3	1.2	1.3	1.6
1924	10.4	5.1	0.5	0.3	1.4	1.3	1.9
1925	11.4	5.5	0.5	0.3	1.3	1.7	2.1
1926	12.1	5.6	0.7	0.3	1.4	1.9	2.1
1927	12.0	5.2	0.7	0.4	1.5	2.0	2.4
1928	11.6	4.8	0.8	0.3	1.4	1.9	2.5
1929	10.8	3.6	0.9	0.3	1.6	1.8	2.5
—							
1933	2.9	0.5	0.2	0.0	0.3	0.3	1.6
—							
1936	6.5	1.6	0.3	0.2	0.5	0.5	3.5
1937	7.0	1.9	0.5	0.2	0.7	0.6	3.1
1938	7.0	2.0	0.2	0.2	0.6	0.6	3.4
1939	8.2	2.7	0.3	0.2	0.7	0.6	3.8

[a] From U.S. Departments of Labor and Commerce, *Construction Volume and Costs, 1915–1954*, statistical supplement to vol. I of *Construction Review*, 1955, pp. 1–6.

[b] Includes warehouses, offices, stores, restaurants, and garages, and religious, educational, social, recreational, hospital, and institutional building.

in living habits that it inspired. The large volume of commercial building reflected the great expansion in the trade, service, and finance industries that occurred during the 1920's.[13] Only a minor part of total construction was required by expanding industrial production. Table 25 also reveals how completely these various stimuli disappeared in the 1930's. Failure of building to return even

[13] Cf. Gordon, *op. cit.*, p. 206.

close to the level of the 1920's was one of the important reasons for the disappointing nature of the recovery after 1933.

MONETARY FACTORS

There is some difference of opinion as to the precise role of monetary developments in the boom of the 1920's and the depression that followed. It is fairly clear that the supply of bank credit was fully adequate for the needs of business. It is also obvious that credit ex-

TABLE 26. Loans, Investments, and Deposits of Commercial Banks, 1914–1939
(In billions of dollars) [a]

Year (June 30)	Loans and Investments			Adjusted Demand Deposits and Currency Outside Banks [b]	Time Deposits of Commercial Banks
	Total	Loans	Investments		
1914	16.9	13.2	3.7	11.6	4.4
1921	34.2	26.1	8.1	20.8	10.9
1923	37.1	26.9	10.2	22.7	13.4
1929	49.4	35.7	13.7	26.2	19.6
1933	30.4	16.3	14.0	19.2	10.8
1937	39.5	17.4	22.0	30.7	14.5
1939	39.4	16.4	22.9	33.4	15.1

[a] Taken from Board of Governors of the Federal Reserve System, *Banking and Monetary Statistics*, 1943, pp. 19, 34–35.

[b] Adjusted demand deposits exclude interbank and United States government deposits and cash items in process of collection.

pansion helped to finance the speculative boom in securities and real estate. It is not clear, however, that the boom was due primarily to easy credit conditions or that a different monetary policy in 1927–1929 or earlier could have prevented the depression, although it might have helped to make it less severe.

The American economy began the 1920's with a money supply about double that of 1914. Between 1921 and 1929 total loans and investments of commercial banks, as well as deposits, increased still further, as is indicated in Table 26. The expansion in loans after 1921 took primarily the form of loans on securities, and there was

also a significant increase in real-estate loans. A substantial part of the increase in investments went into corporation securities. In short, the credit expansion of the 1920's served to support the large volume of security flotations during the period and the growing speculation in the stock market. It is interesting to note in this connection that demand deposits rose only moderately; the chief increase was in time deposits. These was a growing tendency, encouraged by the banks, to classify relatively idle demand deposits as time deposits, on which a higher rate of interest was paid. The increasing volume of such deposits resulted in part from the fact that high security prices encouraged firms to float securities in excess of their immediate needs. Thus the banks' financing of security operations helped to bring these time deposits into existence, made large firms less dependent on the banks for working-capital requirements, and created a reservoir of loanable funds in the hands of nonbank lenders which could find their way into the stock market. To the extent that credit expansion took the form of time deposits, which required lower reserves than demand deposits, the lending capacity of the banks was increased.[14]

By earlier standards, interest rates were not unduly high during the 1920's. High-grade bond yields and short-term interest rates were relatively stable and showed a moderate tendency to decline between 1923 and 1927. In 1928–1929 all classes of interest rates rose under the stimulus of the large-scale speculative demands for credit and the attempts finally made by the Federal Reserve authorities to bring the boom under control.

The Reserve authorities have frequently been criticized for not taking any decisive action to curb the boom until 1928. In 1924 and again in 1927 the Federal Reserve Banks reduced rediscount rates and bought securities in the open market with the twofold objective of alleviating the mild recessions that occurred in those years and creating conditions favorable to the restoration of monetary stability in Europe. The system sold in 1928 somewhat more than the securities it had bought in 1927, and the New York Bank's rediscount rate was raised in several steps from 3½ percent in January, 1928, to 5 percent in July. But more than a year went by before the

[14] And the excess cash holdings of business firms made the latter less susceptible to control from the monetary authorities.

rediscount rate was raised further, to 6 percent in August, 1929. From the middle of 1924 to January, 1928, the rediscount rate did not go above 4 percent.

It is doubtful whether these monetary developments were of primary importance in creating the boom of the 1920's. In the terms used by some business-cycle theorists, the "natural rate" of interest was higher than the market rate, and part of the large volume of investment was financed by credit expansion. But the chief reason for this lay in the nonmonetary sphere—in the developments discussed earlier that made the marginal efficiency of capital high and in the wave of speculative optimism that raised it still higher. In addition, the unwise lending practices of the commercial banks encouraged speculation and unsound promotions and weakened the banking system's ability to withstand the strains that were to come after 1929. It was the nature rather than the amount of lending that led to later trouble. It is doubtful whether the reserve banks could have taken any action then within their power that would have prevented the stock-market boom and yet permitted a level of investment high enough to maintain business activity close to full employment levels.

SPECULATION AND FINANCE

The most spectacular aspect of the "New Era" was the stock-market boom.[15] The extent of the rise in industrial stock prices is indicated in Figure 39. Over a billion shares changed hands on the New York Stock Exchange in 1929 and nearly that amount in 1928, compared to an annual average of about 250 million shares during 1922–1924 and even less in the prewar years. With this rise went an enormous expansion in brokers' loans to finance security purchases on margin.

The rise in stock prices, together with the investment opportunities described previously, stimulated the offering of a tremendous volume of new security issues. More than 30 billion dollars' worth of new issues were put on the market in the three years 1927–1929 alone. The relative importance of the different types of issuers is suggested by the following figures (in billions of dollars):[16]

[15] For an intertaining account of the stock-market boom and collapse, see J. K. Galbraith, *The Great Crash*, 1955.

[16] From Board of Governors of the Federal Reserve System, *op. cit.*, p. 487.

Issues to raise new capital, 1927–1929		
State and municipal	4.3	
Federal agencies	0.2	
Corporate		
Bonds and notes	7.6	
Stocks	10.4	
Foreign	3.3	
		25.7
Refunding issues		5.5
Total		31.2

The magnitude of the amount of new capital raised by corporations, particularly in the form of stocks, is especially to be noted.[17] The figures also point up the fact that state and local governments were borrowing heavily in this period.

By no means all of the new capital issues in these years went into real capital formation. Indeed, the major part, particularly from 1926 on, seems to have gone into erecting a financial superstructure of holding companies, investment trusts, and other forms of intercorporate security holdings that was to come crashing down in the 1930's. Investment bankers were active in the promotion of companies to hold the securities of other companies, and commercial banks were involved also, through their own investment-banking affiliates and through loans on securities. Large promoters' profits were made, and capital gains from the sale of securities by the former holders inflated consumers' demand and spread the speculative fever in widening circles.

Similar speculative developments involving the inflation of capital values occurred in the real-estate field. Homes, apartment houses, office buildings, and hotels were built with almost reckless abandon under the spur of promoters' profits and the ease with which securities could be sold to finance the cost of construction. Banks loaned heavily on bonds and mortgages without adequate safeguards as to amortization and later found themselves with "frozen assets," the values of which had to be drastically scaled down.

The consequences of these financial developments need no great elaboration. One result was a good deal of real investment that was

[17] Issues of bonds by corporations exceeded their stock issues until 1928.

not justified in terms of long-term profit possibilities. Capital goods were created which were to "hang over the market" and discourage further investment for a decade after 1929. The banking system was seriously weakened. Many weak business ventures were saddled with a load of fixed charges that could lead eventually only to the bankruptcy court. Business expectations in many fields became geared to a level of capital values that could not be maintained indefinitely. The rise in security prices created capital gains, particularly in the upper income groups, and thus put an artificial support under the demand for luxury and durable goods that would collapse with the eventual and inevitable break in security prices.

AGRICULTURE AND RAW MATERIALS

American agriculture did not share in the boom of the 1920's to the same degree as the rest of the economy. The disruption of European agriculture had led to an unprecedented boom in American farm production, prices, and exports during the war and immediate postwar years, a boom that was bound to end once European agricultural output returned to something like its prewar level. The situation was made worse by the speculative rise in farm land prices and a heavy increase in farm mortgage debt. The increase in debt, which reached a peak in 1921, proved a heavy burden after the collapse of farm prices in 1920–1921, and foreclosures of farm property were high throughout the middle 1920's.

Nonetheless, American agriculture cannot be said to have been depressed during the 1920's, though seeds of difficulty lay under the surface that were to ripen all too rapidly once the domestic demand for farm products began to fall after 1929. Between 1921 and 1929, income in agriculture showed a percentage increase not greatly below that of total national income. But the rise in farm income was largely over by 1925, the movement being approximately horizontal from then until 1929. In short, American agriculture had by 1925 made a substantial recovery from the very low level reached in 1921, but it contributed little to the expansion in total income thereafter. These indications of a stable demand for agricultural products were a resultant of two sets of forces: rising domestic consumption and declining foreign demand. The percentage of total exports made up of agricultural products showed a significant decline over the decade as a whole. Although this result was accentuated by the growth of agricultural protectionism in Europe, it reflected chiefly

a tendency, associated with the growing industrialization of this country, that had been going on for decades.[18]

Technological developments increased investment and output in agriculture during the 1920's, but the effects were not so marked as elsewhere in the economy. Use of tractors, trucks, and automobiles spread rapidly, and the "all-purpose" tractor tended to encourage the movement toward large-scale farming. The full effect of these and other technological developments, however, was not fully felt until the 1930's. During the 1920's productivity in agriculture increased much less rapidly than in industry, one result being that the ratio of farm to nonfarm prices tended to rise. The combined effect of technological developments and contracting export markets, however, did tend to depress the prices of some important types of farm products, particularly from 1925 on.

The pressure of increasing supplies was even more noticeable in the world markets for a variety of important agricultural products and raw materials. Following the breakdown of a number of commodity control schemes, particularly from about 1925 on, prices of primary products registered important declines. This was true, for example, of coffee, raw silk, petroleum, lead, and zinc—while the price of crude rubber dropped more than 70 percent in the four years following 1925. These falling prices, taken in conjunction with increasing production and, in some cases, rather rapidly accumulating stocks, suggest that a condition of world overproduction in some primary products was developing before the 1929 downturn. For many of these primary products, both demand and supply were inelastic—which meant that falling prices would have little effect in either reducing supply or increasing consumption.

As important as these developments were, they did not bring on or determine the timing of the downturn in 1929. Their cyclical significance lies in the contribution they made to the severity of the Great Depression. This is true also of the international financial and monetary developments in the 1920's, to which we now turn.

INTERNATIONAL DEVELOPMENTS

The years 1922–1929 were marked by expanding economic activity throughout the world. The relative increase in industrial produc-

[18] Cf. H. Barger and H. H. Landsberg, *American Agriculture, 1899–1939*, 1942, p. 293; Margaret S. Gordon, "International Aspects of American Agricultural Policy," *American Economic Review*, vol. 36, September, 1946, p. 598.

tion outside the United States (but excluding Russia) in these years was, on the average, not greatly different from that in this country. But the rate of expansion was not uniform. Europe was slow to recover after the war, whereas countries undergoing rapid industrialization, such as Japan, Russia, and Canada, showed the largest rates of increase.[19]

American loans played a major role in encouraging expansion in other countries. The war had made the United States a great creditor nation, and, as suggested by Table 27, this country continued to export capital on a large scale through the 1920's. Net payments due to other countries for freight, tourists' expenditures, and other "invisible" items did little more than offset the large sums due the United States on account of interest and dividends. (See Table 27.) There remained the large export surplus from the United States to be paid for, and, in effect, it was financed by loans. In this connection, it is interesting to note that the expansion in American exports between 1922 and 1929 was in automobiles and the products of the capital-goods industries and that most of the increase went to the relatively less industrialized parts of the world. Exports to Europe scarcely increased at all during this period.

The nature of international lending during the 1920's did not make for long-run stability in the world. Many of the American loans were unwisely made. American investment bankers, inexperienced in the international field, encouraged firms and governments to borrow more than they could productively use. Germany was the largest borrower during this period, needing capital for internal reconstruction and foreign exchange with which to make reparation payments. The ease with which foreign loans could be secured led to unwise public expenditures, particularly by local German governments, impaired Germany's ability to export by inflating her costs and prices, and temporarily concealed her inability to pay the existing scale of reparations in the absence of continued borrowing. Similar overborrowing, though on a smaller scale, took place in some of the Latin American countries.

The flow of American capital was sharply reduced after the first part of 1928. The stock-market boom and rising interest rates led

[19] Cf. U.S. Department of Commerce, *The United States in the World Economy*, 1943, p. 150; League of Nations, *Industrialization and Foreign Trade*, 1945, pp. 134–135; and Lewis, *op. cit.*

American investors to keep their capital at home and, indeed, even induced a movement of foreign funds to the United States. The abrupt decline of capital exports immediately created difficulties in the debtor countries, particularly those that relied on exports of

TABLE 27. Balance of Payments of the United States, 1920–1937
(In millions of dollars) [a]

Type of Transaction	1920	1924	1928	1929	1932	1937
A. Current transactions						
Exports	8,228	4,591	5,128	5,241	1,611	3,349
Imports	5,278	3,610	4,091	4,399	1,323	3,084
Balance of merchandise trade	+2,950	+981	+1,037	+842	+288	+265
Net freight payments	+271	−46	−88	−119	−84	−130
Net travel expenditures	−123	−226	−327	−344	−194	−213
Net personal remittances	−516	−268	−288	−288	−182	−142
Net interest and divds.	+468	+462	+647	+652	+325	+282
Other service items	−363	+72	+46	+43	+6	−93
Balance of service items	−263	−6	−10	−56	−129	−296
Balance of all current transactions	+2,687	+975	+1,027	+786	+159	−31
B. Capital movements						
Net long-term movements	−832	−700	−847	−278	+225	+521
Net short-term movements	[b]	+119	−348	−4	−446	+356
Net capital movements	−832	−581	−1,195	−282	−221	+877
C. Net gold movements	+50	−216	+272	−120	−11	−1,271
D. Unexplained items	−1,905	−178	−104	−384	+73	+425

[a] From U.S. Department of Commerce, *The United States in the World Economy*, 1943, Appendix. Plus signs represent a demand for dollars; minus signs, a supply of dollars (or demand for foreign currencies).

[b] Not available.

agricultural products and raw materials, the prices of which were already falling in world markets. Even in 1928 some of these countries could not export enough to meet the service charges on their external debt, and by 1929 the situation had become acute in a number of countries. Australia and the Argentine were forced off the gold standard late in 1929, and several other countries found their curren-

cies depreciating before the end of the year. Early signs of difficulty appeared also in Germany.

The inherent instability in the international situation was made all the more serious by the increasingly important role being played by short-term capital movements. Short-term funds were held in the leading world money markets by European banks and investors (particularly French), who tended to shift their balances from country to country in response to actual or anticipated changes in money market conditions or exchange rates. Foreign balances in New York increased during the 1920's and were quite large by 1929. The situation in Great Britain was particularly serious. That country had returned to the gold standard in 1925 at too high a value for the pound (thus making British goods relatively expensive in foreign markets) and then sought to protect her weakened balance-of-payments position by maintaining high interest rates in London. This policy, together with the re-establishment of London as a world financial center after the return of the pound to a gold basis, attracted large amounts of foreign funds to Britain. The sudden exodus of these funds when confidence in the pound was impaired in 1931 was to drive England off the gold standard.

Germany was also a heavy short-term borrower. Here again the flight of short-term capital was to bring a crisis in 1931. France, on the other hand, was an important short-term creditor; and the return of capital to France after the official stabilization of the franc at too low a level in 1928 tended to drain gold out of England and other countries.

It is, of course, impossible to say what the course of the business cycle would have been in the rest of the world if the American boom had not collapsed in 1929. A downswing of some sort was probably called for by 1929 or 1930 merely as a result of the weak position of many world primary markets, and serious difficulties in the rest of the world would certainly have eventually affected the United States. In retrospect, we can see that the turning point in some countries with balance-of-payments difficulties came before that in the United States. But there is no evidence that the weaknesses in the international situation brought the American boom to an end. The immediate causes of the downturn in the United States lay in domestic developments, and these we shall look at further in the next section. But the international developments described in the preced-

ing paragraphs did play a crucial role in determining the extent and severity of the depression once the downswing had begun.

THE TURNING POINTS, 1923–1929

Before examining the critical turning point in 1929, we shall pause briefly to discuss the minor reversals in business activity in 1923–1924 and 1926–1927.[20] According to the reference dates published by the National Bureau of Economic Research, the cyclical phases following the low point in 1921 were as follows (compare Figure 39):

Expansion	July, 1921—May, 1923
Contraction	May, 1923—July, 1924
Expansion	July, 1924—October, 1926
Contraction	October, 1926—November, 1927
Expansion	November, 1927—August, 1929

These are the phases of the minor cycles that were superimposed on the major upswing of 1921–1929. The recessions in 1923–1924 and 1926–1927 were quite mild and brief and were associated, particularly in 1923–1924, with changes in short-term business expectations. In both cases, the continuance of long-term investment opportunities tended to maintain employment and lead to quick recovery.

The downturn in 1923 stemmed largely from the rapidity of the rise after 1921. With the debacle of 1920–1921 fresh in their minds, businessmen became concerned over the rapid rise in prices in 1922 and the early months of 1923. This concern was reinforced by a slight rise in Federal Reserve discount rates and by warnings in the financial press.[21] Building activity fell off slightly in response to rising costs. Beginning in the late spring of 1923, business firms began to curtail output, and a general decline set in, interrupted by a brief abortive recovery in the first two months of 1924. A change in short-term expectations seems to have been the chief factor bringing on the recession. Minor horizontal maladjustments affecting par-

[20] For a more detailed analysis of these turning points, as well as that in 1929, see the studies by Slichter, Wilson, and Schumpeter listed in footnote 1 of this chapter.

[21] The reserve banks also sold securities in the open market during the first half of 1923, but the effect on interest rates was minor. Commercial loans continued to expand through most of 1923.

ticular industries contributed to this downward shift in anticipations.

The low point in business was reached in July. Money conditions were easy, partly in response to open-market purchases and a reduction in rediscount rates by the Federal Reserve system. Inventories had been reduced, and retail sales had fallen relatively little. Building activity increased, and new security issues were in large volume. A number of foreign developments reacted favorably on business sentiment, and large domestic crops associated with poor harvests abroad had a favorable influence on agricultural incomes. Most important of all, underlying investment opportunities continued favorable. Indeed, on the basis of annual data, output and employment in 1924 were not greatly below 1923. Production recovered sharply in the latter part of 1924 and then continued to expand at a more moderate pace in 1925 and 1926.

"The recession which began in the fall of 1926 was so mild that one hesitates to regard it as a recession in general business."[22] The Federal Reserve index of industrial production showed a total decline of only about 6 percent. The recession in manufacturing output was entirely in durable goods, which declined about 4 percent between September, 1926, and May, 1927, when the Ford Motor Company closed down in the course of changing to a new model, and then fell a further 10 per cent to November, after which a rapid recovery set in. The production of nondurable goods scarcely showed any recession, and neither did department store sales. According to Table 23, total GNP (corrected for price changes) was slightly larger in 1927 than in 1926. Since most of the decline occurred after the Ford shutdown, it is a fair inference that, in the absence of this occurrence, no noticeable recession would have developed. It is worth pointing out, however, that there were some elements of weakness in the latter part of 1926. Residential construction had shown a moderate decline from its 1925 peak; there had been some reduction in automobile output since the latter part of 1925; wholesale prices had been falling since the middle of 1925, with some resulting decline in farm incomes; and there was a tendency for our foreign trade balance to decline.

The weakness in prices was world-wide and was connected with developments that were discussed in the preceding section. These

[22] Slichter, *op. cit.*, p. 11.

deflationary forces were offset for the time being by the easy-money policy of the Federal Reserve authorities in 1927, which induced a temporary outflow of gold and capital from the United States, and by accentuation of the boom in 1928–1929.

By the end of 1927, Ford had announced his new model, and automobile production rose rapidly thereafter. Prices had begun to move upward earlier in the year, and agricultural incomes improved; construction activity was again expanding; credit was easy; new security issues had continued at a high level; and stock prices had continued their upward course. Business expectations, which had scarcely felt any setback, improved further; and thus began the final spurt before the collapse of 1929.

The expansion in industrial output in 1928–1929 was extremely rapid. (See Figure 39.) The Federal Reserve index of industrial production rose by 24 percent between November, 1927, and July, 1929, the rise in durable-goods production being particularly marked. Gross capital formation expanded sharply between 1928 and 1929. All of the increase was in producers' durable goods and inventory accumulation. Consumption rose, but at a slower pace; the marginal and average propensity to consume was lower in 1929 than in 1928.

The sharp upsurge in production was in the face of declining prices from the fall of 1928 on. Nonagricultural prices had been drifting downward slowly since 1925. Nonetheless, profits were high and continued to rise till about the third quarter of 1929. Wage rates rose relatively little, and the increase in factory payrolls was not so great as that in factory output. The expansion in business was also in the face of a rise in interest rates after the beginning of 1928, which was associated with the mounting speculative demands for credit and with attempts of the Federal Reserve system to bring the stock-market boom under control.[23] High-grade bond yields rose moderately after the beginning of 1928 following an almost steady though mild decline since 1923. New security issues rose to a new high in 1929, and the volume of stock issues—particularly of investment trusts and holding companies—was abnormally large.

[23] Federal Reserve pressure was exerted in three stages: sale of securities and increase in rediscount rates in the first half of 1928; the Board's famous warning against bank loans for speculative purposes in February, 1929; and increase in the rediscount rate of the New York Reserve Bank to 6 percent in August, 1929.

The stock-market crash came in October, but most observers put the turning point in business several months earlier. The National Bureau's date is August, and the peak in the index of industrial production came at about the same time.[24] Some other indicators turned down even earlier. Retail sales, however, did not decline until the last quarter of the year. We may view the period from about March to October as representing the "turning point zone" or critical period, within which the forces making for deflation gradually came into ascendancy.

It is impossible to give a complete and precise statement of the immediate causes of the downturn. Certainly the full explanation of the extent and severity of the Great Depression is not to be found merely in the sequence of events during 1928–1929; we must look at the boom of the 1920's as a whole and at the course of developments during 1930–1933.

Nonetheless, it is possible to find in the situation in 1929 important elements of weakness that were sufficient to create a depression more severe than that of 1924 or 1927. It is clear that the rise in output of durable goods in 1928–1929 was too rapid to be long maintained. Excess capacity was developing in a number of lines, and this meant a decline in demand for further capital goods. As a matter of fact, new orders for some types of durable goods declined fairly early in 1929. The automobile market was clearly oversold; in addition, the industry's capacity exceeded even the peak production of 1929. The tire industry had been overbuilt, and tire production had fallen sharply in the latter part of 1928. The textile industries had been suffering from overcapacity for some time. Residential construction had been declining sharply since the beginning of 1928, and an overbuilt situation obviously existed in that area. Some of these developments may be described as the result of the belated and rough working of the acceleration principle, although it should

[24] The National Bureau's dates for the turning point in several other countries are: France, March, 1930; Great Britain, July, 1929; Germany, April, 1929. In a more comprehensive compilation, the Brookings Institution lists the turning point in different countries by quarters as follows: first quarter, 1929: Poland; second quarter: Canada, Argentina; third quarter: United States, Belgium, Italy, Egypt; fourth quarter: Switzerland, Netherlands, Austria, Czechoslovakia, India, British Malaya; first quarter, 1930: United Kingdom, Japan, New Zealand, South Africa; second quarter of 1930 or later: France, Sweden, Ireland, Yugoslavia, Norway, Denmark. See *The Recovery Problem in the United States,* 1936, chart following p. 28.

be emphasized that we can trace no simple correlation between the short-period changes in the rate of increase in output and in the demand for capital goods.

The tendency for buyers' markets to develop in 1929 probably weakened business expectations, and some concern was created by fears arising out of the final excesses of the stock-market boom. Accumulating surpluses in agricultural products and raw materials also created a vulnerable situation, which, however, was not fully revealed until after the stock-market crash. There was increasing pressure on the balance of payments of various foreign countries as the stock-market boom and high interest rates in the United States stopped the outward flow of capital, but the fundamental weaknesses in the international economic situation were not to be fully revealed until the depression had gathered momentum in 1930–1931.

Except in a few sectors, the decline in activity was not severe until after the stock-market collapse, which undoubtedly led to a sharp downward revision in expectations and had a considerable effect on the demand for luxury and durable goods. As would be expected, the decline, once it began, was particularly severe in the durable-goods industries.

Although we cannot complete our explanation of the causes of the Great Depression until we look at developments during the 1930's, we can dispose now of a number of possible hypotheses as to the major cause of the downturn in 1929. It was clearly not due to an encroachment of costs on profits. Wages rose relatively little, and raw-material prices showed some tendency to sag. While finished-goods prices were also declining as supply pressed on demand, profits were high and rising through the third quarter of the year.

Nor can the downturn be explained by monetary developments. The rise in interest rates was not great enough to discourage business borrowing; the Federal Reserve authorities were careful not to restrict credit for legitimate business purposes; we have already seen that business was becoming increasingly independent of the banks; and commercial loans did not begin to decline until after the stock-market crash. The tightness in credit affected speculation, but this is another matter. It is also clear that a shortage of capital (savings) was not responsible. Had there been such a shortage, prices and wages should have risen markedly as the capital- and consumers'-

goods industries bid against each other for labor and materials. Clearly, also, there was no bumping against a ceiling, as is called for in Hicks' version of a multiplier-accelerator theory. Nor did the capital markets show any inability to absorb new security issues until after the break in stock prices. True, there was a moderate rise in bond yields, but this was scarcely enough to discourage much new investment.[25] Indeed, new security issues were in surprisingly large volume even in 1930; the decline in that year was in stock issues, more new capital being raised by bonds than in 1929. We shall return to the problem of causation in the final section of this chapter.

ECONOMIC COLLAPSE, 1929–1933

The downswing that began in 1929 developed into the worst depression of modern history. It spread throughout the world, feeding on the weaknesses in the international economic situation that were discussed in an earlier section. While world-wide, the depression was more severe, and developed more rapidly, in the United States than in most other industrial countries. Its severity in the United States is suggested by the series portrayed in Figure 39 and by the summary data in Table 28.

The GNP in real terms declined by about 30 percent between 1929 and 1932, and by about 45 percent when expressed in current prices. Industrial production was reduced by nearly one-half. Private investment sank far below the rate needed for replacement, and the output of durable goods at one point was less than a third of the 1929 high. The fall in prices was catastrophic both because of the magnitude of the overall decline and because of the distortions created in the price structure. Farm prices declined to less than 50 percent of their 1929 peak, while industrial prices fell much less. In this country the decline in prices and capital values was so great and so extensive as eventually to threaten the collapse of our entire banking system and to jeopardize the solvency of many of our financial institutions. Internationally the depression completely destroyed the monetary stability painfully built up in the 1920's, demoralized trading and financial relationships between countries, and generally created a condition of international economic paralysis. Domestically the damage done is summarized in the estimate that during

[25] It may have had some effect on construction, but here overbuilding was clearly more important than high interest rates.

most of 1932 and 1933 twelve to sixteen million persons were unemployed in the United States.

The National Bureau of Economic Research places the lower turning point in March, 1933. The monthly series listed in Table 28 reached a low point either in June–July, 1932, or in February–

TABLE 28. Measures of the Severity of the Great Depression and of the Extent of Recovery by 1937

	1929 High	1932–1933 Low	1937 High
Annual data			
In 1939 prices, billions of dollars [a]			
GNP	85.9	61.5	87.9
Gross domestic investment	14.9	1.1	11.4
In current prices, billions of dollars [a]			
GNP	103.8	55.8	90.2
Gross domestic investment	15.8	0.9	11.4
Disposable income	82.5	45.2	71.1
Consumer expenditures	78.8	46.3	67.1
Total new construction	10.8	2.9	7.0
Outside bank debits [b]	645	266	441
Outside demand deposits	18.3	11.3	19.3
Loans, all commercial banks	35.7	14.9 [c]	17.4
Unemployment (million persons)	1.6	12.8	7.7
Monthly data			
Industrial production (1935–1939 = 100)			
Total index	110	58	113
Durable manufactures	132	41	122
Nondurable manufactures	93	70	106
Wholesale commodity prices (1926 = 100)			
All commodities	95.3	64.8	86.3
Farm products	104.9	48.2	86.4
Stock prices (1935–1939 = 100)	238	36	137

[a] Data for the GNP and its components in both constant and current prices are the estimates of the Department of Commerce. These figures have been slightly revised since this table was originally prepared.

[b] For demand deposits. Debits to time deposits are excluded.

[c] Figure for June 30, 1935, when low for this series was reached.

March, 1933. If we use the National Bureau's dates, the downswing continued for 43 months, longer than any other business contraction since the Civil War except one. (The downswing following the crisis of 1873 lasted 65 months.) But if we take into account the extent of the decline, the amount of distress caused, the interna-

tional ramifications, and the slowness of recovery after 1933, we need have little hesitation in rating the depression of the 1930's the most severe in our history.[26] It was truly the "Great Depression."

That a depression of unusual severity was developing did not become clearly apparent until the second half of 1930. After the sharp decline in the last quarter of 1929 there was a slight, abortive recovery in the early months of 1930, associated particularly with a partial recovery in automobile production and some improvement in nonresidential construction.[27] Wage rates were well maintained throughout 1930, and there were public statements against the desirability of wage cuts. But the decline in prices and production and the collapse in the speculative boom which had already taken place had by mid-1930 led to a marked downward revision in both short-term and long-term expectations. The decline began to uncover the serious weaknesses in the domestic financial situation, in the position of some of the basic agricultural and raw-material markets, and in the international balance-of-payments position of various countries—particularly those exporting primary commodities and those that had borrowed heavily during the 1920's.

Prices continued to decline through 1930; the rise in automobile production proved short-lived; drought conditions added to the effect of collapsing agricultural prices in reducing farm incomes; nonresidential construction fell off sharply after the middle of 1930, and other private investment declined markedly also. Inventories decreased sharply after the third quarter, and the decline in equipment expenditures was accelerated in the latter half of the year. The freezing of bank loans associated with the fall in agricultural, real-estate, and security prices began to sap the public's confidence in the banking system, bank failures increasing particularly toward the end of the year. Business continued to contract despite a decline in interest rates and a general and substantial easing of credit conditions.

[26] This is clearly the case for the period since the Civil War, before which our records are too incomplete for a meaningful comparison. Cf. A. R. Eckler, "A Measure of the Severity of Depressions, 1873–1932," *Review of Economic Statistics*, vol. 15, May, 1933, pp. 75–81; and J. B. Hubbard, "Business Declines and Recoveries," in the same journal, vol. 18, February, 1936, pp. 16–23.

[27] Apparently a considerable amount of investment planned in 1929 was carried over into 1930. The railroad and electric power industries spent more for plant and equipment in 1930 than in 1929.

Contracting world trade uncovered oversupply positions in various agricultural and raw-material staples. As a result, prices of these commodities dropped sharply, in turn reducing the demand by agricultural and raw-material-producing countries for the products of industrial nations. By the middle of 1930, the depression was worldwide. In the face of sharply falling prices, the position of debtor countries—in central Europe, Latin America, and the Far East—began to be intolerable, and the situation was made worse by the cessation of international lending. A few countries had been forced to depreciate their currencies as early as December, 1929, and others followed in 1930, although the acute financial crisis did not come until 1931.[28] Passage of the American Smoot-Hawley Tariff in 1930 induced a wave of retaliation against American trade and set in motion an epidemic of restrictive trade measures by various countries which tended further to strangle world trade. These measures became more severe and discriminatory as the depression developed. Most of them were still in effect at the beginning of World War II.

In the early months of 1931, the American economy again seemed to be attempting to stage a recovery. Production in a number of lines expanded; the decline in retail trade leveled off; and there was some minor temporary improvement in private investment. It seemed as if the extreme liquidation of the preceding months might have been sufficient to induce some recovery. At this point, in the late spring of 1931, the international financial structure collapsed completely, and a financial crisis starting in Europe began a new wave of liquidation through the world and deepened the depression in the United States.

The crisis began in May with the failure of the Credit Anstalt, the largest bank in Austria, spread to Germany and the rest of central Europe, and then precipitated a "run" on the pound sterling that led to the suspension of the gold standard by England and the other countries of the "sterling area" in September. This in turn started a run on the dollar. Gold left this country in large volume in the wake of a flight of short-term capital; interest rates generally increased;

[28] For a picture of the steady abandonment of the gold standard in the 1930's, see the useful summary table in Margaret S. Gordon, *Barriers to World Trade*, 1941, pp. 40–41. This entire book provides an excellent survey of the breakdown of the world trading system in the thirties. For an excellent study of international monetary relations during the entire interwar period, see League of Nations, *International Currency Experience*, 1944.

and rediscounts with the reserve banks rose sharply as the crisis led to renewed currency hoarding and to additional pressure on a banking system already weakened by the impairment of its earning assets. The number of bank failures increased sharply, and business confidence deteriorated still further.

The European financial crisis of 1931 was the direct result of the weak foundations on which the world's financial structure had been built in the 1920's. The recovery of central Europe in the 1920's, as well as Germany's ability to pay reparations, rested on foreign loans, which declined rapidly after 1928. Also, as we saw at an earlier point, short-term capital movements played a much more important role in the 1920's than they did before World War I. Short-term capital could quickly move out of a country and, in so doing, put a violent pressure on its balance of payments. The spread of the crisis in the summer of 1931 was marked by this sort of panicky flight of capital as confidence in currency stability became impaired in one country after another. It was the flight of short-term capital from England that drove that country off the gold standard in 1931, though the steady decline in world prices and in her exports might well have forced her to take this step eventually.

The United States withstood the assault on her monetary standard, but the slight recovery in the first part of 1931 was wiped out. All components of aggregate demand began again a steady decline that continued until the summer of 1932. It was not until this phase of the depression that wage cuts became substantial, but the reductions were made piecemeal, and each cut merely led to expectations of further reductions in wages and prices.

This phase of the contraction was marked by extreme monetary and financial liquidation—frequent bank failures, currency hoarding, a sharp upsurge of business bankruptcies, severe inventory liquidation, another run on the dollar in the spring of 1932, sharply falling stock prices, and other signs of the almost complete destruction of business confidence.[29] The wave of liquidation subsided in the summer of 1932. Beginning in the third quarter, noticeable improvement began to be evident in the United States and other coun-

[29] The Federal Reserve system bought more than a billion dollars in government securities in March–June, 1932. This action enabled the banking system to withstand better the deflationary pressures mentioned in the text, but the easing policy came too late and had no effect in halting the deflationary forces then at work.

tries. The increase in production was particularly pronounced in textiles and was undoubtedly due in part to the need to replace depleted inventories after a prolonged period in which production had fallen much more than retail sales. Commodity prices began to show resistance to further declines, and even construction activity began to show signs of improvement.

This recovery in the United States was struck a severe blow at the beginning of 1933 by an outbreak of bank closings, beginning in the Middle West and spreading rapidly through the rest of the country. A final wave of hysteria undermined completely the foundation of confidence on which modern banking rests, and by the end of the first week of March all banks in the United States were closed. For many series, this month marks the low of the depression. For many others, the summer of 1932 was the low point. Most other countries date their recovery from 1932 rather than 1933. The behavior of American economic series in early 1933 is associated particularly with the final liquidation of the weak spots in our banking system —and to a minor degree, perhaps, with uncertainty arising out of the imminent change of administration in Washington.

In March, the Roosevelt administration came into power, at the climax of the banking crisis. Prompt and energetic action led to rapid reopening of solvent banks, and confidence in the banking system returned immediately. With this hurdle cleared, recovery in the United States unmistakably began.

THE CYCLE OF 1933–1938

It is impossible to place our finger precisely on those causes that finally brought the long downswing to an end. By the fall of 1932, a number of factors were beginning to make for recovery: depleted inventories, the need for some replacement of equipment, the elimination of weak firms and finally a sharp decline in the number of bankruptcies, the tendency of many prices to become stabilized, indications of improvement in some foreign countries and in some world primary markets, and so on. In addition, government measures such as creation of the Reconstruction Finance Corporation in 1932 alleviated some financial distress, and sizable government deficits in 1931 and 1932, even if unwillingly incurred, provided a mild stimulus. Wage rates had declined significantly in 1931–1932, but it is difficult to say whether the stimulating effect of the cost re-

ductions thus achieved offset the depressing effect of business expectations that wages might go still lower.

In any event, the banking crisis previously described overwhelmed the tendencies making for recovery at the end of 1932. A new, and this time sustained, recovery began with the successful efforts of the Roosevelt administration to reopen the banks. Business sentiment immediately improved and then became actively optimistic in response to the further measures of the government to raise prices and incomes.

The upswing that now followed was completely unlike any in our history, and its course aroused more controversy than that of any earlier business cycle.[30] The expansion was the longest on record up to that time—50 months elapsing between March, 1933, and the National Bureau's date of May, 1937, for the peak. Yet the recovery was weak and irregular, and at the peak in 1937 total output had barely recovered to the 1929 level. Output per capita was less than in 1929. Even during the boom of 1936–1937 there were in the neighborhood of seven to eight millions unemployed. The expansion took place in a setting of far more government intervention in economic affairs than this country had ever before experienced; it was accompanied by unprecedented peacetime government deficits and by a storm of controversy over far-reaching measures of social reform; and it occurred in a world setting of restrictive trade barriers and of mounting political tension that eventually ignited the flames of World War II.

CHARACTERISTICS OF THE UPSWING

The main features of the cycle of 1933–1938 are summarized in the chart at the beginning of this chapter, and additional information is supplied by Tables 28 and 29. We shall first look at the general characteristics of the expansion and then go on to consider the boom in 1936–1937 and the apparent causes of the downturn that followed.

Employment and output rose sharply in the second quarter of

[30] For more extended discussion of this cycle, see: League of Nations, *World Economic Survey* (annual); T. Wilson, *op. cit.*, chap. 18; S. H. Slichter, "The Downturn of 1937," *Review of Economic Statistics*, vol. 20, August, 1938, pp. 97–110; J. A. Schumpeter, *op. cit.*, vol. 2, pp. 1011 ff.; A. H. Hansen, *Full Recovery or Stagnation?*, 1938, chaps. 16–17; and K. D. Roose, *The Economics of Recession and Revival*, 1954.

1933 under the spur of renewed business confidence and anticipations of rising prices resulting from various government measures.[31] A speculative boom ensued which collapsed in the latter part of the year. Business activity recovered further in the first half of 1934, suffered some setback in the latter half, and then began a fairly steady advance that accelerated into a moderate boom in 1936 and the early months of 1937. Even at the peak, however, there was con-

TABLE 29. Gross National Product and its Components, 1929 and 1933–1939
(In billions of dollars) [a]

	1929	1933	1934	1935	1936	1937	1938	1939
Total GNP	104.4	56.0	65.0	72.5	82.7	90.8	85.2	91.1
Consumption expenditures	79.0	46.4	51.9	56.3	62.6	67.3	64.6	67.6
Durable goods	9.2	3.5	4.2	5.1	6.3	6.9	5.7	6.7
Nondurable	37.7	22.3	26.7	29.3	32.8	35.2	34.0	35.1
Services	32.1	20.7	21.0	21.9	23.5	25.1	25.0	25.8
Private domestic investment	16.2	1.4	2.9	6.3	8.4	11.7	6.7	9.3
Construction	8.7	1.4	1.7	2.3	3.3	4.4	4.0	4.8
Producers' durables	5.9	1.6	2.3	3.1	4.2	5.1	3.6	4.2
Inventory change	1.7	−1.6	−1.1	0.9	1.0	2.2	−0.9	0.4
Net foreign investment	0.8	0.2	0.4	−0.1	−0.1	0.1	1.1	0.9
Government expenditures	8.5	8.0	9.8	10.0	11.8	11.7	12.8	13.3
Federal	1.3	2.0	3.0	2.9	4.8	4.6	5.3	5.2
State and local	7.2	6.0	6.8	7.1	7.0	7.2	7.5	8.2

[a] From U.S. Department of Commerce, *U.S. Income and Output*, 1958, p. 118.

siderable unemployment. Wholesale prices rose sharply in 1933 and more moderately in 1934, and then remained relatively stable until a final spurt in the latter part of 1936 and the first part of 1937. Even in 1937, prices were considerably below the levels of the late 1920's.

The failure of private long-term investment to achieve anything approaching a full-employment level is the outstanding characteristic of the 1933–1937 upswing in the United States. The deficiency was particularly great in construction expenditures. The recovery in producers' durable goods was more satisfactory, though here

[31] Including emergency banking legislation and suspension of the gold standard, the Agricultural Adjustment and National Industrial Recovery Acts, measures aimed at providing financial relief to various types of debtors, etc.

again the 1929 level was not regained. Exports also lagged, and net foreign investment created little in the way of employment opportunities between 1933 and 1937.

It is clear that the marginal efficiency of capital was relatively low in the 1930's. As a result, recovery tended to lag, and large government deficits failed to prime the pump. The chief gains between 1933 and 1937 were in consumption rather than investment; and, compared to 1929, the output of consumers' and nondurable goods made a better recovery than did the production of durable and capital goods.

New private construction in 1937 was about 4 billion dollars less than in 1929 and about 5 billions less than in 1926. (See Table 25.) Construction was particularly laggard in the residential, public utility, and commercial fields. Residential building alone was in 1937 about 3.5 billions below the peak rate in 1925–1926. Although factory building was also at a low level, the deficiency here was not so great, either absolutely or relatively, as in the other areas mentioned.

Business displayed a notable unwillingness to undertake long-term investment projects, either in new directions or in those lines that had chiefly attracted investment funds in the 1920's. The fact that equipment expenditures made a better showing than business construction suggests that firms were willing to make capital expenditures only as necessary to replace and modernize equipment and to meet relatively minor changes in demand. In manufacturing, sufficient plant capacity had apparently been built during the 1920's to satisfy the demand for products from existing industries, and the demand for capital from new or young industries was not very great. The chief difficulties, however, seem to have been in fields other than manufacturing. Residential construction had obviously been overdone in the 1920's; and the normal corrective forces, retarded by a decline in the rate of population growth, had not yet generated the upper levels of a new building cycle. Similar factors were at work to hold down the amount of commercial building (stores, hotels, etc.). Declining rates of growth, calling into play the long-run working of the acceleration principle, restricted investment in some of the industries that had expanded rapidly during the 1920's—the public utilities, for example, and the types of activity that depended on the production and use of automobiles.

In addition to these underlying influences, the business fears and

antagonism created by various government measures undoubtedly tended to retard investment in some directions, and it has been frequently argued that the New Deal's regulation of the capital markets and tax legislation interfered with the flow of savings into particularly the more risky types of long-term investment. We can safely dismiss the argument that New Deal policies were the *sole* factor retarding investment; but it is highly probable that they did play some, and perhaps an important, restrictive role. The most serious effect may have been on public-utility investment.

TABLE 30. Excess of Federal Government Expenditures over Receipts, Including and Excluding Social Security Accounts, 1933–1939
(In millions of dollars) [a]

Year	(1) Federal Deficit, Excluding Social Security Accounts	(2) Surplus on Social Security Account	(3) Net Deficit
1933	1,368	58	1,310
1934	2,899	49	2,850
1935	2,600	62	2,538
1936	3,756	281	3,475
1937	1,628	1,452	176
1938	3,044	1,084	1,960
1939	3,358	1,145	2,213

[a] Derived from Tables 5 and 10 of the *National Income Supplement, Survey of Current Business*, July, 1947. Column (1) is computed by adding columns (2) and (3). These figures do not precisely correspond with the data shown in the various Treasury statements.

The "deflationary gap" left by private investment was partly filled by government deficits. The effect of the federal budget on the economy is suggested by the figures in Table 30. The federal deficit, in response to increased expenditures on relief, public works, agricultural benefits, and so on, rose from 1.4 billions in 1933 to 3.8 billions in 1936. The deficit in 1936 was swollen by the large soldiers' bonus paid in that year. In addition to the ordinary budget, we must consider the effect of the government's social security transactions, which were relatively unimportant until 1937, when the full effect of the large tax payments required by the Social Security Act of 1935 began to be felt. The excess of these receipts

over benefit payments was a major offset against the ordinary deficits incurred by the government from 1937 on.

On the monetary side, the 1933–1937 upswing was characterized by exceptionally easy credit conditions, despite which both commercial loans by the banks and the volume of new security issues remained abnormally low. Heavy gold imports added to the excess reserves of member banks, which reached a peak of about three billion dollars at the end of 1935. Member bank rediscounting at the reserve banks practically ceased after 1933. Commercial loans failed to show any increase until 1936 and then remained far below the level of the 1920's. Bank deposits rose, partly as a result of gold imports but chiefly through purchases of government bonds by the banks. Bonds rather than business loans became the chief earning assets of the banks.

The lack of private demand for credit and the extreme liquidity of the banking system led to a marked decline in all types of interest rates in the face of expanding business activity. While firms took advantage of the low interest rates to refund outstanding security issues bearing higher interest coupons, the volume of corporate issues to raise new capital was extremely low. Stock issues were particularly restricted during the 1930's.

Profits were relatively low throughout the upswing of the 1930's and were lower in 1936–1937 than in 1928–1929. This was true both before and after payment of taxes. On the other hand, under the stimulus of favorable legislation and the spread of collective bargaining, wage rates rose until in 1937 average hourly earnings of industrial workers were higher than in 1929. The relative rise in money wage rates between 1933 and 1937 was greater than that either in the cost of living or in the wholesale prices of finished products. The rise in wage rates was only partly offset by increases in labor productivity. Unit labor costs in manufacturing rose rather sharply between 1933 and 1934 and again between 1936 and 1937. As a result of these tendencies operating to increase wages, to some extent at the expense of profits,[32] wage and salary income rose as

[32] According to Spurgeon Bell's data, the ratio of unit labor costs in manufacturing to prices of finished manufactured goods was higher in 1937 than in 1929, though this was not true in 1936. (*Productivity, Wages, and National Income*, 1940, p. 270.) However, more recent data in the *Handbook of Labor Statistics*, 1947 ed., suggest that, if anything, unit labor costs in 1937 were slightly lower relative to manufactured-goods prices than in 1929.

fast as the GNP and somewhat faster than total income payments or disposable income. Despite this indication of a shift toward lower incomes with a higher propensity to consume, personal savings in 1936 and 1937 were about as large as in 1929. Corporate saving, however, was far smaller.

The factors thus far discussed that impeded recovery in the United States were apparently absent in most other countries. Nearly everywhere the expansion following 1932 was more vigorous than in the United States, though the advance was far from uniform. The main exceptions were those countries, of which France was the most important, that maintained a deflationary policy after 1932 in an attempt to remain on the gold standard at pre-depression parities. By 1937 the world index of manufacturing production (excluding Russia) was above the 1929 level. This was true of the index of capital goods as well as that of consumers' goods.

The substantial advance in production and national income in the rest of the world occurred despite the continuance of international trade restrictions on a far greater scale than in the 1920's. Currency management and restrictions on imports permitted various countries to adopt measures to stimulate domestic expansion; and the process was aided, to different degrees in different countries, by public works programs, rearmament, and continued industrialization.

The combination of domestic and foreign developments had a marked effect on the American balance of payments. Neither exports nor imports recovered to the level of the late 1920's. The volume of exports was restricted by the failure of agricultural exports to revive, and imports rose only as rapidly as did domestic activity. The net effect was to reduce our "favorable balance" on current account, including invisible items, until it became negative in the years 1935–1937. On the other hand, there was a tremendous inflow of capital, both short- and long-term. This capital movement was responsible for the huge import of gold from 1934 on and reflected both the failure of the United States to resume foreign lending and the flight of capital from Europe.

THE BOOM AND TURNING POINT

Beginning in the spring of 1936, hesitant recovery gave way to rapid expansion. Industrial production increased by nearly 30 per-

cent between March, 1936, and March, 1937, and then moved virtually horizontally through August. Prices rose sharply in the latter part of 1936 and the first few months of 1937. The increase was particularly marked in farm and raw-material prices. At the same time commercial loans began to expand rapidly; there was a general and pronounced rise in inventories; and other characteristics of a speculative boom began to be evident. Wage rates rose sharply in late 1936 and early 1937, other costs increased, and some bottlenecks began to appear, though total unemployment remained substantial. In the middle of 1936, payment of the veterans' bonus increased incomes and added to the upward movement.

The peak was reached about May, 1937, but a sharp decline did not begin until the autumn. The "turning point zone" or critical period lasted roughly from about March through August. The rise in raw-material, farm, and stock prices came to an end in March, and industrial production ceased to advance after that month—though no significant decline occurred till September. A number of other series also show an approximate plateau during a good part of the second and third quarters, including retail sales. Business expectations continued generally to be favorable, but falling prices, large inventories, and other elements of uncertainty led businessmen to exercise caution by curtailing orders for new goods. Beginning about September, production and employment began to fall as old orders were filled and new orders failed to appear, and an extremely sharp downswing got under way which lasted until about June, 1938. (Compare Figure 39.) The recession was primarily an American phenomenon. Production in some other countries fell moderately, but in others expansion continued with little or no interruption.

The question is immediately raised: Why was the recovery cut short when business activity was still considerably below the full-employment level? Before attempting to answer this question, it is necessary to review several developments that occurred during the boom and the ensuing critical period.

The Federal Reserve authorities had for some months been uneasy over the inflationary possibilities inherent in the large excess reserves of member banks, and these fears were accentuated by the beginnings of the speculative boom. Hence, in August, 1936, reserve requirements were increased by 50 percent, and they were raised a

further 33⅓ percent in two steps between March and May, 1937. The total effect was to double reserve requirements (the maximum increase permitted by law) and to reduce excess reserves from about three billions to less than a billion. In December, 1936, the Treasury reinforced this action by announcement of a program to "sterilize" gold imports—i.e., to prevent them from adding to bank reserves. These restrictive measures led to large-scale selling of bonds by the banks, particularly in the first half of 1937. As a consequence, there was some increase in bond yields, which may have had an unfavorable effect on the long-term capital market. Commercial loans and short-term interest rates (except on government securities) were not appreciably affected. The consensus among authorities is that monetary policy was not a major cause of the downswing, though it probably had some effect, especially through psychological channels.[33]

The reduction in the federal deficit in 1937, the second development to be noted, was more clearly deflationary. As noted in Table 30, the ordinary deficit was reduced from 3.8 billions in 1936 to 1.6 in 1937, and the latter item was almost completely offset by the influx of social security taxes. As a result, including social security funds, the federal government reduced its net contribution to aggregate demand from 3.5 billions in 1936, when the deficit was swollen by the bonus payment, to virtually zero in 1937. This deflationary force was particularly in operation from about March on; i.e., it occurred in the critical period when business was already hesitating for other reasons. As a result, the decline in the deficit was not offset by an increase in private investment. Indeed, by tending to depress incomes and consumption, it probably tended to reduce private investment also, since, as we saw earlier, the latter was closely geared to short-term expectations.

In an important sense, federal fiscal policy was exceptionally blundering in 1936–1937. The large bonus payment added to the boom in 1936, and the sudden reduction in the deficit in 1937 helped to tip the balance at a time that business was hesitating.

It is probable that wage and labor developments also played

[33] Kenneth Roose has concluded that "Federal Reserve policy cannot be cleared of important responsibility in the recession" because it made more difficult the flotation of any except the highest grade securities and unfavorably affected business expectations. *Op. cit.*, p. 117.

a role of some importance in 1936–1937. The organizing drives of the C.I.O. and union pressure generally led to a sharp increase in wage rates from the end of 1936 to the middle of 1937, and the number of strikes increased rapidly in the first few months of 1937. Higher wages led to anticipations of higher prices, which at first tended to expand advance ordering by business. During the critical period from March on, however, the continuance of wage increases and the rise in some important material costs tended to cut into profits and to reduce profit expectations, at least in some lines. At the same time, increased building costs had an unfavorable effect on the volume of construction. Under more favorable conditions, the increase in costs could have been absorbed by an expanding demand fed by a rising volume of private investment. In 1937 the depressing effect of the rise in costs was not offset by continued expansion of demand.

DOWNSWING AND RECOVERY

After September an extremely sharp decline in business activity took place. Industrial production declined by 30 percent in the eight months between September, 1937, and May, 1938. Prices, particularly of farm products, fell rapidly, and so did stock prices. (Compare Figure 39.) Unemployment rose by roughly five millions. The decline in output considerably exceeded the reduction in retail sales and consumption expenditures, with the result that inventories were reduced rapidly.

Production of nondurable goods ceased to decline after the end of 1937, and expectations in these lines gradually improved. Output of durable goods continued to fall until June, 1938, but at a decreasing rate after the beginning of the year. Nonresidential construction held up relatively well; and residential building, which had shown a significant decline during the latter part of 1937, rose rapidly after January, 1938. In April, the federal government announced a new "spend-lend" program, which undoubtedly improved business expectations even before incomes were enlarged by increased government spending from June on. Exports held up, and a decline in imports increased the export balance. Credit conditions eased rapidly. Reserve requirements were reduced in April; the Treasury stopped sterilizing gold and began to spend the pro-

ceeds of the gold previously sterilized; and total and excess bank reserves rose throughout 1938.

The low point in business was reached about May or June. The turning point may be ascribed to the normal corrective forces operating in a minor depression, particularly liquidation of inventories, plus the stimulating effect of renewed large government deficits. These factors caused short-term expectations to improve. No serious maladjustments or financial weaknesses developed to delay recovery, and there was no fundamental change in long-term expectations. The latter were relatively poor in 1937—this is why the economy was so vulnerable then—and they were not much better in 1938, except that the passage of time steadily accumulated more replacement needs and gradually reduced the excess capacity inherited from the 1920's.

We may now turn back and try to summarize the causes of the downswing of 1937–1938. The setting was that of a minor cycle against a background of a deficiency of long-term investment. Short-term expectations were inflated by the speculative boom in 1936, and the increase in inventories and the absence of long-term investment incentives made business highly vulnerable to unfavorable developments that might affect short-period expectations. The fall in commodity and stock prices in the spring of 1937 had just this effect, and so, to some extent, did the action of the monetary authorities. Labor difficulties and the increase in costs also tended to depress profit expectations. A leveling off in retail sales, partly occasioned by some consumers' resistance to rising prices, made the economy all the more vulnerable in view of the previous accumulation of inventories. In these circumstances, given the general unsatisfactory level of profits in many lines and the unwillingness or inability of business to invest in long-term projects on any considerable scale, the sharp decline in the government deficit probably provided the factor that tipped the balance. The extent to which private investment was geared to short-term expectations largely explains the sharpness of the decline.

THE SITUATION IN 1939

Business recovered rapidly in the latter half of 1938. Minor hesitation in the first half of 1939 was followed by a new upsurge

after the outbreak of the war in Europe, which marks the end of the period to be discussed in this chapter. By the end of 1939, business in the United States had reached or exceeded the peak attained in 1937. For the year as a whole, the GNP and its various components were approximately the same as in 1937.[34] (See Table 29.) Internationally, the business scene in 1938–1939 was dominated by preparation for and the final outbreak of war, and these developments are reflected in the high level of exports from the United States and in the tremendous influx of gold and short-term capital into this country.

A LOOK BACK

We are now ready to try to summarize the causes of the Great Depression. Two questions need to be answered. First, what were the factors immediately responsible for the downturn in 1929? Second, how do we account for the length and severity of the downswing that followed? The second question is related to a third, dealing with the cycle of 1933–1938: What caused the recovery in the United States after 1933 to be so weak and halting and to stop so far short of full employment?

The immediate causes of the 1929 turning point have already been suggested and may be summarized as follows:

1. There was a weakening of short-term expectations associated with (a) the development of buyers' markets in particular lines (horizontal maladjustments) and (b) concern over the stock-market boom. Oversupply in the automobile industry was particularly important in this connection.
2. Deflationary pressure accumulated as a result of the decline in residential building which had been going on since early 1928. This was offset only as long as other forms of investment were increasing.
3. Most important, the abnormally high level of investment in 1928–1929—on top of the substantial investment in the several preceding years—was beginning to create conditions of overcapacity in particular industries. This was one of the causes of the weakening of short-term expectations previously mentioned, but, more important, it led to a change in long-term expectations.

[34] The chief differences were the absence of marked inventory accumulation, larger government expenditures, and an increase in foreign investment.

4. The developments mentioned were sufficient to begin the downswing. Then in October the stock-market crash provided the *coup de grâce*—depressing expectations, removing a cheap source of long-term capital, and reducing consumers' demand for luxuries and durable goods.[35]

The international factors—the cessation of international lending and the pressure of supply on the prices of important primary products—were, in a sense, an independent set of causes operating chiefly on other countries, at least so far as the immediate causes of the downturn are concerned. They were not important in bringing the boom in the United States to an end.

The situation in 1929 was quite different from that prevailing at the time of the earlier downturns in the 1920's. Whereas in 1921, 1924, and 1927 the maladjustments could be corrected by a brief curtailment of output and liquidation of inventories, in 1929 businessmen came to doubt the profitability of continuing to invest in new plant and equipment at the rate such investment was being made in 1928 and the first half of 1929. Speculative optimism and technology had inspired in the middle and late 1920's a rate of investment in particular lines that could not be indefinitely maintained, and the acceleration of investment expenditures in 1928–1929 aggravated this tendency. Unlike the earlier downturns in the 1920's, a downward shift in short-term expectations in 1929 involved also a fundamental change in long-term expectations.

It is more accurate to say that the downturn in 1929 was due primarily to "overinvestment" than to ascribe the difficulty to "underconsumption." True, overinvestment was in relation to the demand for final products. But it is difficult to conceive of any increase in *total* consumption that would have maintained investment in a number of areas at the rate that had been reached before the turning point. It is true that wages did not rise as rapidly as productivity and that the propensity to consume apparently fell somewhat in 1929; but it can scarcely be argued that a moderately higher level of consumption could have prevented for very long, if at all, a decline in investment in residential and commercial building, in the automobile and related industries, and in other areas that had been expanding most rapidly. There was overinvestment

[35] These factors have been stressed by Galbraith, *op. cit.*, pp. 191–193.

in the late 1920's in the sense that capacity in numerous lines had been expanding at a rate that could not be indefinitely maintained.

The chief immediate cause of the downturn, then, was probably the impact of "partial overinvestment" on business expectations. This, however, is not sufficient to account for the length or severity of the depression or for its international ramifications. What we have said in earlier sections suggests that the following factors were chiefly responsible for the magnitude of the catastrophe that occurred.

1. The exhaustion of investment opportunities resulting from (a) the working of the acceleration principle in industries approaching maturity and (b) the creation of considerable excess capacity, particularly in residential and commercial building.
2. The financial excesses of the 1920's, which at the same time led to too rapid a rate of real investment in some industries and created a superstructure of inflated capital values the collapse of which weakened the banking system and led both borrowers and lenders to take a pessimistic view of the feasibility of further investment.
3. The unwise lending policies of the commercial banks, which created "frozen assets" on such a scale as to undermine the public's confidence in the entire banking system.
4. International balance-of-payments difficulties arising out of (a) the decline in American foreign lending, (b) the erratic movement of short-term capital, and (c) the serious oversupply situation in world primary markets, including some of the principal products of American agriculture.

It was these weaknesses, particularly the last three, that continued and deepened the downswing from 1931 on. The combination of these factors made the depression more severe than the other "major depressions" that we have had since the Civil War.

There remains the question: Why was the recovery of the 1930's so slow and halting in the United States, and why did it stop so far short of full employment? We have seen that the trouble lay with the inducement to invest. Even with abnormally low interest rates, the economy was unable to generate a volume of investment high enough, given the propensity to consume, to raise aggregate demand to the full-employment level. What made the general propensity to

invest so low is a question that is still being vigorously debated.

One answer points to the reform measures of the Roosevelt administration. We have already expressed our own view that New Deal policies cannot be held completely responsible, though they undoubtedly did have a restrictive effect on long-term business expectations. There is no reason to believe that different government measures would have restored residential or commercial construction to the inflated levels of the 1920's, and the federal government can scarcely be blamed for the flattening out of growth curves in particular industries or for the "once-burned-twice-shy" attitude created in many investors' minds by the financial collapse after 1929.[36] But we may grant, without further analysis, that willingness to invest in long-term projects was impaired to some but an unknown extent by the way business reacted to the activities of the federal government during these years. Needless to say, this conclusion carries no implications regarding the social desirability, from one or another point of view, of the measures that were taken.

Although much work remains to be done on the interwar period before we can evaluate with reasonable accuracy the importance of the different influences operating on business incentives in the 1930's, this writer is at present inclined to believe that government policies were not the most important factor holding back investment in those years. There seems to have been a lack of underlying investment opportunities, apart from the depressing effects of government actions and attitudes.

One group of writers holds that a fundamental change in the character of the American economy had occurred by the 1930's, with the result that the level of investment could be expected to be normally deficient in the future, given the existing propensity to consume. According to Hansen, the leading adherent of this view, rapid population growth and the opening up of new territory in various parts of the world were responsible for perhaps half of the total net investment in the nineteenth century. Technology was

[36] Schumpeter offers a persuasive argument that New Deal policies helped to crystallize a "climate of opinion" unfavorable to business and that this largely explains the disappointing nature of the recovery of the 1930's. It was, according to him, also still the downswing of a long wave, but this should not have made the situation any worse than major upswings superimposed on declining phases of earlier long waves. Cf. *op. cit.*, vol. 2, pp. 1038–1050.

responsible for the remainder. In the 1930's, according to this view, we began to see the effect of the decline in population growth and in territorial expansion. Technology alone could not be expected to generate as high a rate of investment as all three stimuli working together; and in the 1930's even technology did not provide as strong a stimulus as it did in earlier decades, when the railroad or the automobile and electric power were expanding most rapidly.[37] Thus, in the 1930's the American economy was suffering from "secular stagnation" or "economic maturity"—a drying up of private investment opportunities. If the argument is granted, the solution is obvious. If aggregate demand is equal to the sum of $C + I + G$, and if I remains too low for full employment, the answer lies in increased government spending or measures to raise the propensity to consume (or both).

This is not the place to attempt a critical valuation of the secular-stagnation thesis, which has attracted much less attention in the inflationary years following World War II than it did in the 1930's.[38] There is unquestionably an important element of truth in the argument—particularly that part of it which stresses the effect of a declining rate of population growth on the opportunities for investment—for example, in residential building and in the further expansion of old industries. Even if technology prevents a secular decline in the level of investment, "underemployment equilibrium" may still be a danger. For as technology raises the full-employment level of real income, it is not sufficient that the level of investment remain constant. The volume of investment must rise to absorb an

[37] Cf. A. H. Hansen, *Fiscal Policy and Business Cycles,* 1941, chap. 17; also his *Full Recovery or Stagnation?* 1938, chaps. 16–18, and his and other testimony in Temporary National Economic Committee, *Hearings,* Part 9. More recently, in *Business Cycles and National Income,* Hansen does not deal with secular stagnation as a separate problem but merges it into his discussion of cycles, long waves, and the conditions necessary for steady growth. The resemblances between the 1930's and the 1870's and 1890's are emphasized more, the big difference being the decline in the rate of population growth in the 1930's. But he no longer assumes that this decline in population growth will automatically continue. He still believes that "there is the danger that we may not achieve, on a sustained basis, our growth potential." *Ibid.,* p. 488.

[38] The most vigorous attack on the secular-stagnation thesis has been made by George Terborgh in *The Bogey of Economic Maturity,* 1945. For a range of views on this subject, see the Twentieth Century Fund symposium, *Financing American Prosperity,* 1945, *passim.* For further discussion and bibliography, see Benjamin Higgins, "Concepts and Criteria of Secular Stagnation," in *Income, Employment, and Public Policy: Essays in Honor of Alvin H. Hansen,* 1948.

increased flow of saving. Else the propensity to consume must increase or government spending must absorb the increased savings not absorbed by an expansion in private investment.

While "underemployment equilibrium" may be a future danger, there are difficulties in accepting the secular-stagnation thesis as an explanation of the disappointing behavior of investment in the 1930's. First of all, looking to the future as well as the past, the development of new industries and the expansion of old ones *may* still provide the necessary investment incentives, particularly if prices can be adjusted to achieve the maximum stimulation of demand. So far as the 1930's are concerned, the secular-stagnation argument is weak because it deals with secular—i.e., gradually operating—forces. Less than a decade separated 1928–1929 and 1936–1937, scarcely enough time for long-run forces to create the differences that existed between the two periods.[39] It is not sufficient to reply that secular stagnation was also operating in the 1920's but was temporarily offset by a speculative boom and by the investment opportunities created by the spread of the automobile and electric power.

Our own view is that stagnation existed in the 1930's, but that it did not necessarily have secular significance. Investment opportunities were restricted then because they had been so thoroughly exploited in the 1920's and because the severity of the financial liquidation after 1929 led businessmen and investors to view with a jaundiced eye the opportunities that were available. We would add that, given such a situation, the relative inflexibility of some prices (for example, building costs) prevented investment from being as high as it might otherwise have been. And, as noted before, the reaction of business to New Deal policies made the situation still worse.

By the end of the 1930's a good deal of excess capacity had been liquidated; residential construction was showing encouraging signs of revival; and technology was creating new investment opportunities. Had the war not intervened, private investment might or might not have increased to the point where full employment would have been possible without major government intervention. We shall never know, of course. The next decade or so, if the world can remain at peace, may throw additional light on the subject.

[39] This point is emphasized by Schumpeter, *op. cit.*, vol. 2, p. 1036.

In the years since World War II, the combination of private investment and a high level of government expenditures has been sufficient to give us sustained prosperity interrupted by only very mild recessions—although, as we noted in an earlier chapter, the rate of growth through most of the 1950's was somewhat disappointing and the level of unemployment did not fall below 4 percent in the latter half of the decade.

CHAPTER 15

FROM WORLD WAR II TO KOREA

HITLER'S march into Poland touched off the second world war in a generation—and initiated a new set of business-cycle and growth influences that were to dominate the course of business activity in the United States and most other countries in the two decades that followed.[1]

1940–1960: AN OVERALL VIEW

The period to be discussed in this chapter begins with the German invasion of Poland in 1939 and ends with the invasion of South Korea in 1950. For purposes of analysis, in the United States these 11 years fall into three clearly defined periods: partial mobilization (1940–1941), the all-out war effort (1942–1945), and postwar readjustment and inflation (1946–1950). The first two periods were dominated by the tremendous rise in government spending and by the inflationary pressures that were thereby released. From 1946 to 1950, business activity was again geared to the behavior of private spending. Strong inflationary pressures were operating on the economy both during World War II and in the period that followed (at least until 1948).

In the next chapter we shall review the decade of the 1950's. The first part of this period was dominated by the Korean conflict, which brought a new upsurge of government expenditures and a renewed but brief spell of rapidly rising prices. Then followed, during 1954–

[1] For the period after World War II (to 1958) the best general source for the United States is Bert G. Hickman, *Growth and Stability of the Postwar Economy*, 1960. See also the annual reviews published in the *Economic Report of the President* and in the *Survey of Current Business*. For business-cycle developments in Europe after World War II, see the annual surveys of the Economic Commission for Europe, a United Nations agency.

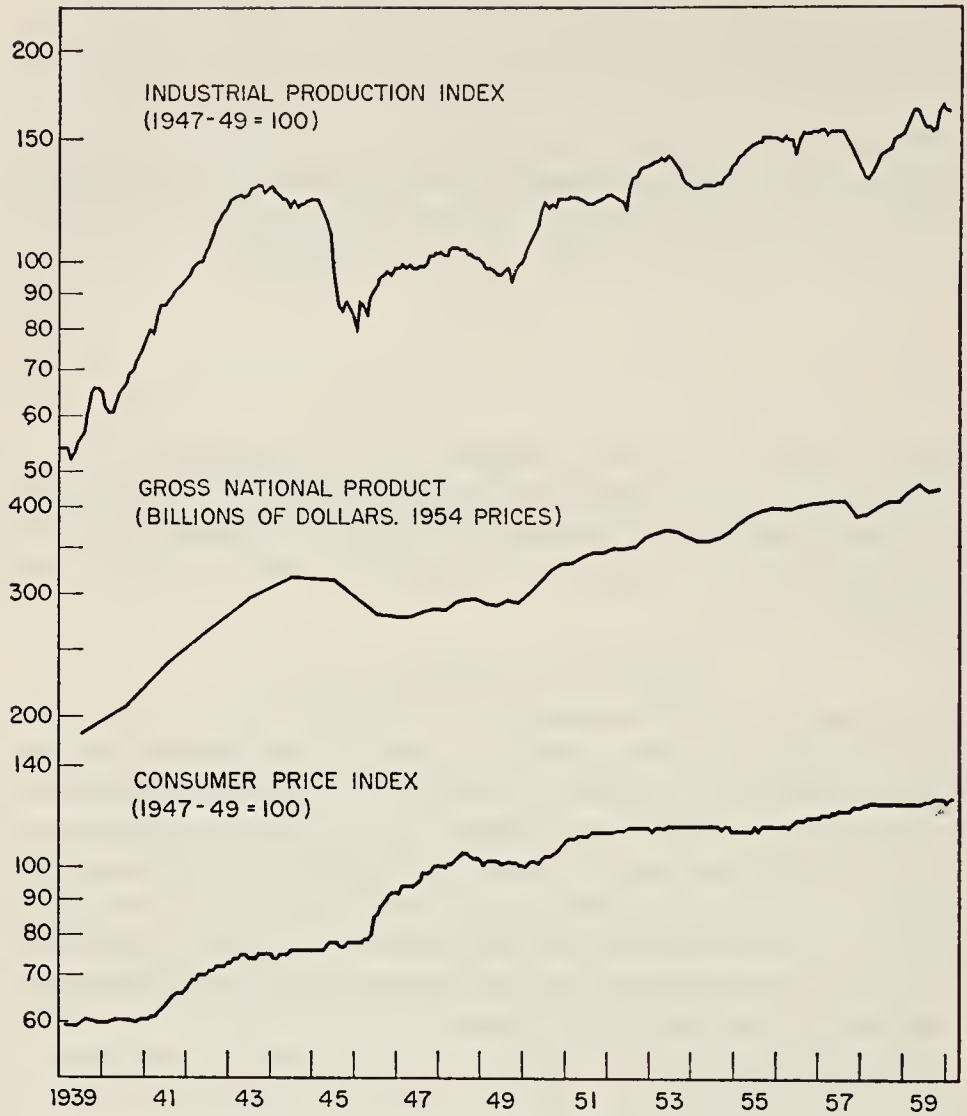


FIGURE 40. Industrial Production, Gross National Product, and Consumer Prices, 1939–1959.

1958, the first "normal" business cycle of the postwar period—normal in the sense that it was dominated neither by war nor by the immediate aftermath of war. Prices rose steadily during the latter half of the decade—although nowhere nearly as rapidly as in 1946–1948—and this "creeping inflation" was accompanied by a noticeable slowing down in the rate of growth and by a level of employment that was not fully satisfactory.

The movements of industrial production, prices, and GNP during

the 1940's and 1950's are presented in Figure 40. Here we can follow the wartime boom in output, the successive stages in the rise in prices over the two decades, and the course of the business cycle in the postwar years.

No *complete* major cycle occurred in the United States in the more than a quarter century that intervened between 1933 and 1960. The decline of 1937–1938 was moderately severe but very brief. Since the war boom began in 1940, the United States has experienced nothing worse than short and very mild recessions. Important structural changes were built into the economy that greatly increased its resistance to severe contractions (see pages 215–217), and a variety of short-run and secular forces have operated during the past 20 years to cause a fairly steady rise in aggregate demand and output.[2] We shall examine these forces in detail in this and the following chapter, but we can list the more important ones now: the pent-up consumer and investment demands and the excess liquidity that resulted from World War II, the upsurge in population growth, the redistribution of population (particularly the vast movement to the suburbs), the protracted housing boom that resulted not only from the pent-up demands and population changes already mentioned but also from the greater availability of mortgage credit, a boom in commercial building in the 1950's, a new scientific-technological "revolution," the impact of the cold war on federal government expenditures, and the sharp rise in nonmilitary public expenditures (for roads, schools, urban redevelopment, etc.) that inevitably went with the growth and shift in population and the general increase in income.

During the 20 years following 1938, the United States experienced four complete business cycles. The dates given by the National Bureau of Economic Research for these four cycles are as follows:[3]

Trough	*Peak*	*Trough*
June, 1938	February, 1945	October, 1945
October, 1945	November, 1948	October, 1949
October, 1949	July, 1953	August, 1954
August, 1954	July, 1957	April, 1958

Each of the postwar cyclical expansions listed above was longer than the prewar average, and each of the recessions was shorter than the average cyclical contraction before World War II. Thus, while

[2] For a more detailed treatment, see Hickman, *op. cit.*, chap. 8.

[3] See Table 22, p. 251, above.

we have continued to have business cycles since World War II, months of business expansion have outnumbered those of contraction by a very substantial margin—actually in the ratio of about 3.5 to 1.

Let us now go back and examine the decade of the 1940's in some detail. This is the decade of World War II and its aftermath.

THE DEFENSE PERIOD, 1940–1941

World War II did not begin to have its full effect on the American economy until after the German invasion of western Europe in the spring of 1940. The outbreak of war in September, 1939, had led to a spurt in production and prices, which reflected in large part accumulation of inventories in anticipation of higher prices as well as a substantial increase in exports. In the early months of 1940, this spurt in business activity subsided. Aggregate demand and the GNP continued to expand during the first half of 1940, but not so rapidly as during the latter part of 1939.

All this was radically changed by Hitler's triumphant march to the Atlantic and the consequent decision of the United States to arm itself as rapidly as possible. The American defense program dates from June, 1940. From this point on, the rapid acceleration of military spending injected a powerful new stimulus into the American economy.

BEHAVIOR OF GOVERNMENT AND PRIVATE SPENDING

The factors operating to expand business activity during 1940–1941 can be traced through the quarterly data on the components of the GNP in Figure 5 on page 42. While defense appropriations increased rapidly after May, 1940, the expansion of military orders did not result in a marked increase of government spending until the fourth quarter of the year. From then on, the rise in government outlays was very rapid.

Working through the familiar multiplier effect, the increased government expenditures led to a rise in incomes and in consumers' spending. Private investment expanded rapidly also. The demand for military goods called for new specialized facilities, particularly machinery and equipment. The rapid influx of workers into the war-production centers created a demand for new housing. Increased incomes and sales of civilian goods called for new produc-

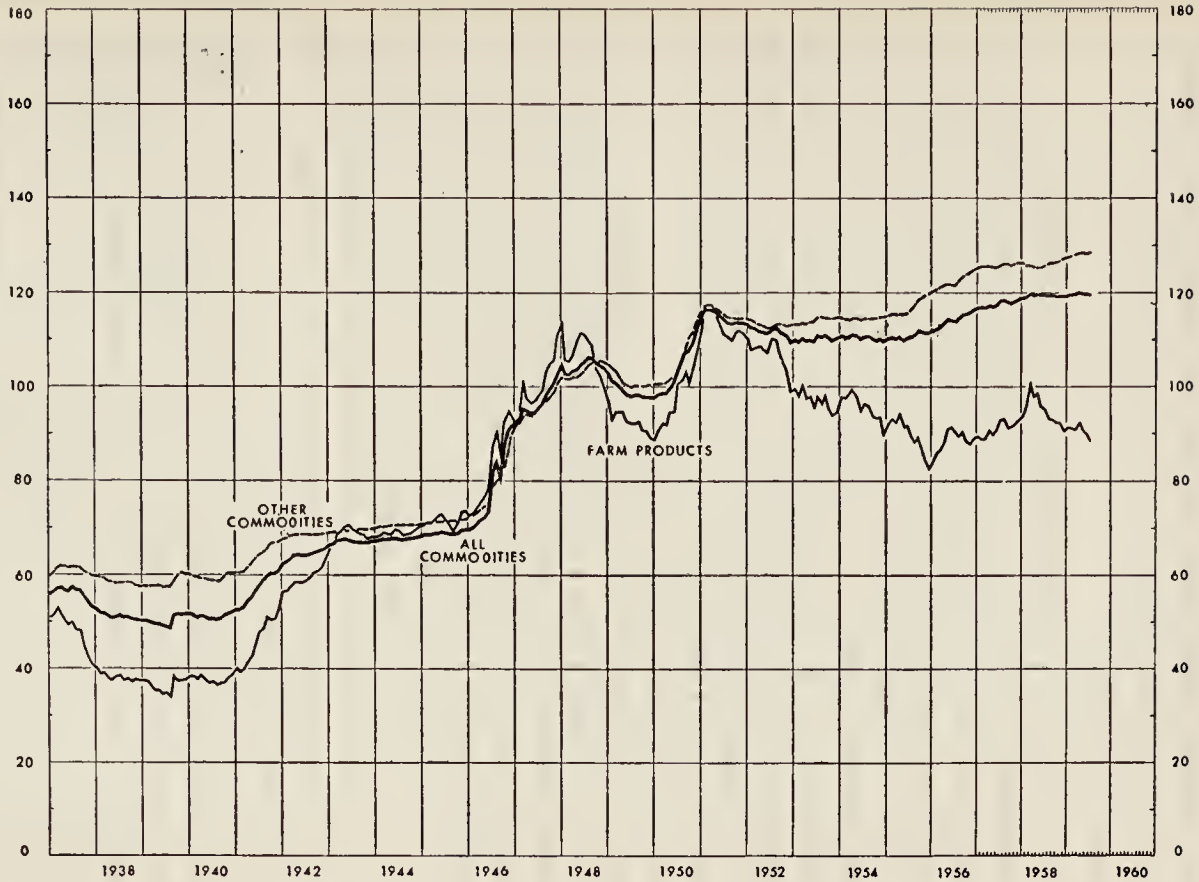


FIGURE 41. Wholesale Commodity Prices, 1937–1959 (1947–1949 = 100).

"Other commodities" are all except farm products and foods. Data are from Bureau of Labor Statistics. Reproduced from *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics.*

tive facilities, and fears of later shortages led business firms to accelerate their plans for expansion, modernization, and replacement. At the same time, with prices rising and the growing threat of shortages, business firms ordered ahead and accumulated inventories as rapidly as they could.

PRODUCTION AND PRICES

The effect of these powerful new stimuli on industrial production can be seen in Figure 40 on page 452. The Federal Reserve index of industrial production began a spectacular rise that continued almost without interruption until the end of 1943. Commodity prices began to increase toward the end of 1940, and the rise was quite sharp during 1941. Farm prices, in particular, shot up rapidly (Figure 41). It is interesting to note that much the largest part of the increase in the wholesale price index during the decade of the 1940's occurred either before Pearl Harbor or after the end of the war. During 1942–1945, price control succeeded in holding the rise in prices fairly well in check.

Naturally, employment rose rapidly during 1940–1941, and unemployment declined. Even at the end of 1941, however, the American economy had not yet achieved a position of full employment. Total unemployment in December, 1941, was estimated at 3.8 million, or about 7 percent of the total civilian labor force.[4] Unemployment did not fall below the two-million mark until the second half of 1942. Even so, labor shortages in some occupations and industries and in some sections of the country began to be felt even in 1941. Wage rates began to rise and continued to rise rapidly until they were finally stabilized in 1943.

When aggregate demand increases as rapidly as it did in 1940–1941, prices and wages are likely to rise considerably even before labor and other resources are fully employed. Short-period supply curves may be quite inelastic, particularly when sellers anticipate that demand, costs, and prices will continue to rise. The supply of agricultural products is relatively fixed in the short period, and farm prices react sensitively to upward shifts in demand curves. We have seen that there was a rapid accumulation of inventories in 1941, which had the effect of both increasing demand and reducing supply for various types of goods. We know also that, especially when labor

[4] *Survey of Current Business,* January, 1943, p. 5.

is well organized, wage increases may begin well before full employment is reached. This is particularly likely to be so when the increase in demand is concentrated in certain areas or industries. Local labor shortages lead to localized wage increases to attract labor from other industries or sections of the country, and these wage increases are likely to spread rapidly.[5]

FROM PEARL HARBOR TO RECONVERSION

THE EXPANSION IN SPENDING AND OUTPUT

The main outlines of the tremendous expansion in the American economy during the years 1942–1945 are summarized in Figure 40. The total GNP, in current prices, jumped from 126 billions in 1941 to 214 billions in 1945, when it was more than twice the figure for 1940. (See Figure 5 on page 42.) Government expenditures on goods and services rose to 96.5 billions in 1944, which was more than the entire gross national product during any year of the 1930's. In terms of constant (1954) prices, the GNP increased from 189 billions in 1939 to a peak of 318 billions in 1944. In the same constant prices, government purchases of goods and services were 30 billions in 1939, 48 billions in 1941, and 152 billions in 1944. Despite this increase in the share of total output taken by government, private consumption was able to expand moderately (in constant prices) between 1941 and 1945. While gross private domestic and foreign investment declined sharply, the total of private consumption and investment during the war years never fell below the level of 1939.

These figures suggest the great elasticity of the American productive system at the time of Pearl Harbor, despite the rise in output that had occurred during the two preceding years. The great increase in war production was secured primarily by expanding total output and, to a minor extent, by reducing private investment. Total consumption did not decline at all. However, neither did it rise as rapidly as disposable incomes. And, of course, some types of con-

[5] We can restate all this in terms of the aggregate-supply analysis presented in Chapters 2 and 4. The aggregate supply curve may bend upward even before full employment is reached. (See, for example, Figure 9 or 10.) In addition, aggregate demand and aggregate supply are not independent. Particularly if aggregate demand rises rapidly, the aggregate supply curve itself is likely to tilt upward under the pressure of wage increases that begin even before full employment is reached.

sumers' goods did suffer a reduction in supply or were not available at all. Thus, despite the maintenance of consumption, a backlog of unsatisfied or "pent-up" consumers' demand accumulated all during the war. This was to have an important effect on the course of developments after the war was over.

The expansion of output during the war years was heavily concentrated in the manufacturing industries, and particularly in the production of durable goods. This explains why the index of industrial production nearly doubled during the war, while the total GNP in constant prices rose by only about 60 percent. The production of durable goods nearly trebled between mid-1940 and the end of 1943, after which it began to decline—well before the end of the war. The rate of output needed to equip the armed services and to fill the pipelines was greater than that needed for replacement and further expansion in the later months of war.

Toward the middle of 1942 the United States reached a position of full employment for the first time since 1929. Even before, labor shortages were becoming acute in some areas and occupations. From 1942 on, the expansion of the armed forces was greater than the increase in the total labor supply, with the result that the *civilian* labor force declined. Unemployment averaged much less than the normal frictional amount during the years 1943–1945.

FINANCING GOVERNMENT EXPENDITURES

The federal deficit mounted rapidly as tax receipts lagged behind the rise in expenditures. The government financed less than half its total expenditures by taxation during the war years. For the entire period from mid-1939 to June, 1946, 45 percent of the total expenditures was met by taxes. About 35 percent was borrowed from nonbank investors and did not directly add to the money supply; about 20 percent came from the banking system and gave rise to a corresponding increase in the supply of money in the hands of the public.[6]

The gross public debt rose from about 50 billion dollars at the end of 1940 to more than 275 billion dollars at the end of 1945. Something more than 20 billions of this increase represented merely the building up of Treasury deposits, which were used after the war to cancel a corresponding part of the outstanding debt. If we adjust

[6] H. C. Murphy, *The National Debt in War and Transition,* 1950, p. 258.

for this item, we get the results shown in Table 31. Between June, 1939, and June, 1946, the federal government borrowed a total of 215 billion dollars, of which 38 percent was provided by the banking system and the remainder by nonbank investors.

By the standard both of past wars and of the experience of other countries in World War II, this was a relatively good performance, although both Canada and the United Kingdom met a larger per-

TABLE 31. Net Amounts Borrowed by the Federal Government from Each Principal Investor Class, June 30, 1939–June 30, 1946 [a]

Investor Class	Billions of Dollars	Percentage of Total
Commercial banks	60.2	28
Federal Reserve Banks	21.2	10
Total banking system	81.4	38
Individuals	53.9	25
Insurance companies	19.1	9
Mutual savings banks	8.5	4
Other corps. and assocs.	22.8	10
State and local governments	6.1	3
U.S. govt. agencies and trust funds	23.2	11
Total nonbank investors	133.6	62
Net amount borrowed	215.0	100

[a] Reproduced by permission from p. 259 of *The National Debt in War and Transition*, by H. C. Murphy. Copyright, 1950, by McGraw-Hill Book Company, Inc.

centage of their expenditures by taxation than did the United States. Nonetheless, the (roughly) 200 billions of deficit financing created a huge inflationary potential which could not be held in check once the war was over.

The magnitude of this inflationary potential can be shown in various ways. Total adjusted deposits and currency in circulation more than doubled between 1939 and 1945.[7] The expansion in the money supply was much greater than the increase in private spending, with the result that the income velocity of money (adjusted

[7] Adjusted deposits are total demand and time deposits minus checks in process of collection and minus U.S. government and interbank deposits.

to exclude both government spending and the government's deposits) fell sharply. The public found its supply of cash rising rapidly relative to its spending. "Excess" cash was accumulating which firms and households would try to spend once direct controls were terminated and goods were again freely available.

In addition to the increase in the money supply, two other monetary effects, both inflationary, resulted from the government's deficit-financing program. First, the public's total supply of liquid assets was increased far beyond the expansion in the money supply. As is shown in Table 31, nonbank investors absorbed more than a hundred billion dollars in government bonds between 1939 and 1946. These bonds were readily convertible into cash, either by direct redemption (as in the case of savings bonds) or by sale in the market. Once the war was over, this greatly increased supply of "near-monies" made the public much more willing and able to spend than would otherwise have been the case. As we shall see later, consumption did rise much faster than disposable income after the war, and there is no doubt that the large accumulation of liquid assets in the hands of the public had a great deal to do with this.

MONETARY POLICY AND DEBT MANAGEMENT

The other inflationary result arose out of the obligation the Federal Reserve authorities assumed, and carried over into the postwar period, to support the government bond market. This meant that the Federal Reserve Banks could not engage freely in open-market operations to limit the supply of bank reserves and thereby the volume of bank lending. This was not a major issue during the war. The Federal Reserve Banks had to make available to the banking system the reserves required to meet the great expansion of currency in circulation and the rise in deposits that resulted from government bond sales to the banks. The trouble came after the war. Since the Federal Reserve had to maintain government bond prices, it could not sell government securities freely in order to reduce bank reserves or even to offset the effect of the large gold imports after the war. And it also had to support the bond market while insurance companies and other business firms liquidated part of their bondholdings in order to use the funds in ways that increased the volume of private spending. One result was that the commercial banks were able to expand their loans rapidly after the

war, secure in the knowledge that their large holdings of government securities were a highly liquid secondary reserve that could be turned into cash reserves as needed.[8]

By agreement with the Treasury, the monetary authorities maintained a fixed pattern of interest rates on the various maturities of government securities. With this pattern of rates fixed by intervention in the market to the extent necessary, the Federal Reserve authorities then proceeded to supply the banking system with the additional funds needed to meet the large drain of currency into circulation and to provide the additional reserves required by the steady expansion in bank deposits. This was done through open-market operations. Between Pearl Harbor and the end of 1945, the Federal Reserve Banks increased their holdings of government securities from about two to approximately twenty-four billion dollars, thus permitting total reserves to rise by no more than the amount required to back the expansion in deposits. *Excess* reserves declined steadily through 1943 and remained approximately constant thereafter.

THE CONTROL OF CONSUMPTION AND INVESTMENT

Both consumers' expenditures and private investment were far less during the war years than they would have been in the absence of government controls. Gross private domestic investment declined by about 70 percent between 1941 and 1943. During America's participation in the war, private capital expenditures were not large enough to take care of current wear and tear and obsolescence, although this deficiency was offset to some extent by new government-financed facilities which are not included in the figures for private investment. Private capital expenditures were held to a low level by a battery of direct controls. Under a system of priorities and allocations, scarce materials were not available for uses other than those considered essential to the war effort; nonessential building was prohibited; and the production of some types of goods was limited by direct order or forbidden outright.

On the basis of the prewar relationship of consumption to disposable incomes, all types of consumers' expenditures were abnormally low during the years 1942–1945. (See Figure 15 on page 99.)

[8] The issues thus created for monetary policy are discussed further in Chapter 19.

A number of factors contributed to holding down consumption. The mere fact that some goods were unavailable or of inferior quality led many consumers to save a larger part of their income. (Of course, to some extent the public merely diverted its expenditures to other goods that were available.) Price control also tended to hold down consumption and to increase personal savings. In the absence of price ceilings, the excess of demand over supply would have led to much higher prices, and consumers' expenditures would have risen correspondingly without any increase in the actual quantities of goods available. Thus the combination of scarcities and price control tended to hold down consumers' spending and to increase saving, although some income did spill over into black markets.

Price control needed the support of a rationing program, and a variety of consumers' goods, chiefly foods, were rationed. This again held down consumption and increased the proportion of consumers' incomes that was saved.

The government also acted to stimulate saving directly, through a vigorous campaign to sell bonds to nonbank investors. An integral part of this campaign was the savings bond designed for the small investor and the program for systematic purchase of these securities through payroll deductions.

So far, we have mentioned the main ways in which the propensity to consume disposable income was reduced during the war. In addition, of course, the government attempted to slow down the rise in disposable incomes by increasing taxes. Personal tax payments to the federal government rose from two billions in 1941 to about 20 billions in 1945. In 1943 the federal income tax was put on a pay-as-you-go basis, a major reform which was carried over into the postwar period.

WAGE AND PRICE CONTROLS

We have already commented on the sharp rise in wholesale prices in 1940–1941 and on the continued rise of retail prices through 1942. Wages rose steadily during 1941–1942 and were not effectively stabilized until 1943. With the inflationary gap creating a large excess of demand over supply in the private sector of the economy and with wages and prices pursuing each other upward in an inflationary spiral, all-embracing wage and price controls became necessary.

Price ceilings were imposed on a number of scarce commodities

during 1941, but general price control was not instituted until the spring of 1942. Most farm prices were not brought under effective control until 1943, until they had achieved parity or the other minimum levels specified by Congress.

Price control could not hope to be effective without stabilization of wages. Wage rates were finally stabilized by application of the "Little Steel Formula," which limited the rise in wage rates to 15 percent above the level of January, 1941, with adjustments permitted to remove inequities and to meet special needs. Despite the eventual stabilization of *wage rates,* average hourly earnings continued to rise as the result of the upgrading of workers to higher-paid jobs and an increasing amount of overtime. With longer hours, weekly earnings rose faster than hourly earnings. Despite higher taxes, the take-home pay of the average American factory worker rose significantly more during 1941–1945 than did the cost of living. Part of this gain, however, was due to the increase in the length of the work week. When the work week was again reduced to 40 hours after the war, labor sought to offset the consequent reduction in take-home pay through higher wage rates.

SUMMING UP

Price and wage control in World War II was surprisingly successful, particularly in view of the inflationary gap created by the government's deficit spending. There were many inequities; some black markets flourished; and perhaps some goods would have been in larger supply or of better quality in the absence of price control. On net balance, however, price and wage controls clearly aided the war effort. As long as it was politically impossible to meet the full cost of the war by taxes, price and wage controls prevented a sharp inflationary spiral that would have created grave political and social tensions and would have distributed the burden of the war effort in a highly inequitable manner. Unfortunately, the success was only temporary. The end of price control in 1946 let loose the inflationary pressures which had been bottled up, with the result on prices shown in Figures 40 and 41.

On net balance, despite the mistakes that anyone can point to, this country managed the economic side of the war quite well. We were able to put forth a prodigious productive effort, which resulted in a tremendous flow of armaments of all types. We were able

to do this while putting 12 million men and women into the armed services and at the same time maintaining the living standards of the civilian population. Resources were diverted promptly to war uses by a system of direct allocation controls. While deficit financing, particularly from the banks, created a large inflationary potential, prices were held in check and the real burden of the war was distributed reasonably equitably through direct controls involving rationing and price and wage controls—as well as through heavy, progressive taxation (which should, however, have been heavier than it was). As one economist has put it, this country did a good job of managing a "disequilibrium system" during the war.[9] An "equilibrium system" would have entailed a much heavier burden of taxation to cut down consumers' demand or a much higher level of prices to equate the inadequate supply of civilian goods to the inflated level of demand.

THE PROCESS OF RECONVERSION

Government expenditures reached their peak in the first quarter of 1945, and so did a number of other business indicators. Government war orders declined rapidly after V-E Day and precipitously after the Japanese surrender. Total government purchases of goods and services fell from an annual rate of nearly 100 billions in the first three months of 1945 to about 35 billions in the first quarter of 1946. (See Table 32.) Thus, in the short period of a year, the government reduced its contribution to aggregate demand by an amount equal to almost 30 percent of the total GNP at the beginning of the period.

It is not surprising that many observers doubted the ability of the economy to adjust itself to a deflationary force of such magnitude without a serious recession in business. In the summer of 1945 the belief was fairly widely held in Washington that unemployment would be a serious problem during the winter of 1945–1946, and a strong deflationary tendency was predicted.

THE MILDNESS OF THE RECESSION

These forecasts, of course, were completely wrong. Total GNP declined by considerably less than half the drop in government spend-

[9] J. K. Galbraith, "The Disequilibrium System," *American Economic Review*, vol. 37, June, 1947, pp. 287–302.

ing and had about regained its wartime peak by the end of 1946. (See Table 32.) Unemployment did not rise as high as three millions. Consumers' expenditures did not decline at all and, indeed, rose with startling rapidity after V-J Day. Beginning early in 1946, as controls were relaxed, prices began to rise at a disconcertingly

TABLE 32. Components of the Gross National Product for Selected Calendar Quarters, 1945–1946
(Seasonally adjusted annual rates; billions of dollars) [a]

	First Quarter, 1945	First Quarter, 1946	Fourth Quarter, 1946
Consumption expenditures	119.0	137.2	156.1
Gross private domestic investment	7.7	24.5	30.3
Net foreign investment	−2.7	2.6	4.8
Government expenditures [b]	98.6	35.4	29.2
Gross national product	222.6	199.7	220.4
National income	191.8	169.7	190.3
Less: Corporate profits	23.5	14.7	21.5
Other deductions [c]	5.8	6.9	5.8
Plus: Govt. transfer payments	3.5	12.0	9.7
Other additions [d]	8.5	10.2	10.4
Personal income	174.4	170.3	184.2
Less: Personal taxes	21.3	17.8	19.6
Disposable income	153.1	152.5	164.6
Consumption expenditures	119.0	137.6	156.1
Personal saving	34.1	15.3	8.5

[a] From *Survey of Current Business*, July, 1950, pp. 30–33. These figures have since been slightly revised, but the changes are not important and in no way alter the interpretation offered in the text.

[b] On goods and services, for all levels of government. Excludes transfer payments

[c] Chiefly social security contributions.

[d] Includes government interest payments, corporation dividends, and business transfer payments.

rapid rate. In 1946 as during the war, the problem was one of insufficient supplies of goods and excessive demand. Even with the decline in government spending, aggregate demand was sufficient to maintain full employment. By the middle of 1946, a vigorous inflationary boom was in full swing.

The question now before us is: Why was the "reconversion re-

cession" so brief and mild—in the face of such a tremendous decline in government spending and in the face of a decline of 35 percent in industrial production between February, 1945, and February, 1946? One way of answering this question is to examine in detail the behavior of the components of the GNP during this period. Table 32 provides a basis for doing this.

THE BEHAVIOR OF PRIVATE INVESTMENT

Despite the tremendous decline in government spending in the latter half of 1945 and the first part of 1946, both consumption and private investment rose rapidly. Between the first quarter of 1945 and the first quarter of 1946, private spending increased by about 40 billion dollars. As can be seen in Table 32, consumption, domestic investment, and foreign investment all contributed to this expansion. Let us look first at the forces operating on private investment.

The outstanding influence here, of course, was the existence of tremendous pent-up demands. Expenditures had to be made as quickly as possible to convert facilities to civilian production, to acquire inventories of raw materials and work-in-process, and to build up stocks of finished goods at all stages of distribution. Deferred replacement and modernization had to be undertaken. There was a desperate shortage of housing, and also a shortage of other types of buildings. In addition, capacity had to be expanded in many lines to meet the high level of demand that existed or was anticipated.

Businessmen were encouraged to go ahead by the quick lifting of wartime controls and by an efficient machinery for the handling of war-contract cancellations. As a result, industry was free to absorb labor and materials almost as quickly as they were released by the decline in government orders. Firms reconverting from war to civilian production retained most of their employees to aid in the reconversion process, which helps to explain why factory employment did not decline as much as factory output. In addition, workers laid off in some plants were quickly hired by other manufacturing plants and by the trade and service industries, which had been starved for manpower during the war, or else they were absorbed into the burgeoning construction industry. Another factor moderating the decline in employment was the shortening of the work week.

In addition to the expansion of domestic capital formation, Table 32 indicates that there was a large increase in net foreign investment, as privately financed exports spurted to meet the insistent demand of foreign countries for American goods. The rise in privately financed shipments offset in part the decline in Lend-Lease exports, which had been included in government expenditures in the GNP accounts.

THE SPURT IN CONSUMERS' EXPENDITURES

The increase in private domestic and foreign investment was only a modest offset against the much larger decline in government expenditures on goods and services. In view of this, the behavior of consumption in this period is quite remarkable. There was no multiplier effect leading to a decline in consumers' expenditures. On the contrary, consumption increased rapidly in the face of a decline in the GNP. Here lies the main part of the answer to the mildness of the reconversion recession.

There were two main reasons for this behavior of consumption. First, disposable incomes of individuals remained virtually constant during the year despite the decline in total GNP. Second, there were powerful forces operating to increase the ratio of consumption to disposable income and to reduce the volume of personal saving.

The stability of disposable income in the face of a decline in GNP is explained primarily by three facts, all of which are brought out in Table 32. First, there was a sharp rise in government transfer payments, which increased disposable income without affecting the GNP. This represented chiefly benefits paid to veterans as they were released from the armed services at a rapid rate. Second, and less important, a decline in personal taxes helped to support disposable income. The third factor had to do with the behavior of profits. Corporate profits fell by nine billion dollars between the first quarter of 1945 and the same period of 1946, but dividend payments did not decline. To this extent the decrease in GNP had no effect on disposable income. We shall encounter this same stabilizing behavior of profits and dividends in the later postwar recessions.

Now we have to ask: Why did the ratio of consumption to disposable incomes rise sharply in the latter part of 1945 and the first part of 1946? Here we get into the most important set of in-

fluences holding up demand during the reconversion period—and perhaps the most important cause of the inflated level of demand during the postwar boom that followed.

As we have already seen, consumers' expenditures were abnormally low during the war years; the propensity to consume was held down by shortages, rationing, price control, and the wartime savings campaign. The deficiency in consumption was most striking in the case of durable goods. By the end of the war there had accumulated a tremendous amount of pent-up consumers' demand, and an unprecedented amount of liquid assets to make this demand effective.

Once the war was over and rationing was brought to a quick end, consumers went on a buying spree that lasted well into the post-reconversion period. What surprised economists was the extent to which this pent-up demand made itself felt in the field of nondurable goods—with the result that expenditures on such goods rose rapidly to a level well above the prewar relationship to disposable income.[10] Expenditures on durables increased rapidly, also, but continued shortages limited the amount that could be spent. Returning veterans, supplied with accumulated savings and demobilization benefits, added to the swelling demand for consumers' goods. It was particularly the buying splurge in nondurable goods that made retail sales behave so favorably during the reconversion period, a development that added to the already optimistic expectations of businessmen. The continued buoyancy in the demand for nondurables and steadily rising expenditures on durable goods made consumers' demand an active inflationary force during the postwar boom that followed.

THE POSTWAR BOOM

The low point of the mild reconversion recession came in the last quarter of 1945. There followed a vigorous inflationary boom that carried the money value of the GNP well above the peak reached during the war. This boom was characterized by rapidly rising prices and wages, an unprecedented volume of private investment of all types, an unparalleled volume of exports, and a seem-

[10] See, for example, the scatter diagrams relating expenditures on durables and nondurables to disposable income in *Survey of Current Business*, March, 1959, p. 22.

ingly insatiable demand for consumers' goods. Signs of tapering off appeared in 1948, to be followed by a brief and mild recession that reached its low point in the latter half of 1949. A vigorous recovery got under way in the first six months of 1950, and by mid-1950 the level of output achieved in 1948 had been exceeded. At this point, the outbreak of hostilities in Korea unleashed a new set of inflationary forces.

COMPARISON WITH THE BOOM AFTER WORLD WAR I

The period 1946–1949, therefore, marks off one complete minor cycle, superimposed on a major inflationary boom that had not yet run its course by the time rising military expenditures again came to dominate the business scene in 1950–1951. The period 1946–1949 has certain resemblances to the postwar boom of 1919–1920, and also some even more striking differences. The mild recession of 1945–1946 was matched by the mild reconversion dip of 1918–1919. In both cases, a sharp inflationary boom ensued which had its basic causes in a great pent-up demand by consumers and businessmen, a swollen foreign demand for American goods, a greatly increased money supply and stock of liquid assets in the hands of the public, and an elastic credit supply which expanded rapidly as business firms sought to enlarge their inventories and plant and equipment. But here the resemblance largely ends. The boom following World War I was heavily speculative and closely geared to short-term expectations, and the backlog of pent-up demand and the accumulation of liquid assets were nowhere as great as after World War II. In addition, the expansion of 1919–1920 was subject to three deflationary shocks that were largely absent in the later period. Government spending continued at the wartime level well into 1919 and then was sharply curtailed. (In 1945–1946, this deflationary force was out of the way before the postwar boom began.) Secondly, the monetary authorities deliberately and substantially tightened credit in 1920, whereas monetary ease was the rule throughout the period 1946–1950. And finally, foreign demand weakened significantly in 1920, whereas after World War II the Marshall Plan and other forms of American aid helped to maintain exports for a much longer period.

As a result of these factors the 1919–1920 boom was quickly over; and, because of its short-term, speculative character, it was followed

by the quite severe depression of 1921. The outstanding characteristic of this episode was the extreme decline in prices that occurred. Nothing like this occurred after 1945. Though the boom was highly inflationary and prices rose substantially, there was little speculative accumulation of inventories. While bank credit expanded, business firms remained in a healthy and relatively liquid position. The pent-up demand for consumers' goods, the great demand for plant and equipment for replacement, modernization, and expansion, and the acute housing shortage, all backed by the great increase in the money supply and in the public's stock of liquid assets, created a set of expansionary forces that were largely independent of changes in short-term business expectations and other minor deflationary shocks. Essentially, this is why the underlying boom persisted so long, and why the minor deflationary forces operating in 1948–1949 led to only a mild and brief dip in business activity.[11]

GENERAL FEATURES OF THE BOOM

Perhaps the outstanding feature of the postwar boom was the rise in prices—which lifted wholesale prices by about 60 percent and retail prices by about 35 percent between the end of 1945 and the summer of 1948. Both wages and prices began to rise rapidly early in 1946, and the upward course of prices was accelerated after the abandonment of price controls in the latter half of that year. Protected by a steadily expanding aggregate demand and spurred by the sharp increase in the cost of living, organized labor successfully negotiated a series of wage increases which were quickly passed on in higher prices. Food prices, however, rose faster than industrial prices.

Because of the much higher price level, the money value of the GNP quickly surpassed its wartime peak, but the physical volume of production remained below the highest levels reached during the war. This is to be explained in large part by the return to a shorter working week and by the fact that labor productivity in many peacetime lines was lower than in the highly mechanized war industries. The economy operated at or above the full-employment

[11] It is also true that, largely because of the greatly increased strength of organized labor, wages and prices were more inflexible downward than they had been before World War II. This inflexibility showed itself particularly in the mild recessions in the 1950's.

level. Frequently, during the years 1946–1948, unemployment fell below the two-million mark, which may be taken as a crude but conservative estimate of the "frictional unemployment" going with a civilian labor force of 60 million workers.

The rise in aggregate demand during the postwar years reflected the behavior of private spending, both consumption and private investment. Although government expenditures on goods and services remained much higher than before the war, they averaged about the same fraction of the total GNP as in 1939. On a cash basis, the federal budget showed a substantial surplus during the period of most rapid expansion in private spending. There was also a significant decline in the size of the public debt between the end of 1945 and the end of 1948. However, bank loans expanded rapidly, so that, despite the decline in the commercial banks' holdings of government securities, the money supply in the hands of the public showed a net increase during 1946–1947.

BEHAVIOR OF CONSUMPTION

Consumers' expenditures were more than a passive influence during these postwar years. Consumption reacted, via the multiplier, to the rising volume of private investment, but, more important, it increased faster during 1946–1947 than could be explained by the rise in income. In short, a rise in the propensity to consume was an independent influence inflating the level of aggregate demand. The final tapering off in the expansion of consumers' demand in 1948 undoubtedly was an important factor in the mild business recession that occurred in 1949.

As we should expect, there was a heavy pent-up demand for consumers' durable goods, particularly automobiles, which had not been available during the war. Sales of consumers' durables rose as rapidly as expanding production permitted, but it was not until 1948 that consumers' expenditures on durable goods approximated the prewar relationship to disposable income. Even then, the backlog of deferred demand for some goods, particularly automobiles, had not been fully satisfied.

Consumers' demand for durables (as well as for housing and for nondurables) was supported by the large accumulation of liquid assets carried over from the war. Individuals also borrowed heavily, and the volume of installment credit outstanding increased by nearly 10 billion dollars between the end of World War II and the

outbreak of the Korean crisis. Even so, the ratio of such credit to personal income was not so high as during 1939–1941, though much higher than in 1929.

The rapid rise in expenditures on durable goods was not surprising. What was surprising was the level of demand for nondurable goods. We have already commented on the spurt in this type of expenditure in 1946. Retail sales of nondurables rose in 1946 to a level considerably above the prewar relationship to disposable income and continued at an abnormally high level during the several years following.

This is not the place to explore in detail all the possible reasons for this behavior of nondurable-goods consumption. Here again, the large volume of accumulated liquid assets probably played a role of some importance. The distribution of income had changed in favor of the low-income groups, and this tended to raise the overall propensity to consume. For a variety of reasons, the consumption of food per capita was higher in the postwar period than before the war, and the public did not readily reduce its consumption as prices rose. The percentage of disposable income spent on food was significantly higher during 1946–1947 than before the war.[12] Other factors also undoubtedly played a role in expanding consumption: the desire to spend freely after three years of shortages and rationing, the needs of veterans as they were reabsorbed into civilian life, and generally a free-spending psychology induced by high and rising incomes, full employment, and anticipation of still higher prices. The expansion in consumers' expenditures began to level off in 1948, and the volume of personal saving began to rise after the middle of 1947, suggesting that some of the forces operating to create an abnormal level of consumers' demand were perhaps beginning to lose their strength. We shall look further at the behavior of consumption and saving in the next section.[13]

PRIVATE INVESTMENT IN THE POSTWAR PERIOD

Perhaps at no time in our history were the forces tending to expand private investment as strong as they were after World War II.

[12] For an analysis of food consumption and expenditures in the immediate postwar period, see the *Survey of Current Business,* January, 1948, pp. 12–16.

[13] See also Irwin Friend, "Personal Saving in the Postwar Period," *Survey of Current Business,* September, 1949.

(See Figure 42 on page 475.) As a result, gross private domestic investment rose sharply between 1945 and 1946, from 10 to 28 billions, and then continued a further rapid expansion to a temporary peak in 1948. Even if we allow for the higher level of prices after World War II, private investment during 1946–1948 was considerably larger than in 1929.

American business emerged from World War II with a heavy pent-up demand for plant and equipment of all types. There had been little expansion or modernization of capacity during the 1930's, and stringent controls prevented much private investment during the war years. Against this background, business faced a demand for goods after the war which was much greater than anything that had been experienced before. Thus there were two driving reasons to invest in new plant and equipment: to satisfy the accumulated needs for replacement and modernization and to expand capacity to meet the new high level of demand.

As a result of these forces, expenditures on nonfarm plant and equipment exceeded 20 billion dollars in each of the years 1947–1950—more than double the level of 1929 and 1941, the two previous peak years. (Of course, allowance must be made for the higher level of prices after World War II.) The rise in plant and equipment expenditures was particularly marked in manufacturing and in electric power and gas.

By 1949, some evidence was beginning to accumulate that the most urgent part of the pent-up demand for plant and equipment had been met, although a substantial backlog of demand for replacement and modernization still existed. In some industries—for example, electric power—the need for further expansion of capacity continued to be urgent. It seems likely that, if the Korean crisis had not occurred, business expenditures on plant and equipment would have remained at a high level for several more years but with some tendency to recede gradually from the peak rate reached in 1948.[14]

The high level of residential building was another important factor in the postwar investment boom. An acute housing shortage

[14] For further analysis of the situation with respect to business long-term investment at the end of 1948, see the "Annual Economic Review" of the Council of Economic Advisers in *The Economic Report of the President*, January, 1949, pp. 17–18, 55–61. See also the *Survey of Current Business*, June, 1949, pp. 8–18.

existed at the end of the war, and residential construction mounted rapidly despite the rise in building costs. During 1948–1949, residential building amounted to about 10 billion dollars a year, and the figure jumped to 14 billions in 1950. These dollar figures are well above the peak rate reached in the building boom of the 1920's.

In addition to the shortage inherited from the war, a number of other factors operated to raise the demand for housing: a continued high marriage rate, the high level of incomes, the large volume of liquid assets, and the liberal terms under which mortgage credit could be secured. Here was a strong and largely autonomous stimulus to investment that provided powerful support to the postwar level of business activity and was largely immune to minor deflationary shocks.

During 1946–1948, builders concentrated on higher-priced single-family dwellings for home ownership. By 1948, supply was beginning to catch up with demand for this type of housing, suppliers of mortgage credit became somewhat more cautious, and a short decline in building activity resulted. This was enough to bring about some reductions in building costs and to stimulate builders to enter the lower-priced field. In addition, federal legislation made easier the financing of multifamily rental projects. As a result, residential construction began to rise very rapidly early in 1949, and the expansion continued into 1950 at an accelerated rate. This behavior of residential building is one of the main reasons for the mildness of the 1949 recession.

The accumulation of business inventories also helped to support the postwar investment boom. Our chief interest in investment in inventories, however, is in the year-to-year changes. In 1947 and again in 1949 the tendency toward inventory accumulation was reversed. (See Figure 42.) In 1947 other factors offset the decline in inventory investment, and no general recession in general business activity occurred. In 1949 the decline in inventory investment was larger; this was not offset by expansion in other types of investment; and a mild recession resulted.

Although there was a substantial net accumulation of inventories between V-J Day and the Korean crisis, nearly all of this was required to fill the pipelines between producer and consumer. Despite the rise in prices, there was little speculative accumulation of

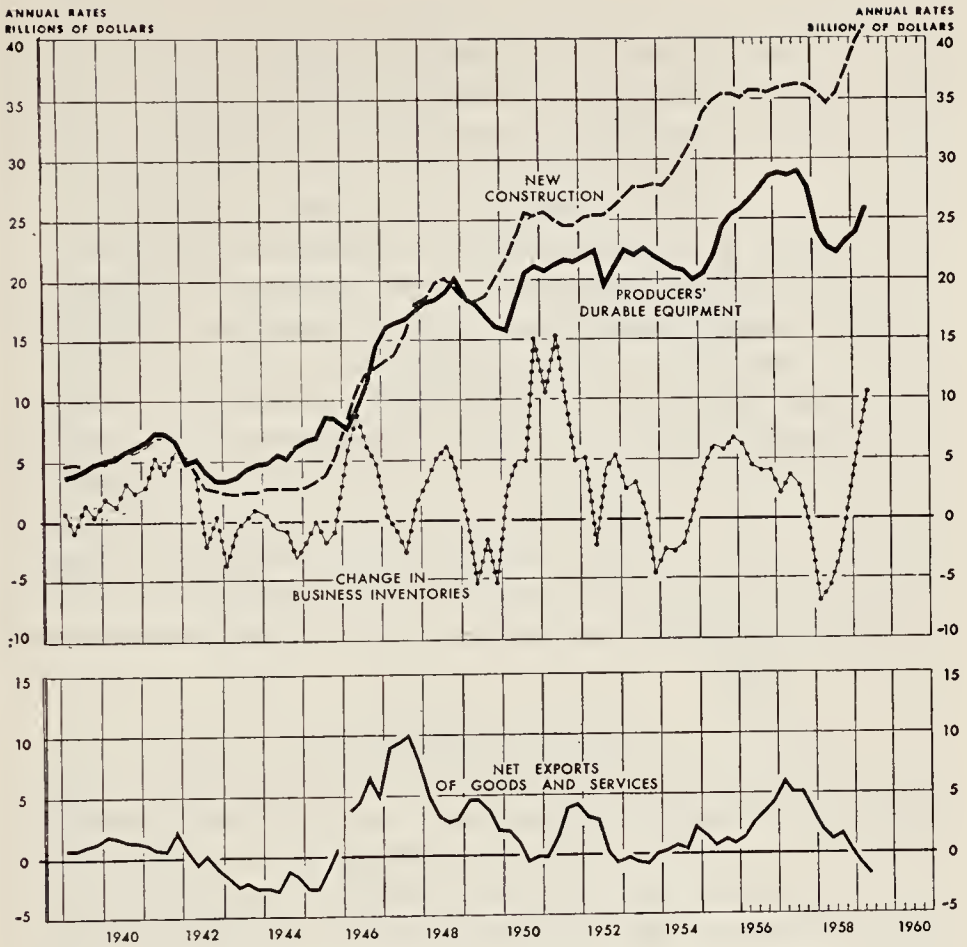


FIGURE 42. Components of Gross Capital Formation, Quarterly, 1939–1959.

From *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics.*

inventories, and businessmen generally followed a conservative policy with respect to the ratio of inventories to sales.

The foreign demand for American goods also played a role of considerable importance in supporting the level of business activity after the war. Both total exports and our export surplus rose rapidly to a peak in 1947. While both declined thereafter, the volume of exports continued to be abnormally large.

By 1947, foreign countries had virtually exhausted their free reserves of gold and dollars, which they had been using to pay for needed imports from the United States. This situation and con-

tinued recovery abroad, which lessened the need for American goods, were responsible for the decline in exports after 1947. The world-wide dollar shortage continued after 1947; and, despite American assistance, Great Britain and many other foreign countries were forced to devalue their currencies in 1949. The marked rise in American imports in 1950 made a substantial contribution toward relieving the dollar shortage in the rest of the world.

The six-billion dollar decline in the net export surplus from 1947 to 1948 undoubtedly contributed to the mild recession of 1948–1949; but it was only one of the contributing factors, and not the most important one. (Figure 42 shows that the peak in the export surplus was reached in the third quarter of 1947, a year or more before the downturn in business began.) The course of business activity was largely dominated by domestic influences. But the decline in foreign demand did serve to moderate somewhat the inflationary pressures operating on the American economy in 1947–1948, and this in itself contributed to the business downswing that began in the closing months of 1948.

THE 1949 RECESSION

The rapid rise in prices came to a halt in 1948; farm prices reached a peak early in the year and the general wholesale index reached its pre-Korean maximum in August. (See Figure 41.) In the latter part of the year, a general business recession set in which was not reversed until the latter half of 1949. The recession was mild and brief. Industrial production fell only by about 10 percent; GNP declined only by a little over 3 percent; and consumers' expenditures hardly declined at all. In all, the contraction lasted 11 months, from November, 1948, to October, 1949. By June, 1950, even before the invasion of South Korea, virtually all the ground lost in 1949 had been regained.[15]

CAUSES OF THE DOWNSWING

The causes of the 1949 decline can be summarized fairly briefly. Three sets of influences were operating in 1948 to bring the steady rise in aggregate demand and in the price level to at least a temporary halt.

[15] For a useful summary of the developments leading up to the 1948 downturn and an appraisal of the factors responsible for the recession, see Hickman, *op. cit.*, chap. 4.

1. Increasing supplies of goods, both in the United States and abroad, exerted a downward pressure on prices and led eventually to a reaction to the rapid rise in prices which had been in progress. The reaction was sharpest where the rise in prices had been most rapid, i.e., in the case of farm prices.

2. The abnormal expansion in consumers' demand began to level off, and personal saving began to rise. There were a number of reasons for this. The backlog of demand for most durable goods (except automobiles) had been or was close to being satisfied. In addition, the inflationary stimulus arising from the swollen supply of liquid assets was beginning to lose some of its force. The rise in prices had reduced the real value of the public's stock of liquid assets, and there had also been a movement of these assets from those who wanted to buy goods to those, in the higher income groups, who were willing to hold them. At the same time, the steady rise in consumers' debt eventually began to act as a check on the continued expansion in consumers' spending. Also, as consumers satisfied their most urgent needs, there was increasing resistance to high prices.[16]

3. Finally, private investment stopped expanding at its former rate and showed a tendency to decline. Residential building fell off in the final quarter of 1948 for reasons that we have already examined, and industry was beginning to satisfy its most urgent demands for new plant and equipment.[17] New firms had been an important source of demand for plant, equipment, and inventories during 1946–1947, but in 1948 the number of new businesses declined significantly.[18] There was also a sharp decline in our export surplus between 1947 and 1948.

These deflationary forces all came from the private sector of the economy. Paradoxically, the federal budget had been a deflationary force in 1946–1947, when the inflationary boom was at its height,

[16] Cf. D. Hamberg, "The Recession of 1948–49 in the United States," *Economic Journal*, vol. 62, March, 1952, pp. 1–14; also Hickman, *op. cit.*, chap. 4.

[17] Thus, in January, 1949, the Council of Economic Advisers stated, perhaps too sweepingly, that: "Business investment has entered a phase in which added investment will be largely determined by the effort to reduce costs and to improve product rather than by the purpose to expand capacity." See the Council's "Annual Economic Review," p. 18, in *The Economic Report of the President*, January, 1949. The role of declining plant and equipment expenditures in the 1949 recession has been particularly emphasized by C. A. Blyth in "The 1948–49 American Recession," *Economic Journal*, vol. 64, September, 1954, pp. 486–510.

[18] See *Survey of Current Business*, February, 1949, p. 2.

and turned inflationary during 1948, shortly before the downturn in business. Congress reduced tax rates in the spring of 1948, and during the same year adoption of the Marshall Plan and a stepped-up defense program led to increased government expenditures. The excess of cash receipts over expenditures for all levels of government declined from an annual rate of 16 billions in the fourth quarter of 1947 to 2 billions in the final quarter of 1948.[19]

What happened in 1949 is a good example of the distinction, which we emphasized in earlier chapters, between major-cycle and minor-cycle influences. The decline in prices, the leveling off in consumers' demand, and the return to a buyers' market in many lines led to a change in *short-term* expectations, which was reflected in a rapid decline in inventory investment. (See Table 33.) But underlying long-term investment opportunities remained favorable and, in the important field of residential construction, took a marked turn for the better. Also, government spending remained at a high level. Thus, despite the decline in inventory investment, the maintenance of government spending and private long-term investment served to prevent incomes and consumers' demand from falling very far. Actually, consumers' expenditures rose slightly in face of a small decline in disposable income.[20]

THE MILDNESS OF THE RECESSION

The nature of these influences can be readily seen in Table 33. Here are presented the main components of the GNP for the fourth quarters of 1948 and 1949, which marked the peak and trough of the decline, and for the last quarter before the Korean crisis.

The upper half of the table indicates clearly why this is called an inventory recession. The decline in inventory investment (9.6 billions) was slightly greater than that in total GNP. Consumers' expenditures rose during 1949,[21] while there was a small drop in

[19] *Ibid.*, p. 7.

[20] As Blyth has pointed out in the article previously cited, there was a 20 percent decline in expenditures on producers' durables between the fourth quarters of 1948 and 1949. (See Table 33.) Some part of this decline was geared to short-run expectations, although part certainly reflected satisfaction of backlog needs for capital goods. Significantly, the decline in nonresidential construction was much milder. Taking the economy as a whole, overall investment opportunities were still very favorable.

[21] Expenditures on nondurables declined by about three billions, but this was more than offset by an increase in spending on durable goods and on services.

TABLE 33. Components of the Gross National Product for Selected Calendar Quarters, 1948–1950
(Seasonally adjusted annual rates; in billions of dollars) [a]

	Fourth Quarter, 1948	Fourth Quarter, 1949	Second Quarter, 1950
Consumption expenditures	180.8	184.0	189.9
Durables	23.1	26.3	27.9
Nondurables	99.2	96.3	97.7
Services	58.5	61.5	64.3
Gross private domestic investment	43.9	30.6	46.9
New construction	19.4	19.9	23.6
Producers' durables	20.1	16.0	18.4
Change in inventories	4.3	−5.3	4.9
Net exports	3.0	2.1	1.1
Government expenditures [b]	38.2	40.3	36.5
Gross national product	265.9	257.0	274.4
National income	229.2	214.8	233.6
Less: Corporate profits	32.4	26.2	33.5
Other deductions [c]	5.4	5.7	6.7
Plus: Govt. transfer payments	9.9	11.8	14.2
Other additions [d]	13.0	13.3	13.9
Personal income	214.4	207.9	221.5
Less: Personal taxes	20.4	18.6	19.9
Disposable income	194.0	189.3	201.7
Consumption expenditures	180.8	184.0	189.9
Personal saving	13.2	5.3	11.8

[a] From U.S. Department of Commerce, *U.S. Income and Output*, 1958.
[b] On goods and services. Excludes transfer payments.
[c] Chiefly social security contributions.
[d] Includes government interest payments, corporation dividends, and business transfer payments.

private long-term investment (i.e., investment other than in inventories). A moderate decline in expenditures on nonresidential construction and equipment was partly offset by a rise in residential construction in the latter half of the year.

The lower part of the table helps to explain why consumption was so well maintained. Although the GNP fell about nine billions, disposable income declined by less than five billions. Three "automatic stabilizers" account for the failure of disposable income to fall fur-

ther. First, corporations maintained dividends despite a decline in profits. Second, some increase in unemployment and the need to support farm prices led to a rise in government transfer payments. And third, personal taxes were lower. Thus, disposable income declined by much less than did the GNP.[22] Consumption, however, did not fall at all. Instead, consumers reduced their savings and slightly increased their rate of spending.

With consumption so well maintained, manufacturers and retailers were able quickly to liquidate excess stocks of goods. By the beginning of 1950 they were finding it necessary to replace depleted inventories. Industry began to expand its orders even before the end of 1949, and a widespread and vigorous recovery spread through the economy after the turn of the year. The extent of the recovery up to the time of the Korean crisis is indicated by the last column in Table 33, as well as by Figure 40 on page 452. An outstanding feature of the upswing was the almost spectacular increase in residential building—a rise that had got under way in the first half of 1949 before the low point in general business activity was reached. The recovery was also supported by larger plant and equipment expenditures. However, the most important single factor in the expansion of private investment in the first half of 1950 was the cessation of inventory liquidation.

SUMMARY AND EVALUATION

This, then, was what we have called a minor recession, which was reflected primarily in a sharp change in inventory investment. Long-term investment opportunities remained favorable throughout the downswing; hence there was only a modest decline in long-term investment. Since government spending was also maintained, incomes could not fall very far. There was no serious impairment of business expectations, both because long-term investment opportunities remained favorable and because businessmen did not believe that a

[22] A comparison of the declines in GNP and disposable income does not tell the whole story. Actually, two strong destabilizers were also at work. Capital consumption (i.e., depreciation) increased by nearly two billions, and indirect business taxes rose by a billion. In addition, there was a change of nearly three billions in the statistical discrepancy. Thus the decline in national income was greater than in GNP. The stabilizers mentioned in the text led to the result, shown in Table 33, that disposable income declined only about a third as much as national income.

catastrophic decline in prices was "in the cards." A number of factors were responsible for this latter belief: the government's support program for farm prices, the easy credit situation and the favorable financial position of business generally, and the general feeling that high wages and other elements in the cost structure would keep prices high. The moderate price decline that did occur probably contributed to the removal of some maladjustments in particular markets and thus helped to bring on recovery.

Though the change in expectations was of the short-term variety, it should not be forgotten that there was the beginning of a change in the underlying situation in 1948. In particular, there was evidence that the most urgent pent-up demands by consumers and business were being satisfied. Apparently this required at the time no more than a tapering off in the previous rapid rate of advance. Underlying investment opportunities, especially, were still sufficiently widespread and profitable to call for a continued high level of investment. This was most strikingly the case in the field of housing, but it seemed also to be true for business plant and equipment.

Wages continued to rise during 1949, and, although consumer prices fell, the decline was very modest. As we shall see in the next chapter, the wage-price structure became increasingly invulnerable to downward pressure in the mild recessions of the 1950's. Another troublesome feature of the 1949 recession was the slowness with which excess unemployment was eliminated once recovery began. By the middle of 1950, both industrial production and the GNP in constant prices were above their 1948 peaks; yet the unemployment rate was still as high as 5 percent. The economy had not achieved a position of full employment at the time the Korean War broke out in June, 1950. While 1948 output levels had been exceeded, the substantial increase in the labor force since 1948 had not yet been fully absorbed.

CHAPTER 16

THE AMERICAN ECONOMY IN THE 1950's

BY JUNE, 1950, the American economy had almost fully recovered from the mild recession of 1949. Industrial production had already exceeded its 1948 peak; substantial inroads were being made into the unemployment left from the recession; and prices were again rising moderately in the fashion typical of cyclical expansions. Then came the invasion of South Korea; the United States found itself again in a "shooting war"; and thus a new set of inflationary pressures were released.

The decade of the 1950's spans two complete business cycles and the beginning of a third. The expansion phase of the first—from October, 1949, to July, 1953—was heavily influenced by developments associated with the Korean conflict. The ensuing recession was again mild, and this was followed by a vigorous capital-goods boom which was the outstanding feature of the long cyclical expansion from August, 1954, to July, 1957. Again the economy escaped with a mild and very brief recession. As the decade of the 1960's opened, a new cyclical expansion had been going on since April, 1958.

Price inflation continued to plague the United States—and, indeed, most of the rest of the world—through the larger part of the decade. The Korean war touched off a new but brief wave of sharply rising prices in 1950–1951. (See Figures 40 and 41 on pages 452 and 455.) Then came a lull that lasted through 1955, after which prices began to rise again, fairly rapidly in 1956–1957 and more slowly thereafter, with no offsetting decline during the 1958 recession. Concern over "creeping" inflation was being widely expressed as the decade ended.

It was also a period of substantial growth, in both total output and

real income per capita, but there were also signs that the rate of increase was slowing down in the later years of the decade. On balance, it was clearly a prosperous decade, and the fruits of prosperity were widely shared. Even so, the economy did not completely achieve the goal of continuous full employment. The record with respect to stability was on the whole good; the two recessions of the decade were both mild and brief. But unemployment continued to be something of a problem during the latter half of the decade. It steadily exceeded 4 percent of the labor force from 1954 on; and unemployment was a serious problem in some depressed areas, among Negroes, and among young people who had not yet acquired needed skills and seniority rights. It is significant that, as the 1950's ended, a vigorous debate was going on as to how the American economy could best reconcile the possibly conflicting goals of full employment and rapid growth, on the one hand, and price stability, on the other.[1]

THE KOREAN PERIOD

The Communist invasion of South Korea immediately unleashed a wave of anticipatory buying by both consumers and businessmen; and, as a result, prices rose rapidly during the remainder of 1950 and into the early months of 1951.[2] Although the Federal government immediately moved to build up defense expenditures as rapidly as possible, this was a process that took time. As a result, the inflationary pressures during the latter half of 1950 and the early months of 1951 came chiefly from the private sector of the economy, supplemented by the impact of the initial placement of orders for armaments which led private manufacturers to increase their orders for raw materials and semifinished goods and to step up their hiring of workers. But the big rise in actual government expenditures came in 1951 rather than 1950.

Anticipating price increases and future shortages, consumers sharply increased their purchases in the third quarter of 1950. In effect, the consumption function shifted upward. Indeed, the increase in retail sales was so large that, for the time being, business was

[1] See the numerous volumes of hearings and staff reports published by the Joint Economic Committee under the general title, *Study of Employment, Growth, and Price Levels* (86th Congress, 1st and 2nd Sessions, 1959–1960).

[2] For a more detailed analysis of developments during the Korean War, see Bert G. Hickman, *Growth and Stability of the Postwar Economy*, 1960, chap. 5, and the *Economic Reports of the President* for the appropriate years.

frustrated in its attempts to build up inventories. In the fourth quarter, retail sales fell off, and a huge increase in inventory investment occurred. (See Figure 42.) In the meantime, business was expanding its capital budgets to provide for larger expenditures on plant and equipment.

The federal government moved promptly to control the inflationary pressures that had been unleashed. Personal and corporate income taxes were substantially increased in September; an excess profits tax was enacted; controls were reimposed on installment credit; and new restrictions on mortgage credit were put into effect to curb the housing boom. Since tax revenues increased immediately, as a result of both the rise in incomes and the higher tax rates, while government expenditures rose much more slowly, a large budgetary surplus accumulated in the second half of the year.

These measures were not enough to bring the rise in prices to a prompt halt, nor were they able to prevent a new wave of anticipatory buying and inventory accumulation which broke out after the Chinese Communists entered the war in November.

This second buying wave subsided by the end of the first quarter of 1951. Early in the year, general price and wage controls were reimposed on the economy. From the end of the first quarter of 1951 until the fall of 1952, the economy was surprisingly stable. Wholesale prices declined gradually through the rest of 1951 and all of 1952. While the Consumer Price Index did not fall, its rise after early 1951 was very moderate. (See Figures 40 and 41.) The rise in consumers' expenditures became more gradual, a substantial decline in purchases of consumers' durables being somewhat more than offset by continued expansion in expenditures on nondurables and services. Inventory accumulation fell off sharply after the second quarter of 1951,[3] and noninventory investment began to move more or less horizontally, with the result that total gross private domestic investment declined steadily through the latter half of 1951 and the first half of 1952.

What is remarkable about all this is that it was just during this period, after the end of 1950, that the big rise in defense expenditures

[3] The decline in inventory investment was concentrated in wholesale and retail trade and in manufacturing industries producing consumers' goods. The defense build-up required a rapid accumulation of inventories of defense goods and materials throughout the year. Cf. *Survey of Current Business*, February, 1952, pp. 18–19.

occurred and a substantial budgetary surplus was converted into a sizable deficit.[4] Federal government expenditures on new goods and services more than doubled, from an annual rate of 22.7 to a rate of 47.4 billions, between the fourth quarters of 1950 and 1951. (They continued to rise, though more slowly, to a peak of 58.9 billion in the second quarter of 1953.)

Undoubtedly, government controls had something to do with the stability of the economy following the early months of 1951. Price controls removed some fears regarding runaway inflation and thus reduced speculative buying; wage controls helped to reduce inflationary pressures emanating from tight labor markets and strong trade unions; and the battery of credit and allocation controls worked to hold back consumers' purchases of durables, residential building, inventory accumulation, and those types of business investment not considered necessary to the defense effort.

While all of this is true, the chief reason for the economy's stability during 1951–1952 probably lay in another direction. Consumers and businessmen had stocked up so well in the two big buying waves in the fall and winter of 1950–1951 that some recession in private buying was virtually inevitable.[5] Even with the diversion of resources to military production, capacity to produce civilian goods proved to be ample, and businessmen found that their inventories were larger than their sales required. Significantly, prices of some goods fell below the price ceilings imposed by the price-control authorities.[6]

Thus, to use Hickman's phrase, we had a "divided economy" during 1951–1952. Government expenditures and the production of military-type goods rose rapidly; but consumers' expenditures remained relatively stable, and production of many types of consumers' goods actually declined. While consumers' restraint and inventory disinvestment were largely responsible for these developments, government controls played an important contributing role. As we suggested earlier, higher tax rates restrained the rise in disposable incomes; selective credit and allocation controls helped to curb non-

[4] Cf. *Economic Report of the President*, January, 1952, pp. 89, 91.

[5] As John P. Lewis has put it, referring to the relative price stability after early 1951, "we could not have gotten the sobriety of the lull without the excesses that immediately preceded it." "The Lull That Came to Stay," *Journal of Political Economy*, vol. 63, February, 1955, p. 7.

[6] Cf. Hickman, *op. cit.*, chap. 5. See also the stimulating article by Lewis cited in the preceding footnote.

essential private investment; and wage and price controls not only had some direct influence but also helped to create consumer and business expectations that were favorable to economic stability.

THE 1953 DOWNTURN AND THE ENSUING RECESSION

Under the stimulus of a continued rise in defense expenditures, business activity expanded further during the first nine months of 1952, although a steel strike caused some interruption during the summer. In general, the lull in consumers' spending and in the behavior of prices that prevailed through most of 1951 continued until about September. At that point there was a spurt in private spending, both by consumers and by business; output and employment increased sharply in the closing months of 1952 and then began to level off in the spring of 1953. This was the last phase of a cyclical expansion that finally reached its peak in July, 1953.

CAUSES OF THE DOWNTURN

Following the steel strike in July, 1952, private demand expanded rapidly, especially for durable goods. A number of factors were responsible for this upsurge. Inventories had to be replenished after the steel strike. Government controls had been relaxed as the rise in military expenditures began to level off, capacity continued to increase, and a larger supply of scarce materials could be released for civilian use. Also, after a year and a half of fairly restrained buying, consumers were in a mood—and had the cash or borrowing power—to step up their purchases, particularly of durable goods.

As a result, consumers' expenditures increased sharply in the fourth quarter of 1952, and inventory accumulation by businessmen also rose. (See Table 34.) Rapidly expanding production and payrolls increased disposable income, but consumers' spending increased faster still, implying an upward shift in the consumption function.[7] At the same time, other elements of aggregate demand continued to rise—for example, residential building, national defense expenditures, and the spending of state and local governments.

Retail sales leveled off rapidly in the first half of 1953. *Total* consumers' expenditures continued to rise about in line with the rise in disposable income, but most of the increase went into services rather

[7] Note in Table 34 that personal saving declined in the fourth quarter of 1952 despite the rise in disposable income.

than durable or nondurable goods. Services, however, unlike tangible commodities, do not require a pipeline of inventories in the hands of retailers, wholesalers, and manufacturers.

Given the flattening out in retail sales, the level of output achieved by the early months of 1953 could not be maintained. The rapid rise in production, particularly of durable goods, in late 1952 and early 1953 was geared to (1) the current increase in consumers' demand, (2) expectations of further expansion of demand in 1953, (3) the need to replenish inventories depleted by the steel strike, and (4) the need not merely to replenish inventories but to *increase* them in line with the actual and anticipated increase in sales. The flattening out in sales caused a similar movement in output, and this more or less horizontal movement in output and sales meant that the current level of production could no longer be maintained. Part of current output was going into inventories, and inventories were already adequate if not excessive.

Another factor was emerging that called for a decline in inventory investment. This was the tapering off in the defense program. Defense expenditures reached a peak in the second quarter of 1953, but by then the government had already begun to reduce its new orders to manufacturers in defense industries.

Thus, the ground was laid for an inventory recession, somewhat after the fashion of the Metzler inventory model described in Chapter 12.[8] Actually, the inventory-adjustment mechanism at work was considerably more complicated than Metzler's simple model. First of all, there was an element of mistaken anticipations at work. Particularly in the durable-goods industries, producers expected an increase in demand in the first half of 1953 that failed to materialize. Secondly, and more important, there were autonomous shifts in consumers' expenditures that a Metzler-type model cannot readily take into account. The propensity to consume—particularly the propensity to purchase durable goods—shifted upward in the fall of 1952; and the propensity to purchase (particularly durable) goods shifted downward in the first half of 1953, in part because total consumption returned to a more normal relationship to income and in part because of the shift from goods to services previously noted.

In addition to the consumption-inventory phenomena just described, two other sets of factors played some part in the downturn,

[8] See pp. 351–352.

which we can date from July, 1953. One was monetary tightness; the other was the substantial decline in federal government expenditures that began in the third quarter of the year.

In Chapter 15 we commented on the easy money policy that the Federal Reserve authorities had pursued during the war and immediate postwar years. This period came to an end with the famous agreement, or Accord, between the Federal Reserve and the Treasury in March, 1951. Under this agreement, the Federal Reserve authorities again acquired the ability to pursue a restrictive credit policy if they thought such a policy desirable, even though the result might be a rise in interest rates and a consequent decline in the prices of government bonds. Actually, interest rates had been rising gradually since the end of the 1949 recession. This rise accelerated toward the end of 1952 and became quite sharp in the spring of 1953.[9] The Federal Reserve banks did not supply reserves to the banking system as fast as the demand for bank credit was expanding; as a result, member banks were forced to increase their rediscounts from the Reserve banks, to sell some of their holdings of government securities, and to restrain the expansion in loans. Credit continued to tighten until June, 1953. In general, the rise in interest rates during this period was not extreme; and at their peak short-term interest rates and long-term bond yields were both much below the levels that were reached after 1955.

Concerned about the degree of credit tightness that was developing and the effect it was having on anticipations, the Federal Reserve authorities executed a quick reversal, and credit conditions eased rapidly from June on.

It is doubtful if these monetary developments played more than a minor role in the downturn.[10] First of all, money market conditions began to ease a month or more before the downturn. While falling bond prices in the spring led to postponement of some security issues, there is little evidence that tight money had much effect on actual capital expenditures. Nor, following Hawtrey, can we attribute

[9] Flotation of a large long-term treasury issue in April and increased indications that the federal government would be a heavy borrower during the rest of the year added to the credit stringency.

[10] This is also Hickman's view (*op. cit.*, chap. 6). For a more extended review of monetary developments in 1953, see the *Monthly Review* of the Federal Reserve Bank of New York, July, 1953, and February, 1954; *Federal Reserve Bulletin*, February, 1954; and *Economic Report of the President*, January, 1954, pp. 32–35.

the eventual decline in inventories to high interest rates. As we have already seen, the behavior of sales required a decline in inventory accumulation, regardless of the level of interest rates.

More important in the business downturn than temporary monetary stringency was the cutback in military procurement. Defense expenditures started to decline in the third quarter of 1953; but the flow of new defense orders began to diminish before the business peak in July. Once the downturn began, declining government expenditures acted as a deflationary force for a year or more.

THE MILDNESS OF THE RECESSION

The contraction lasted 13 months, until August, 1954. Industrial production fell about 10 percent; GNP declined only about 3 percent; and the unemployment rate rose from about 2 to nearly 6 percent. Prices moved more or less horizontally, and wages continued to rise, though more slowly than in earlier postwar years.

Examination of Table 34 quickly reveals that the downswing can be explained in terms of declines in just two components of aggregate demand: inventory investment and federal government expenditures. These two together declined by a total of about 17 billion dollars between the second quarters of 1953 and 1954. This was considerably more than the decline in total GNP.

Clearly this is the case of another inventory recession—a recession which might have been so mild as hardly to be noticeable had it not been for the decline in federal defense expenditures. Noninventory investment did not fall at all, a small decrease in expenditures on producers' durables being offset by an increase in construction (both residential and nonresidential); there was a large increase in state and local expenditures; and, despite the decline in GNP and rise in unemployment, consumers' expenditures actually increased by some three billion dollars.

Why did consumers' expenditures rise, and why was long-term investment so stable?

To answer the first of these questions, we need to look at the lower part of Table 34. First of all, the automatic stabilizers were powerfully at work. Corporate profits declined by 6.5 billions, but dividends did not decline. This alone offset about two thirds of the decline in GNP. In addition, transfer payments increased by over two billions.

TABLE 34. Components of the Gross National Product for Selected Calendar Quarters, 1952–1954
(Seasonally adjusted annual rates; in billions of dollars) [a]

	Third Quarter, 1952	Fourth Quarter, 1952	Second Quarter, 1953	Second Quarter, 1954
Consumption expenditures	219.6	227.2	233.3	236.5
Durables	27.5	32.1	33.4	32.2
Nondurables	115.9	117.2	118.6	118.8
Services	76.2	77.9	81.2	85.5
Gross private domestic investment	49.1	52.6	52.9	47.2
New construction	25.4	26.1	27.8	28.9
Producers' durables	19.4	21.2	22.0	20.9
Change in inventories	4.3	5.3	3.1	−2.7
(Nonfarm only)	(3.4)	(4.7)	(4.0)	(−3.2)
Net exports	0.1	−0.7	−0.7	0.8
Government expenditures [b]	78.2	79.5	83.3	74.4
Federal	55.0	55.8	58.9	47.1
State and local	23.2	23.7	24.3	27.3
Gross National Product	347.0	358.6	368.8	358.9
National Income	292.4	300.6	308.7	299.4
Less: Corporate profits	36.0	38.9	39.8	33.3
Other deductions [c]	8.3	8.8	8.8	9.6
Plus: Government transfer payments	12.3	12.5	12.7	14.9
Other additions [d]	15.2	15.2	15.9	16.2
Personal income	275.6	280.6	288.7	287.6
Less: Personal taxes	34.5	35.0	35.9	32.8
Disposable income	241.1	245.6	252.8	254.8
Consumption expenditures	219.6	227.2	233.3	236.5
Personal saving	21.5	18.4	19.6	18.3

[a] From U.S. Department of Commerce, *U.S. Income and Output*, 1958.
[b] On goods and services. Excludes transfer payments.
[c] Chiefly social security contributions.
[d] Includes government interest payments, corporation dividends, and business transfer payments.

On top of this, personal taxes declined by three billion dollars. For the most part, however, this did not represent an automatic stabilizer at work. A tax cut had for some time been scheduled to go into effect in January, 1954; and, following the Korean truce, the Administration decided not to seek to defer the reduction that had been

scheduled. The total tax saving involved was about three billion dollars.[11] Here we have an example of *discretionary* policy which turned out, largely by chance, to have just the right timing.

The net result of all these influences was that disposable income rose slightly in the face of a decline of GNP. In addition, consumers increased their spending a bit more than the increase in disposable income. Thus, there was no multiplier effect at all from the fall in defense spending and inventory investment. Indeed, not only was there no decline in disposable income but there was also a modest upward shift in the consumption function. This was only one of a number of occasions in the postwar years when consumers' spending proved to be an autonomous factor in the movement of aggregate demand. The consuming public has, to a considerable extent, refused to play the purely passive role reserved for it in the simpler types of aggregative models.

Despite the stabilizers and the favorable behavior of consumers' expenditures, the 1954 recession would have been much more severe than it was had it not been for the behavior of long-term investment. The decline in business expenditures on producers' durables was very mild, and new construction actually increased. Clearly, autonomous forces were at work to maintain a high level of long-term investment. Hence, the needed adjustment in inventories was quickly effected, and the economy was able without difficulty to absorb the decline in government defense expenditures. Of the autonomous forces operating to sustain and expand long-term investment, the following particularly should be noted.

1. Residential building has continued to support aggregate demand in each of the postwar recessions. The underlying demand for new housing remained strong throughout the 1950's. Equally important, the short-run behavior of residential building had a stabilizing effect in each of the postwar recessions, falling little if at all after business turned down and then beginning to expand again before or at about the time that general business revival set in. Two sets of reasons account for this stabilizing behavior. First, the federal government acted in 1954, as it did in the other postwar recessions, to stimulate residential building, particularly by liberalizing the terms on which mortgage credit could be secured. More important, the opera-

[11] Offset to a minor extent by an increase in social security taxes. On this, see *Economic Report of the President,* January, 1955, p. 19.

tion of the postwar money and capital markets has had the effect of causing residential building to behave almost as if it were an automatic stabilizer. When interest rates rise and credit becomes tight during a cyclical boom, lenders divert more of their funds from residential mortgages, particularly guaranteed and insured mortgages with relatively fixed interest rates, to other types of securities. When credit becomes easy in a recession and yields on nonmortgage securities fall, life insurance companies and other lenders eagerly seek to invest in residential mortgages. This happened on a large scale in 1954, with the result that mortgage lending and residential building began to expand rapidly from about the beginning of 1954. New housing starts rose about 40 percent between December, 1953, and December, 1954.[12]

2. There was a strong and rising demand for commercial buildings of all types, and in the 1950's the United States had its first commercial building boom since the 1920's.

3. Plant and equipment expenditures declined only moderately. Investment opportunities were still large in a variety of industries under the spur of rapid technological change and continued growth of population and income. While this type of investment did not rise until after the general upturn in business, it did not fall very far in 1954, and it provided a powerful stimulus to expansion during 1955–1957.

4. Capital expenditures of state and local governments continued to rise steadily during the 1953–1954 recession, as they have every year since World War II. The decline in interest rates and the better market for state and local bonds made it easier to meet some of the seemingly insatiable demands for schools, roads and streets, and other essential public improvements.

THE CYCLE OF 1954–1958

Rapid recovery got under way in the latter part of 1954, sparked particularly by the new boom in residential building and the steady expansion in consumers' expenditures, which accelerated in the last quarter of the year. Inventory disinvestment ceased after the third quarter; the decline in federal government expenditures tapered off after the second quarter; and easy credit conditions stimulated

[12] See Leo Grebler, *Housing Issues in Economic Stabilization Policy,* National Bureau of Economic Research Occasional Paper 72, 1960, esp. chap. 2.

building activity and installment buying. Consumers' expenditures rose more rapidly than disposable income. An important stimulus from the fourth quarter on was the public's enthusiastic acceptance of the 1955 model automobiles. Plant and equipment expenditures lagged at the upturn; but, once recovery began, business firms rapidly revised upward their investment plans.[13]

THE DURABLE GOODS BOOM, 1955–1957

Expansion was extremely rapid during 1955. Industrial production then flattened out but remained at a high level during 1956 and most of 1957. (See Figure 40.) A number of features make this cyclical expansion of particular interest.

1. It was the first postwar boom not to be dominated by either pent-up demands inherited from a war (1946–1948) or by actual war (1950–1953).

2. The boom was particularly concentrated in the durable-goods industries. This was associated with a very rapid rise in plant and equipment expenditures, a boom in residential and commercial construction, and the rush to buy automobiles in 1955.

3. The economy teetered on the brink of contraction for an unusually long period, from early in 1956 until the peak was finally reached in July, 1957.

4. The period of price stability that had begun in 1951 finally ended. Both the consumer and wholesale price indices rose rapidly during 1956 and 1957, and they even continued to increase moderately during the short 1957–1958 recession. (See Figures 40 and 41.)

5. The investment boom led to a rapid increase in capacity but, surprisingly, the improvement in labor productivity was unusually small in 1956 and 1957. At the same time, wages rose rapidly, with a consequent increase in unit labor costs not completely offset by rising prices, with the result that there was some deterioration in profit margins in the later stages of the boom.

6. Despite the vigor of the expansion, unemployment did not fall below 4 percent, and the growth in total output during 1956–1957 was not very large.

7. The monetary authorities found themselves in a particularly unhappy situation. The rapid rise in prices and the vigorous invest-

[13] This discussion of the 1954–1958 cycle owes a good deal to the excellent analysis in Hickman, *op. cit.*, esp. chap. 7.

ment boom led them to follow a restrictive monetary policy, while the slow growth in output after 1955 and the existence of pockets of unemployment led many observers to advocate a policy of monetary ease. This dilemma gave rise to a vigorous debate regarding the mutual compatibility of the goals of full employment, rapid growth, and price stability.

Let us now look at the boom in greater detail.

Expansion in 1955 was rapid and extended to virtually all sectors of the economy except the federal government. The lead had already been taken in 1954 by a sharp rise in residential building. This was followed in late 1954 and early 1955 by a big spurt in expenditures on consumers' durables, particularly automobiles, and by a large increase in inventory investment. Beginning in the first quarter of 1955, also, plant and equipment expenditures began to rise rapidly.

The boom flattened out in the closing months of 1955. The peak in consumers' expenditures on durables was reached as early as the third quarter of 1955, and public reception of the 1956 automobile models was disappointing. Residential building also began to decline after the third quarter of 1955, particularly because of tight credit conditions. What kept the boom going during 1956 and into 1957 was the continued rise in plant and equipment expenditures by business (including commercial building), the steady upward trend in state and local government expenditures, and the increase in consumers' expenditures on nondurables and services. Federal government expenditures were stable through 1955 and the first half of 1956 and then rose moderately to a peak in the third quarter of 1957. Net exports also provided an additional stimulus.

The result of these various forces can be traced in the index of industrial production. From a level of 86 in September, 1954 (1957 = 100), the index rose rapidly to an initial peak of 100 in December, 1955, and then moved more or less horizontally. There was a further slight increase at the end of 1956, but at no time during 1956–1957 did the index exceed 102, compared to 100 in December 1955.[14]

By the beginning of 1956 the economy was operating close to full capacity, with bottlenecks appearing in various durable-goods industries, where there was clearest evidence that demand was threat-

[14] These figures refer to the revised Federal Reserve index, including utilities. Cf. *Federal Reserve Bulletin*, December, 1959, p. 1474.

ening to outstrip supply. But, on the whole, the more or less horizontal movement of total output in 1956 and the first half of 1957 was due more to the failure of aggregate demand to continue to expand sufficiently than to an overall insufficiency of supply. Capacity expanded rapidly in 1956–1957, and so did the labor force. A slackening in demand was felt particularly in the automobile industry, in housebuilding, and in related lines.[15]

Deflated for price changes, consumers' expenditures for other than services flattened out markedly in 1956, and the proportion of income saved began to rise. One result was that inventory investment declined all through 1956 and into the first quarter of 1957. (See Figure 42.)

All this suggests that a minor recession might well have begun sometime in 1956. What saved the situation was the continued expansion in business fixed investment, a rapid increase in exports, and a rise in federal government expenditures after the middle of 1956. The continued expansion in private long-term investment in the face of a marked retardation in demand requires explanation, even if we do not place much faith in the short-run working of the acceleration principle.

Long-term investment continued to rise (to a peak in the first quarter of 1957) for two reasons in particular. First, much investment was autonomous and only loosely related, if at all, to short-run changes in demand. This was particularly true of commercial and some other types of construction. It was probably also true of a good deal of investment inspired by technological change and the rise in labor costs. Secondly, substantial lags were involved in many types of investment. Expenditures made in late 1956 and early 1957 represented commitments that had been incurred sometime earlier. Decisions made in the latter half of 1956 to curtail investment would be reflected in an actual decline in expenditures only some months later.[16] It is worth noting that expenditures on producers' durables

[15] For additional evidence on the failure of aggregate demand to rise as rapidly as capacity in 1956–1957, see Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959), pp. 74 ff.

[16] Thus capital appropriations in manufacturing reached a peak in the first half of 1956, although actual investment expenditures for the same companies continued to rise until the third quarter of 1957. Cf. *The Conference Board Business Record,* March, 1958.

reached their peak a half-year earlier than nonresidential building. The former type of investment is subject to shorter lags and is also more sensitive to changes in total output.

Just about the time the boom began to flatten out, from the beginning of 1956 on, prices began to rise rapidly. The decline in farm prices ended. In the industrial sphere prices rose, particularly in the durable-goods industries (steel, machinery, etc.) [17] At the consumers' level, price increases in 1956–1957 were particularly marked in food and in services. Wages rose rapidly throughout the economy, as indeed they had been doing throughout the postwar period (except for brief pauses during recession periods).

The boom was also marked by a considerable degree of monetary tightness. Short-term interest rates reached a low in the middle of 1954 and then began a rise that continued for three years. Long-term bond yields rose sharply beginning early in 1956. The capital-goods boom, continued government borrowing, and the rising need for working capital put heavy demands on the money and capital markets. The Federal Reserve authorities followed a restrictive credit policy, keeping pressure on member bank reserves. As a result, commercial banks were forced to sell government securities and to rediscount at the reserve banks. The money supply rose much less rapidly than the money value of GNP, with a consequent reduction in the liquidity of the economy. By the peak in 1957, interest rates were higher than they had been in more than twenty years.[18]

THE DOWNTURN IN 1957

The expansion showed signs of new life in the latter part of 1956, helped by the recovery from a steel strike, a rise in federal government expenditures, a boom in exports (in part because of the Suez crisis), and an increase in automobile production associated with the introduction of new models. But industrial production failed to rise any further after the beginning of 1957; the economy entered into what we have called a turning point zone; and a cumulative contraction developed in the latter half of the year. The turning point came in July.

[17] Cf. Joint Economic Committee, *op. cit.*, pp. 123–125. See also Hickman, *op. cit.*, chap. 14, for a discussion of price movements during this and the preceding postwar cycles.

[18] For an official review of these monetary developments, see the *Annual Reports* of the Board of Governors of the Federal Reserve System for 1957 and 1958.

A number of deflationary forces were already at work in the first half of the year. Manufacturers' new orders for durable goods were declining, both because of disappointing sales of consumers' durables and because a number of industries were beginning to cut back their investment programs. Investment programs were being curtailed chiefly because, in many lines, some excess capacity was beginning to emerge as sales leveled off or began even to decline. Tight money and declining profit margins may also have played a contributing role. On top of this, the decline in residential building, which was largely the result of tight credit, continued.

Temporarily offsetting these deflationary forces were the continued expansion in consumers' expenditures on nondurables and on services, the high level of exports, and the rise in federal government expenditures during the first half of the year. But beginning around the middle of the year, both federal government expenditures and exports began to decline; the economy's precarious balance was upset; and a cumulative contraction began.

THE RECESSION OF 1957–1958

The recession that followed was notable in a number of respects.[19] It was one of the briefest contractions of the last century, lasting only nine months, till April, 1958. Yet, though brief, the decline was very rapid; and, in terms of the decline in GNP, industrial production, and employment, it was the most severe (although the shortest) of the recessions experienced since World War II. (See Table 35.) There was a significant decline in long-term investment (particularly in producers' durables); yet a major depression did not develop. Prices continued to rise during the contraction in business activity, in marked contrast to the "typical" recession behavior of prices. The recession was felt in a number of other countries, either as an actual decline in output or as a marked retardation in the rapid growth that had been under way for some years.

Once the downswing began, it spread quickly in typical cumulative fashion. As output declined, so did employment and incomes. Retail sales fell off, particularly for durable goods. There was a sharp

[19] In addition to the other sources cited in the preceding and following pages, see H. D. Osborne, "National Income and Product—A Review of the 1957–58 Decline and Recovery," *Survey of Current Business,* November, 1958, pp. 9–17. See also G. H. Moore, "The 1957–58 Business Contraction: New Model or Old," *American Economic Review,* vol. 49, May, 1959, pp. 292–308.

decline in inventory investment, and the decline in plant and equipment expenditures accelerated. Fears that this recession might turn out to be a serious one spread. Yet, as concern mounted in the spring of 1958, recovery suddenly took hold, and rapid expansion characterized the economy during the rest of the year and into 1959.

What accounted for the brevity of the recession and for the fact that it was relatively mild, although moderately more severe than the recessions of 1948–1949 and 1953–1954? The factors making for quick recovery fall under the following headings: the effective work-

TABLE 35. Amplitude of Decline in GNP, Industrial Production, and Employment in Three Postwar Recessions [a]

	Percentage Decline [b]		
	1948–1949	1953–1954	1957–1958
Real gross national product (quarterly, 1954 prices)	2.4	3.6	4.7
Industrial production (monthly, 1947–49 = 100)	10.5	10.2	13.7
Nonagricultural employment	3.8	3.5	4.6

[a] From B. G. Hickman, *Growth and Stability of the Postwar Economy*, 1960.
[b] Based on total decline measured from peak to trough in each series.

ing of the automatic stabilizers, special factors leading to an increase in farm incomes, the cyclical insensitivity of sectors that have been becoming steadily more important, the quick cessation of the declines in federal spending and in exports, the favorable behavior of housebuilding, and the autonomous supports provided by the continued rise in state and local government expenditures and by the relative stability in some important types of long-term investment. The result of these factors was that incomes and consumers' expenditures were well maintained; the necessary inventory adjustment took place quickly; and output began to recover from the spring of 1958.

The automatic stabilizers operated powerfully, and again the most powerful stabilizer of all was corporate profits. (Table 36.) Whereas GNP declined by 16.8 billion dollars between the third quarter of 1957 and the first quarter of 1958, 11.2 billions of this drop was absorbed by the decline in corporate profits while dividends scarcely declined at all. In addition, transfer payments rose by

TABLE 36. Components of the Gross National Product for Selected Calendar Quarters, 1957–1959
(Seasonally adjusted annual rates; in billions of dollars) [a]

	Third Quarter, 1957	First Quarter, 1958	First Quarter, 1959
Consumption expenditures	288.2	287.3	303.9
Durables	40.9	36.9	41.3
Nondurables	139.7	139.5	145.3
Services	107.6	111.0	117.4
Gross private domestic investment	67.9	52.4	70.0
New construction	36.2	35.5	39.7
Producers' durables	29.0	23.8	23.9
Change in inventories	2.7	−6.9	6.3
Net exports	5.1	2.0	−0.9
Government expenditures [b]	86.6	89.3	97.4
Federal	49.7	50.1	53.8
State and local	36.9	39.2	43.6
Gross national product	447.8	431.0	470.4
National income	371.1	355.8	389.4
Less: Corporate profits	42.7	31.5	45.5
Other deductions [c]	14.8	15.4	17.5
Plus: Government transfer payments	20.1	22.8	24.7
Other additions [d]	20.7	20.6	20.8
Personal income	354.5	352.2	371.8
Less: Personal taxes	43.1	41.9	44.4
Disposable income	311.5	310.3	327.4
Consumption expenditures	288.2	287.3	303.9
Personal saving	23.3	22.9	23.5

[a] From *Survey of Current Business*, July, 1959, and *Federal Reserve Bulletin*, June, 1960.
[b] On goods and services. Excludes transfer payments.
[c] Chiefly social security contributions.
[d] Includes government interest payments, corporation dividends, and business transfer payments.

2.7 billions and personal taxes fell by 1.2 billion. The result was that disposable income fell by only 1.2 billion, and consumers' expenditures declined even less.

Farm incomes actually rose during the recession, in part because of reduced supplies of some products with inelastic demands and in part because of an increase in government benefit payments for those crops subject to federal price supports. Wage payments were well

maintained in the distributive trades and actually increased in the service industries and in government. This tended, of course, to support consumers' expenditures on commodities as well as services, with the result that commodity production revived fairly quickly.

Residential building again acted somewhat as an automatic stabilizer, although not so promptly as in 1954. Housebuilding remained stable during the recession and then, from about April on, began to expand rapidly. As credit conditions eased, mortgage funds became much more readily available; and, in addition, the government liberalized the terms for borrowing on guaranteed and insured (FHA and VA) mortgages.[20]

This favorable short-run behavior of residential building was possible only because the underlying demand for housing remained strong with the continued upward trend in new families and in incomes, the migration to the suburbs, an increasing amount of demolition resulting from freeway construction and urban redevelopment, and a resurgence of apartment-house construction in the cities.

Investment opportunities continued to be favorable in other directions, particularly in the public utility industries and in commercial construction (stores, warehouses, office buildings, motels, etc.)[21] Total business expenditures for new plant and equipment declined by about 8 billion dollars or 22 percent in the year following the third quarter of 1957. The decline in manufacturing was 5.5 billions or 34 percent, and there were also large relative declines (although the total decrease was only about 1.8 billion) in mining, railroads, and nonrail transportation. In contrast, fixed investment in the public utility and the "commercial and other" categories declined only from 16.8 to 16.0 billion in the same period.[22] And in the same period, public construction rose from about 13.8 to 14.8 billion.

Thus, favorable investment opportunities in broad sectors of the economy, both private and public, largely offset such (presumably

[20] Interest rates fell rapidly in late 1957 and early 1958 as credit restraint was followed by monetary ease. The Federal Reserve authorities were slower to act, however, than they had been in 1953. Their first positive steps to ease the pressure on bank reserves did not come until October–November, 1957, several months after business had already turned down.

[21] However, electric power investment declined in 1959, after general business recovery was well under way.

[22] Cf. *Economic Report of the President,* January, 1960, p. 190, and Hickman, *op. cit.* "Commercial and other" includes the trade, service, finance, communications, and construction industries.

temporary) impairment of investment opportunities as occurred in manufacturing, mining, and transportation. In addition, the mildness of the decline in total activity, the greater ease in the capital markets, and the early and rapid recovery in the stock market helped to maintain the inducement to exploit investment opportunities in lines where they were still favorable.

Federal government expenditures ceased to act as a deflationary force after the beginning of 1958. Defense expenditures rose slowly after reaching a low point in the fourth quarter of 1957, and a speeding up in the placing of new defense orders, even before money was actually spent, provided a further stimulus. In addition, there was a sharp increase in nondefense expenditures, chiefly to purchase farm commodities under the agricultural support program. The decline in exports also ceased after the first quarter of 1958, but imports then rose more than exports, so that the net export surplus continued to decline throughout the year.

Thus, what began as if it might become the first major depression of the postwar period quickly developed into what was essentially a minor recession. There was a significant decline in private long-term investment, especially in manufacturing, but the decline was largely in producers' durables rather than construction. (By the end of 1958, total construction exceeded its 1957 peak, and it rose much further in 1959.) While total private long-term investment was declining, a substantial part of the fall was offset by the rise in government expenditures, both state and federal. The fact that there was not a complete offset, plus the sharp decline in inventory investment and in net exports, account for the decline in aggregate demand that did take place. (Table 36.) The automatic stabilizers, aided by the special circumstances that led to the increase in farm incomes, kept disposable income and consumers' expenditures from declining significantly.

RETROSPECT AND PROSPECT

By the spring of 1960, the recovery from the 1958 recession had been in progress for two years, although interrupted by the prolonged steel strike in the latter half of 1959. By the end of 1958, the GNP, in both current and constant prices, had exceeded its 1957 peak; in the first half of 1960, the GNP in current prices reached the magic figure of 500 billions.

The 1960's opened with business expanding rapidly after the steel strike and amid generally optimistic expectations regarding the prospective rate of growth in the decade ahead. But there were also grounds for caution and concern. Indeed, a new recession seemed to be developing in the closing months of 1960. There continued to be a hard core of unemployment that kept the unemployment rate above 4 percent. (The seasonally adjusted rate was above 6 percent at the end of 1960.) The growth in output had not been particularly rapid since 1955. Total plant and equipment expenditures in 1960 were barely back to their 1957 peak; more rapid growth in business investment would be necessary if it were to continue to provide support for a rapid rate of growth in total output in the 1960's.

Prospective population trends were favorable, however, and so were long-term prospects for residential building.[23] Technological change and the steady expansion of consumers' demand were continuing to create investment opportunities on a substantial scale.[24] The prospects were for continued rise in state and local government expenditures at as rapid a rate as the financial resources of these governments permitted.

One question that might increasingly be raised as the 1960's wore on concerned the boom in commercial building. This, together with residential building, had been a major support to the economy through most of the 1950's, and it might well continue to do so for some years to come. But past commercial building booms had eventually ended, and such would probably be the eventual fate of this one. When and if that happened, the economy might face some serious readjustment problems—although nothing like the problems that would be created if the time should ever come for a large-scale disarmament program that wiped out the major part of federal defense expenditures.[25]

[23] See, for example, L. J. Atkinson, "Factors in the Housing Market," *Survey of Current Business,* April, 1960, pp. 16–22.

[24] For a highly (and probably overly) optimistic prediction regarding investment opportunities, see the series of articles on the capital-goods outlook for the 1960's in *Fortune,* August–December, 1958.

[25] Another question might be raised regarding investment in the important fuel and power industries (electric and gas, oil, and coal), which accounted for a quarter of all plant and equipment expenditures during 1956–1960. Investment in

Inflation also continued to be a problem as the economy faced the 1960's. The 1958 recession had only retarded, not stopped, the rise in prices that had begun in 1956. Wages were still showing a seemingly inexorable tendency to rise faster than labor productivity.

Despite these concerns about the future, the record of the preceding 15 years suggested that the American economy was much more stable than it had been before World War II, even if the rate of growth might sometimes be slower and the rate of unemployment higher than was generally desired. In Chapter 8, we suggested the main reasons why this was so. Much credit must be given to the automatic stabilizers. But clearly more than this was involved. There were other respects in which the economy had become more resistant to deflationary forces. Among these were the greater stability of our financial institutions, the growing importance of sectors in which employment and incomes do not react sensitively to a decline in aggregate demand (government and the service trades, for example), the steady rise in population that helped to bolster consumption in recessions, the high and rising level of government expenditures, and, although this may be debatable, the tendency for wages to rise even in recessions. Perhaps most important of all, public and private investment opportunities remained large enough to support a rising trend of aggregate demand and to make it relatively easy for the economy to recover quickly from the cyclical contractions that did occur.

A final stabilizing factor was the federal government's readiness to act in recessions in response to the commitment undertaken in the Employment Act of 1946 and the effect that this commitment had on business and consumer expectations. Actually, the federal government had done less than it might have in the postwar recessions, and the instability of federal (particularly defense) spending had had something to do with bringing on some of the postwar contractions. But, on net balance, government policy had been a stabilizing factor and could certainly be counted on to continue to be so in the future.

---

this broad sector flattened out markedly after 1957, chiefly because of the substantial decline in petroleum investment. It is also worth noting that this type of investment seemed to be becoming more sensitive to cyclical contractions. See M. F. Foss and Walter Leibowitz, "Investment in Fuel and Power Industries," *Survey of Current Business,* June, 1960, pp. 18–24.

While at the time this is written it is too early to say, it is possible that the American economy was in the downward phase of a mild Kuznets swing as the 1950's ended, one which showed itself in some retardation in the rate of growth in total output.[26] Such retardation was associated particularly with the slow rate of expansion of private investment after the boom of 1955–1957. It remained to be seen how long the retardation would last and whether the economy could continue to avoid anything significantly worse than the minor contractions of 1945–1960.

[26] See the discussion of shorter secular movements in Chapter 9.

PART III

PREDICTION AND CONTROL

CHAPTER 17

FORECASTING BUSINESS ACTIVITY

ALL economic planning—whether by business, by government, or by consumers—involves making assumptions regarding the future. We cannot escape having to make forecasts. This is true of the business firm budgeting its operations for the coming year; it is equally true of the government as it tries to plan its manifold activities for some period into the future.

In particular, both business and government must make assumptions regarding the future course of business activity. But the question immediately rises: Is accurate business forecasting possible? This is the question to which we shall address ourselves in the present chapter. We shall limit ourselves to the subject of *business* forecasting—how to predict future movements in the level of business activity, as measured in one or more of the ways described in Chapter 7. We shall make no attempt to deal with specialized forecasting techniques that have been developed in order to predict the behavior of specific economic variables, such as common stock prices or the prices of separate commodities. We shall, however, comment briefly on how general business forecasts can be used in attempts to estimate the future level of sales for a particular industry or firm.[1]

[1] The number of useful books on economic forecasting has shown a large increase in recent years. Among the more recent ones, see A. G. Abramson and R. H. Mack, *Business Forecasting in Practice,* 1956; V. L. Bassie, *Economic Forecasting,* 1958; E. C. Bratt, *Business Forecasting,* 1958; and J. P. Lewis, *Business Conditions Analysis,* 1959. The books by Bassie and Bratt also contain some discussion of forecasting techniques for particular economic sectors or variables, such as commodity prices, sales and output in particular industries, stock prices, etc.

IS ACCURATE BUSINESS FORECASTING POSSIBLE?

DEVELOPMENT OF BUSINESS FORECASTING

The study of business forecasting became especially popular after World War I. A number of private agencies developed to provide clients with analyses of current economic conditions and forecasts of the future course of business, and some large corporations set up research departments of their own to carry on this sort of work. With the increasing availability of statistical data, the contributions made by W. M. Persons and others to the analysis of economic time series, and the growing interest in business-cycle analysis, many businessmen and economists came to believe that it was possible to predict fairly accurately the future ups and downs of the business cycle. This belief was encouraged by the sustained prosperity of the 1920's.

Even during the 1920's the leading forecasting services were able to predict the important cyclical turns with only a fair degree of success.[2] Then came the stock-market crash in 1929 and the Great Depression of the 1930's. The business forecasters of that period were unprepared for the catastrophic decline in business activity and stock prices that occurred. Confidence in our ability to forecast business fluctuations waned as the depression dragged on. Nonetheless, considerable progress was made in developing new and useful statistical data and in adding to our understanding of the causes of cyclical fluctuations.

More recently there has been a considerable revival of interest in the possibilities of accurate business forecasting, accompanied by a healthy awareness of what forecasting methods can and cannot do. Increasingly, it is being recognized that the best business forecasting is simply a careful, detailed, and sophisticated application of the lessons of business-cycle theory, that no method of forecasting can be expected to be right all the time, and that business forecasting cannot hope to succeed if it relies merely on a few mechanical, shortcut formulas.[3]

A number of reasons account for this new interest in forecasting.

[2] See G. V. Cox, *An Appraisal of American Business Forecasts*, rev. ed., 1930.

[3] This healthy attitude is now to be found in the writings on business forecasting by business economists as well as by those in the universities. See, for example, Wilson Wright, *Forecasting for Profit*, 1947; Leo Barnes, *Handbook for Business Forecasting*, 1949; and Abramson and Mack, *op. cit.*

There has been an enormous increase in our stock of useful statistical data, including an increasing amount of information on business and consumer expectations; and new analytical techniques have also been devised. The development of the income-expenditure approach in business-cycle analysis and the accumulation of a wealth of data on the components of the GNP have led to an increased understanding of the factors determining the level of aggregate demand—and to the hope that we might be able to predict the behavior of these components in the future. Simultaneously, work being done in the field of econometrics has spurred the search for systematic "patterns of relationship" that might explain the movement of some economic variables in terms of the current and past behavior of other variables. During World War II, economists began to construct "GNP models" or "national economic budgets," which attempted to present internally consistent estimates of the components of aggregate demand for some future period, given an array of assumptions about various government policies and other exogenous variables.[4]

These new forecasting techniques have had a mixed record. There have been some spectacular failures, of which perhaps the best known was the forecast in 1945 by some Washington economists that there would be serious unemployment in the immediate postwar period.[5] Business forecasters today fairly generally set their forecasts in the framework of some sort of GNP model; that is, they consider all the main components of aggregate demand and then call on various sorts of information to estimate the future behavior of

[4] Increasingly, also, business executives have come to appreciate the importance of forecasting in their business operations. Their sad experience during the 1930's, the new conditions created by World War II and its aftermath, and the growing attention paid to budgeting have all contributed to this development. For a survey of forecasting activities by a sample of business firms, see *Business Forecasting: A Survey of Business Practices and Methods,* published by the Controllership Foundation, Inc., 1950; also *Forecasting in Industry,* National Industrial Conference Board Studies in Business Policy No. 77, 1956.

[5] Cf. Michael Sapir, "Review of Economic Forecasts for the Transition Period," in Conference on Research in Income and Wealth, *Studies in Income and Wealth,* vol. 11, 1949, pp. 273–367; E. E. Hagen, "The Reconversion Period: Reflections of a Forecaster," *Review of Economic Statistics,* vol. 29, May, 1947, pp. 95–101; and E. C. Bratt, "A Reconsideration of Postwar Forecasts," *Journal of Business,* vol. 26, April, 1953, pp. 71–83. It should be added that many private postwar forecasts were in error also. See Leo Barnes, "How Sound Were Private Postwar Forecasts?" *Journal of Political Economy,* vol. 56, April, 1948, pp. 161–165.

these components. The worst failures seem to have occurred where the estimates of the components (such as consumers' expenditures) have been based on mechanical extrapolations of past relationships —for example, between consumption and disposable income. There is growing awareness that, while the components of the GNP provide the proper starting point for business forecasts, the future values of these components cannot be estimated mechanically on the basis of unchanging relationships that are presumed to have held for some past period. A wealth of quantitative and qualitative information must be used, in ways suggested by economic theory and a sophisticated and discerning familiarity with current economic affairs. There is some reason to hope that this sort of sophisticated and flexible forecasting procedure, what John Lewis has referred to as "opportunistic model building," can produce significant results.[6]

WHY FORECASTING IS DIFFICULT

The basic difficulty in trying to forecast future business conditions can be put quite simply. The future is unlike the past, and no two business cycles ever repeat exactly the same pattern. Even if we grant that the response mechanism of the economy does not change radically with the passage of time, we can never be sure that we know how consumption or investment will respoud to new stimuli whose effects we have not been able to observe before. The forces making for change at any time are many and complex; they are not always easy to discover and measure; and they may occur in all sorts of combinations. Thus, the past can never be a perfect guide to the future, and forecasting must always remain a somewhat hazardous business.

The only kind of business forecasting that is likely to be reasonably successful represents detailed and systematic analysis of recent and current economic developments. Such analysis should start with a well-developed theory of economic fluctuations. Equally important, it must be based on a thorough familiarity with the relevant statistical data (and any significant interrelationships they have shown in the past) and on a lively appreciation of how nonmeasurable economic and political developments are likely to affect various aspects of the business situation. Business forecasting will probably always remain something of an art. It requires a skilled blending of

[6] Cf. Lewis, *op. cit.*, pp. 367–368.

theory and practice; it entails knowing which new facts are important and which are not, when we can rely on past relationships and when we cannot. Business forecasting should never become the mechanical application of certain theoretical ideas to an unchanging list of variables.

Forecasts can extend for various periods into the future. We can distinguish between short-run forecasts, which are concerned with the course of the business cycle, and long-run forecasts, which attempt to predict the secular movement underlying cyclical fluctuations. Short-term forecasts may be concerned only with the month or year ahead, or they may be extended in order to predict the full duration and amplitude of the current phase of the cycle. Obviously, predictions become more hazardous the longer they are extended into the future. In the rest of this chapter we shall be concerned chiefly with short-run forecasting for, say, not more than a year ahead, but we shall also pay brief attention to methods of making long-term conditional forecasts or projections.

METHODS OF FORECASTING

Most of the techniques that have been used to forecast business activity can be put into one of three categories.[7] The first, which we can dispose of quickly, might be termed the naive method. This consists of either simply projecting into the future what has been happening in the recent past or else taking as one's own forecast some sort of concensus as to what other "experts" think will happen. The simplest type of naive forecast is to predict that the level of business activity next year will be the same as it is now. Or one might predict that the GNP will increase next year by the same amount that it increased last year. We also include in this naive group the procedure of taking a poll as to what others think, where no attention is paid to how the polled predictions were put together.[8]

[7] For a somewhat different classification of forecasting methods, see C. F. Roos, "Survey of Economic Forecasting Techniques," *Econometrica,* vol. 23, October, 1955, pp. 363–395.

[8] For an evaluation of such polls of expert opinion, see Bassie, *op. cit.,* pp. 112–116. We do *not* include in this category the use of the growing amount of information about businessmen's and consumers' expectations and spending plans. Such evidence is valuable raw material that can be used in what we shall call analytical forecasting.

The second method of forecasting is sometimes called the barometric method. It rests on the assumption that past historical patterns will repeat themselves in the future. The past is searched for types of statistical behavior that seem to have been regularly associated with fluctuations in business, and these relationships are used to predict the future course of business activity. A simple example of this method is the use of statistical series whose turning points are supposed to lead the turning points in general business. If, for example, stock prices have usually begun a cyclical decline several months before the turning point in general business activity, this past relationship might be used to predict future movements in business.

The third method of forecasting can be called the analytical. It relies on a more or less detailed analysis of the recent and present business situation; hypotheses regarding causal relationships are explicitly or implicitly introduced in examining current data; and, on the basis of this analysis, judgments are reached as to the probable future course of business. The analytical approach to forecasting can take a variety of forms. It may be as unsophisticated as merely making a list of all the factors in the current situation thought to be favorable and unfavorable (on the basis of some combination of intuitive judgment and conscious analysis) and then striking some sort of balance between them. Or the analytical approach may involve the highly sophisticated use of econometric models. Most applications of the analytic method fall between these two extremes.

The big advance in analytical forecasting in recent years has come from the use of the income-expenditure approach and the resulting emphasis on studying the factors influencing each of the major components of the gross national product. In effect, forecasting is done within the framework of the national income accounts. Both quantitative and qualitative information can be used. Historical patterns are not ignored, and use can be made of past statistical relationships in order to estimate future values of some of the components of aggregate demand. In particular, past relationships with disposable income have been used to forecast consumers' expenditures.

Enough has already been said to indicate that we should be on our guard against the first two of these methods of forecasting. The objections to primary reliance on the naive method are self-evident.

As for the barometric method, with its reliance on the past behavior of various statistical indicators, here also we have to be careful. The future movement of business activity cannot be predicted from simple historical patterns that seem to have held in the past. Forecasting mechanically on the basis of a few simple formulas is almost certain to give poor results a good part of the time. There is no easy road to successful forecasting. That is why we shall emphasize the need for careful and detailed analysis of the whole business situation—that is, for the analytical type of forecasting. But first we shall review briefly some of the barometric techniques that have relied on past historical patterns.

HISTORICAL PATTERNS AND ECONOMIC BAROMETERS

PERIODIC CYCLES

One of the most simple and unsophisticated forecasting methods is that which assumes a periodic cycle repeating itself with the same duration over and over again. The myth of the absolutely periodic cycle refuses to die, and shortly after World War II it was put forward again in a book that had a wide sale among businessmen.[9] The authors, E. R. Dewey and E. F. Dakin, believe that business fluctuations are influenced by at least four major rhythms or periodic cycles, which continue to repeat themselves with unchanging duration. These major rhythms are the Kitchin, Juglar, Kondratieff, and building cycles described in Chapter 9; and to them Dewey and Dakin assign durations of 3½, 9, 54, and 18 years, respectively. Projecting these periodic cycles into the future, they predicted a depression of some severity for the early 1950's.[10]

Detailed criticism of this approach is unnecessary, particularly after the discussion in Chapters 9 and 10. Business cycles are not strictly periodic; there is no conclusive evidence that a long wave of 50 to 60 years exists; and we have seen that major and minor cycles can be combined in all sorts of ways. Indeed, as we saw in Chapters 15 and 16, there has been no complete major cycle in the United States since 1933.

[9] E. R. Dewey and E. F. Dakin, *Cycles: The Science of Prediction,* 1947; reprinted with a "1950 Postscript" in 1949.

[10] See, for example, the chart on pp. 188–189 of their book, which shows a low point for all four "major rhythms" in 1951–1952. See also the "1950 Postscript," written in 1949, which seems to assume that the 1949 recession would continue into the early 1950's.

LEAD-LAG RELATIONSHIPS

A number of forecasting methods have relied on the study of leads and lags among important statistical indicators. This, of course, is a highly persuasive approach. If we can find one or more series whose cyclical turns in the past have regularly preceded the turning points in business, what is more natural than to use these leading series to predict what will happen to business activity? Unfortunately, forecasting is not that simple.

A variety of statistical indicators have been used in this way. One of the best known is stock prices, which tend to reach their cyclical peaks and troughs before the turning points in business activity. This relationship has formed a part of a good many forecasting methods, notably the Harvard-Brookmire method, which attracted wide attention after World War I. Actually, though stock prices do *tend* to lead business activity, they are of only limited value as a forecasting device. The lead is highly variable; on occasion it fails to appear; and, in addition, stock prices go through irregular fluctuations of their own that make it difficult to determine when a significant business-cycle turn has actually occurred.

The most famous of the now discredited forecasting sequences is that which depends on the lead-lag relationships between the stock market, business activity, and the money market. This method was used as early as 1911 by the Brookmire Economic Service, but it was most fully developed by the Harvard University Committee on Economic Research at the close of World War I.[11] The result was the famous Harvard "Index Chart" or "Index of General Business Conditions," consisting of three separate indices representing speculation, business, and the money market.[12] These were usually referred to as the "A," "B," and "C" curves, respectively. For the period before World War I it was found that the "A" curve (speculation) regularly led the "B" curve (business), with the "C" curve lagging behind the latter. Thus, an upturn in the speculation curve, par-

[11] Cf. Harvard University Committee on Economic Research, *The Harvard Index of General Business Conditions: Its Interpretation*, 1923; C. O. Hardy and G. V. Cox, *Forecasting Business Conditions*, 1927, chap. 7.

[12] The composition of these indices was changed from time to time. As last revised, the speculation curve was based on stock prices; the business curve represented bank debits of selected cities outside New York; and the money curve was based on short-term interest rates.

ticularly if accompanied by a fall in the money curve, was taken to mean that the business curve would rise in a few months. When the "A" curve turned down, particularly if the money curve was high or rising, a recession in business was predicted.

This sequence did not work as well during the 1920's as it did before World War I, and the Harvard Economic Service found it increasingly necessary to rely on other methods of forecasting also—for example, analysis based on a continuous review of the current business situation. The Harvard service failed to survive the Great Depression, though the index continued to be published in the *Review of Economic Statistics* until 1941. Thus, this method of forecasting, begun with high hopes and under eminent auspices, has fallen into disuse.

THE NATIONAL BUREAU'S STUDY OF STATISTICAL INDICATORS

Various other statistical series have been used to forecast turning points in business activity—for example, bank deposits, the ratio of loans to deposits, bond prices, building activity, sensitive commodity prices, and so on.[13] By far the most careful work on "statistical indicators" has been done by the National Bureau of Economic Research. This aspect of the National Bureau's work in recent years has been particularly associated with the name of Geoffrey Moore.[14] Two features of this work call for discussion here: the compilation of a list of individual series or "indicators" that consistently lead, coincide, or lag at cyclical turning points, and the development and use of a particular kind of measure called a "diffusion index."

Table 37 presents the list of 21 statistical indicators of revival and recession that Moore has prepared.[15] These series are divided into three groups, depending on whether they tend to lead, coincide, or lag at cyclical turning points. For a series to be included, it had to represent an important economic variable that showed a definite

[13] Cf. Roos, *op. cit.*, pp. 369–379.

[14] See G. H. Moore, ed., *Business Cycle Indicators,* 1960, which contains a number of papers by Moore and others, including his earlier *Statistical Indicators of Cyclical Revivals and Recessions,* National Bureau of Economic Research Occasional Paper 31, 1950. One paper in this volume, by W. A. Beckett, develops a set of statistical indicators for Canada.

[15] The list has been revised somewhat since this table was prepared. See *Fortieth Annual Report* of the National Bureau of Economic Research, 1960, p. 35.

TABLE 37. Average Timing of Selected Statistical Indicators before 1957 and Their Timing in the 1957–1958 Recession [a]

Series	First Year of Period Covered	Number of [b] Reference Peaks	Number of [b] Leads at Peak [c]	Number of [b] Reference Troughs	Number of [b] Leads at Trough [c]	Median Lead (−) or Lag (+) (Months) Peak	Median Lead (−) or Lag (+) (Months) Trough	Lead or Lag at 1957 Peak	Lead or Lag at 1958 Trough [d]
Leading Series									
Liabilities of business failures [e]	1875	20	13	21	16	−6.5	−7	−26	0
Industrial stock prices	1871	21	15	21	14	−3.5	−6	0	−5
New orders, durable goods	1920	9	6	9	6	−5.5	−2	−8	−2
Residential building contracts	1915	10	6	10	6	−14	−6	−31	−2
Commercial and industrial building contracts	1919	9	5	10	5	−8	−2	−16	+2
Average work week, manufacturing	1920	8	5	9	5	−6.5	−4	−20	−2
New incorporations	1860	24	13	24	16	−2	−6	−17	0
Sensitive wholesale prices	1892	17	10	17	9	−2	−1	—	−5
Roughly Coincident Series									
Nonagricultural employment	1889	18	5	18	7	0	0	0	0
Unemployment [e]	1929	6	2	6	0	−1.5	+1	−4	+4

Outside bank debits	1875	20	4	21	14	+2	−3	+1	−1
Freight carloadings	1867	22	6	23	6	+1	0	−15	+1
Industrial production	1889	18	4	18	9	0	−1	−5	0
Wholesale prices, excluding farm products and foods	1913	11	3	11	2	−0.5	+1	—	—
Corporate profits [q]	1920	9	5	9	5	−2	−1.5	−20	−2
Gross national product [q]	1921	8	1	9	4	0	−1	+1	−2
Lagging Series									
Personal income	1921	8	2	9	6	+1	−1.5	+1	−2
Retail sales	1919	9	0	10	2	+3.5	+1	0	−1
Consumer installment debt	1929	6	0	6	1	+4.5	+2.5	+6.5	+5.5
Manufacturers' inventories	1929	6	1	6	0	+3.5	+2.5	+1.5	+5.5
Bank rates on business loans	1919	9	1	10	0	+5	+6	+5	+2

[a] From G. H. Moore, "The 1957–58 Business Contraction: New Model or Old?"; paper presented at the annual meeting of the American Economic Association, December 27, 1958 (mimeo.). These figures were not included in the version of the paper subsequently published in the *American Economic Review*, vol. 49, May, 1959.

[b] Excluding 1957 peak and 1958 trough.

[c] Leads of three months or less are excluded.

[d] Some of these leads and lags are tentative since they were determined before the end of 1958. The reference trough was in April, 1958.

[i] Series is inverted.

[q] Quarterly data only.

tendency toward a persistent timing relationship with the turning points in general business at both peaks and troughs.[16]

The leaders turn out to be liabilities involved in business failures, stock prices, new orders for durable goods, residential building contracts, commercial and industrial building contracts, hours worked per week, new incorporations, and sensitive wholesale prices. Although these series clearly tend to lead at both peaks and troughs by a significant period, no series shows an invariable lead at all turning points in general business. Every leading series shows some cases of lags or rough coincidences to go with the more frequent cases of clearly marked leads.

These indicators were chosen on the basis of the consistency of their timing in business cycles before World War II. They have continued to perform in much the same way since the war. The prewar leaders have continued to lead, and the prewar lagging series have continued to lag—and, in both cases, about as consistently as before the war. Interestingly, however, the postwar leads have, on the average, been longer, and the lags shorter, than the corresponding prewar averages.[17] We also found evidence of longer postwar leads in some of the reference cycle patterns we examined in Chapter 10 —for example, in inventory investment.

It is interesting to see whether the list of statistical indicators in Table 37 would have helped to predict either the downturn in business in July, 1957, or the revival that quickly followed after April, 1958. The necessary information is given in the last two columns of Table 37.

Six of the eight leading series did turn down before the peak in business in July, 1957. But five of these leads were a year or longer, and two of them were over two years. Also, several of the coincident series showed significant leads. All of this, as a matter of fact, is consistent with what we discovered in Chapter 16 about the characteristics of the boom during 1955–1957. The economy was poised on the brink of a downturn for an unusually long time before the recession actually began.

Another aspect of Table 37 is more interesting and important

[16] For a more detailed discussion of how the series were chosen, see Moore, *Statistical Indicators, op. cit.*, pp. 63 ff.

[17] See G. H. Moore's report in *Fortieth Annual Report* of the National Bureau of Economic Research, pp. 33–35.

than the behavior of particular leading series. Although we cannot trust the average lead shown by one or a few series, Table 37 brings out that there is a persistent relationship among groups of series. On the whole, the leading series did tend to lead, and the coinciding series, on the average, turned down later. If an observer had waited until half the coinciding series had turned down before finally accepting the forecast of the leading series, he would have been prepared to predict a downturn about the time, or a little before, it happened. Half of the coinciding series had turned down by March, 1957. Given the delay in the publication of data and in recognizing the turning point in a series once it has occurred, our observer might have been ready to make his forecast just about the time the general downturn began, perhaps a month earlier.

The indicators also performed fairly well at the lower turning point in 1958. Five of the leading indicators led by from two to five months; only one lagged. Five of the eight coinciding series had turned up by the time general recovery began. Thus, allowing for delays in publication and in the recognition of turning points, an observer would have been able to predict an upturn by about July. This was three months after what we now recognize to have been the low point, but this is not a bad record. It was by no means clear to all observers in the summer of 1958 that the improvement in business during the preceding two or three months meant that a cyclical expansion had definitely begun.

Moore has also used these indicators to secure an early judgment as to how severe a recession is likely to be. For example, the magnitude of the declines in the leading indicators during the first four months or so of a recession gives some indication as to how severe the full contraction is likely ultimately to be.[18]

We now come to the National Bureau's use of diffusion indices.[19] If we take a large number of series and tabulate the percentage of the total number that are expanding each month, we get the results

[18] Cf. G. H. Moore, *Measuring Recessions,* National Bureau of Economic Research Occasional Paper 61, 1958.

[19] See Moore, ed., *Business Cycle Indicators,* Part I, especially his paper, "The Diffusion of Business Cycles," which appeared originally in R. A. Solo, ed., *Economics and the Public Interest,* 1955. For critical evaluations and further references, see S. S. Alexander, "Rate of Change Approaches to Forecasting—Diffusion Indices and First Differences," *Economic Journal,* vol. 68, June, 1958, pp. 288–301, and A. L. Broida, "Diffusion Indices," *American Statistician,* June, 1955, pp. 7–16.

shown by the series designated "diffusion index" in Figure 43. The number of series expanding regularly begins to increase considerably before the trough in business and reaches its peak well before the peak in business. This is no more than we should expect. In the later stages of an expansion, as the economy becomes more vulnerable, an increasing number of series reach their peaks and begin to decline. This is the story told by the diffusion index in Figure 43. Similarly, as the downswing progresses, the number of series expanding declines; but then, as maladjustments in various parts of

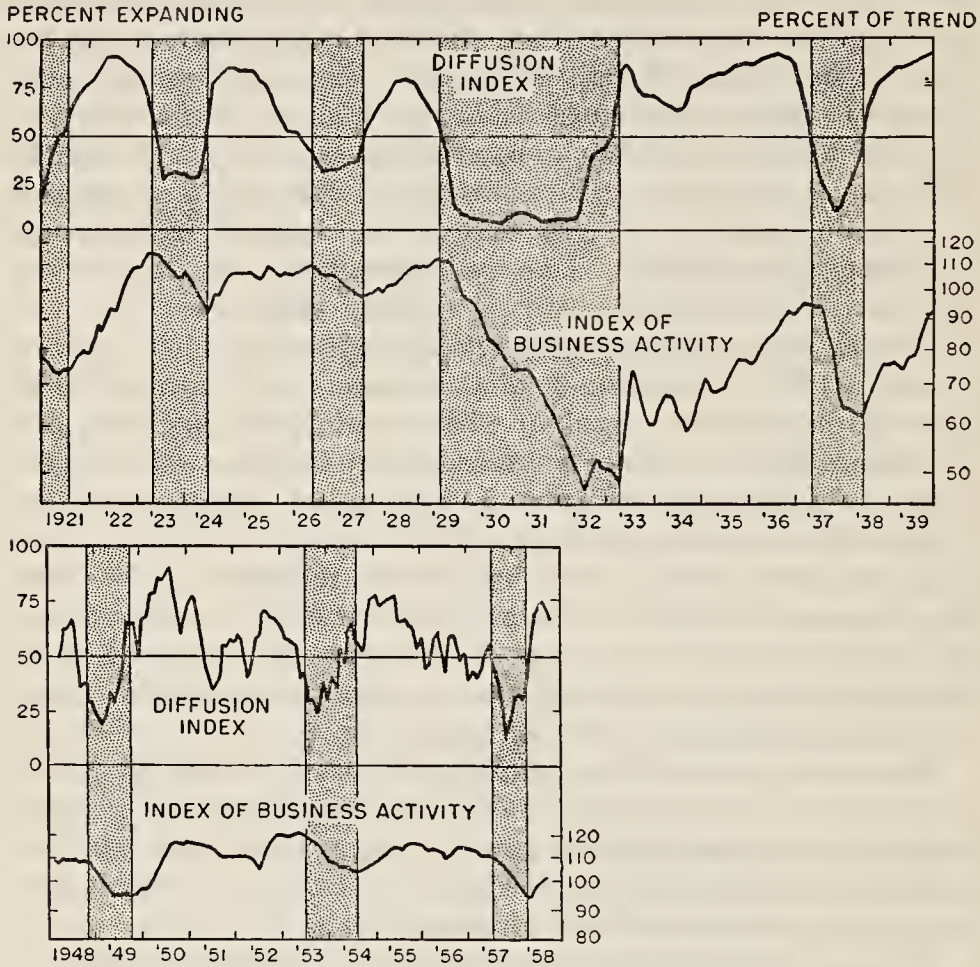


FIGURE 43. Diffusion Index and Index of Business Activity, 1921–1958.

From G. H. Moore, "The 1957–58 Business Contraction: New Model or Old?" *American Economic Review,* vol. 49, May, 1959, p. 298. By permission of the National Bureau of Economic Research, and the American Economic Association. Shaded areas represent business contractions; white areas, business expansions.

the economy are corrected, the number begins to rise again while overall measures of business activity are still declining.[20]

The leads shown by the diffusion index are quite striking, though of irregular duration. In no case was there a lag. If an observer had, during the period 1921–1958, used this series as a preliminary forecasting device but had waited to date the turn until the percentage of series expanding had risen or fallen to the 50 percent line, he would have had a fairly good forecasting record. His chief trouble would have come from the false signals given by the diffusion index in 1951 and 1956. As we saw in Chapter 16, there were special circumstances in both these years that led a number of series to turn down while general expansion was still in progress.

The reader should not jump to the conclusion that we have here an infallible forecasting device. The leads are irregular; use of the 50 percent criterion may lead to important delays in recognizing cyclical turns; further delays are involved if we must wait for the publication of a very large number of series and wait further until we can clearly recognize when each series in our sample has reached a cyclical turn; and occasionally there may be false signals. Nor does this method by itself tell us anything about the *amplitude* of the phase that we are trying to predict. There is no doubt, however, that we have here a technique deserving of further study and refinement. At least one large industrial company has experimented with a forecasting technique based on the sort of information shown in Figure 43,[21] and Moore and others have experimented with ways of summarizing promptly the current behavior of groups of indicators such as those shown in Table 37.[22] It is noteworthy that so judicious a person as Arthur F. Burns has stated that findings such as those summarized in the preceding paragraphs may "prove helpful in

[20] The group of series used in Figure 43 is not made up only of those that usually lead. Lagging and coinciding series, as well as some with irregular timing, are also included. It is possible to prepare a chart like Figure 43 for a sample of series all of which typically lead at business-cycle turns. In this case, a curve showing the percentage of series expanding would lead the comparable curve in Figure 43 and would cross the 50 percent line sooner. However, the smaller and more narrowly selected the sample, the greater are the chances that any particular forecast will be in error.

[21] The company is Standard Oil Company (New Jersey). See C. Ashley Wright, "Business Cycle Research and Business Policy," in Universities–National Bureau Committee, *Conference on Business Cycles*, 1951.

[22] Cf. *Business Cycle Indicators, op. cit.*

predicting reversals in the direction of total economic activity—or at least in identifying them as such promptly."[23]

Diffusion indices have considerable analytical value beyond their possible use in short-term economic forecasting. They provide a direct way of measuring one important aspect of business-cycle fluctuations—that is, their pervasive character. Thus, a diffusion index portrays the gradual spread of, say, a cyclical expansion to include a steadily wider area of the economy; it also records the early beginnings of a contraction as first a few and then a growing number of firms, industries, or types of activity stop expanding. Diffusion indices can be computed for particular types of activity—for example, profits, production, employment, prices, and so on. Thus, a diffusion index for production might show the percentage of the total number of industries represented in the Federal Reserve Index of Industrial Production that were expanding.

ANALYTICAL FORECASTING

"Analytical forecasting" is simply a brief way of referring to what people do when they make a more or less thorough study of the current business situation in order to decide how business activity is likely to behave in the near future. The level and kind of economic and statistical analysis involved in this sort of study of the business situation can vary over a wide range. Virtually all business forecasters use this method in some form or other, either as the entire basis of their forecasts or to supplement their own special formulas. As we have suggested, business forecasting must rely heavily on this method if it hopes to be even moderately successful. Even the promising work on statistical indicators summarized in the preceding section can scarcely do more than supplement the careful study of current business conditions.

THE GNP AS A FRAMEWORK

To an increasing degree, economic forecasting is being carried on in terms of the components of the gross national product. In effect, the attempt is made to forecast each of the main components of aggregate demand—consumers' expenditures on durable and nondurable goods and on services, the various types of domestic private in-

[23] In *Thirtieth Annual Report* of the National Bureau of Economic Research, 1950, p. 27; reprinted in *Business Cycle Indicators.*

vestment, net exports, and government expenditures. Framing the forecast in terms of the GNP statement has numerous advantages. It insures that all the components of aggregate demand will be considered; it facilitates the use of modern business-cycle analysis in forecasting; it encourages the analyst to look for the fundamental influences operating on each component of aggregate demand; and it insures that the various parts of the forecast will be consistent with each other. As one business economist has put it: "The pivotal starting-point of economic change is a shift in the *spending decisions* of one or more of the four main spending groups—consumers, business, government and foreign buyers. . . . Successful forecasting must start by evaluating probable changes in the future spending patterns of these four major groups."[24]

Forecasting the components of the GNP can be done in various ways. One may, for example, merely make a series of horseback guesses about the future course of each major type of spending, based on general impressions as to the forces operating on business investment, consumers' spending, etc. At the other extreme, use can be made of detailed econometric models, in which various components of the GNP are made to depend on other variables in accordance with patterns of relationship that seem to have held in the past.[25] The best-known of these relationship patterns is the consumption function, which we dealt with in some detail in Chapter 5. Other relations that may be used are, for example: the relation between disposable income and GNP, the relation of corporate profits to GNP, the relation of business investment to profits and other variables, the relationship between GNP and employment, and so on.

None of these relationships is so stable that we can use it without question and without the test of other evidence. We have already seen that the short-term relationship between consumption and disposable income is subject to significant changes. Even when we do feel that we can make use of these relationships, we must be clear as to what we can and cannot do with them. Suppose, for example,

[24] Leo Barnes, *Handbook for Business Forecasting*, p. 5. Italics his.

[25] See the brief discussion of econometric model-building in the Appendix to Chapter 13. The Klein-Goldberger model described on page 392 has been used to make current forecasts by a group at the University of Michigan. See, for example, A. M. Okun, "A Review of Some Economic Forecasts for 1955–57," *Journal of Business*, vol. 32, July, 1959, p. 210.

that we want to forecast consumers' expenditures for the next quarter, using the best available estimate of the past relationship between consumption and disposable income. But in order to estimate consumers' spending for the next quarter, we must first estimate the next quarter's disposable income. On what variables does disposable income depend, and how do we go about forecasting the future values of these variables? We seem to be in a vicious circle, in which everything depends on everything else—which, in an important sense, is really the case.

In general, there are two ways out of this dilemma. One is by use of *lagged* relationships. If, for example, next quarter's consumption depends on *this quarter's* disposable income, which is known, then we have found a way of breaking out of the circle. The other way is to assume, as a first approximation, that certain key variables are exogenous—that is, that they are not determined by the variables in our model. We estimate these exogenous variables on the basis of whatever information is available, and then use the estimates to help determine the other, endogenous variables. As a final check, we should then see if our original, independent estimates for the exogenous variables are consistent with the final estimates for the GNP and for the various components derived on the basis of past relationships. For example, we may estimate private investment and government spending independently as a starter, but the final forecasts for the GNP and for corporate profits that eventually result may not be entirely consistent with this level of investment. In that case, some sort of reconciliation will have to be made.

In practice, it is always necessary to consider some variables to be exogenous and to estimate them on the basis of whatever information is available. This is so even if lagged relationships are also used, so that the future values of some GNP components are estimated on the basis of the known present values of other variables.

A great variety of information can be used as a basis of forecasting the various components of aggregate demand, and in this connection the reader is advised to refer again to the array of partial business indicators briefly described in Chapter 7 (pages 175–177). Much of this information can be used in predicting the future behavior of different components of the GNP. For example, data on new orders, sales, and shipments, as well as on the current course of prices, are helpful in predicting the future behavior of inventory invest-

ment. The surveys made by the Department of Commerce and the S.E.C. regarding business plans for future expenditures on plant and equipment are our chief source of information regarding this important component of aggregate demand. Various sorts of information in the building field are helpful in judging the probable behavior of residential construction. And so on.

THE PROBLEM OF FORECASTING PRICES

Business forecasts entail some consideration of the probable future course of prices. One difficulty with some of the past forecasts based on GNP models was that not enough attention was paid to the probable behavior of prices.[26] Changes in the price level affect the components of the GNP in at least two important ways. First, if prices change, the behavior of money expenditures will not reflect accurately the physical flow of goods and services. Thus, if retail sales rise 5 percent but retail prices increase 10 percent, the physical volume of sales will have declined. Obviously, we should be making a mistake if we concentrated our attention exclusively on the dollar figures without any reference to the behavior of prices.

Second, price changes are important because some economic variables are closely associated with the rate of change, or at least the direction of change, in prices. This is particularly true of inventory accumulation. It may also be true of other types of business spending and, to a less extent, of some types of consumers' spending. If we are preparing a forecast for the next three months and fail to allow for the fact that prices will, say, rise during this period, we may seriously underestimate the amount of inventory investment and other forms of anticipatory buying that will probably occur.

Allowance for price changes in a general business forecast is best done in a series of steps. First, preliminary forecasts can be made for the various components of aggregate demand without reference to the possibility of future price changes. This estimate of aggregate demand can be weighed against estimates of potential supply to secure a preliminary impression of the probable course of prices. This can also be done for specific fields—for example, agriculture, in which prices are likely to be particularly sensitive to changes in de-

[26] For an interesting example of this in forecasts for the years 1955–1957, see Okun, *op. cit.*, pp. 205, 208. There was a general tendency in these years to underestimate the rise in prices.

mand or supply. Against this background, account should then be taken of possible "autonomous" price-making influences. For example, are important labor contracts expiring, and is the underlying situation, as we have evaluated it, such as to suggest that substantial wage increases will be granted?

Having come to a final conclusion regarding the probable direction of change in prices and wages, and having made a very rough estimate of the amount of the change, we can then adjust our preliminary estimates of the GNP components to allow for business and consumers' reactions to this sort of price behavior. We can then also estimate the physical volumes that are involved in our adjusted estimates of dollar expenditures.

METHODS OF FORECASTING THE GROSS NATIONAL PRODUCT

As already indicated, there are various ways to approach the problem of forecasting the components of the GNP and other economic variables on which these components may depend. In this section we shall summarize the method that seems to be most commonly used.

This method begins with independent estimates of private investment and government expenditures, and then proceeds to estimate simultaneously consumption and total GNP with the help of the consumption function derived from past data. The method can be illustrated by a simple example. Let us assume that, on the basis of whatever information is available, we estimate that next year private investment (I) will be 80 billion dollars and government expenditures (G) 95 billions. Assume further that we have found that consumers' expenditures (C) are related to total GNP by the equation:[27]

$$C = 25 \text{ billions} + .6\ GNP.$$

We have already estimated I and G and we know that $GNP = C + I + G$. We can substitute $(25 + .6\ GNP)$ for C and write:

$$GNP = 80 + 95 + (25 + .6\ GNP).$$

Solving, we secure 500 billions as our *preliminary* forecast of the GNP for next year. By substituting this figure for the GNP in the

[27] Actually, as we saw in Chapter 5, the short-run relation between consumption and GNP is not at all stable. We shall take care of this problem in the more detailed discussion that follows.

consumption equation, we can also make a forecast of consumers' expenditures. Our estimates would then be checked and revised in ways to be discussed in the following paragraphs.

These are the bare bones of the method. Let us now follow through the steps in a little more detail, adding the necessary complications.

FORECASTING INVESTMENT

Each component of gross private domestic investment is estimated separately.[28] As we have seen, these components are chiefly business expenditures on plant and equipment, residential construction, and changes in business inventories. (Other private building, chiefly by nonprofit institutions, should also be included.)

Business expenditures on plant and equipment are usually estimated on the basis of the valuable quarterly surveys made by the Department of Commerce and the Securities and Exchange Commission and published regularly in the *Survey of Current Business.* These surveys yield information on *planned* expenditures by business for the current and succeeding quarters, and early in each year planned expenditures for the entire year are reported. The forecasting record of these surveys has been very good. An annual survey of business plans for expenditures on plant and equipment is also made by the McGraw-Hill Publishing Company and published in *Business Week.*[29]

What these surveys help us to predict are two items in the GNP accounts: expenditures on producers' durable goods (i.e., equipment) and nonresidential construction. The sum of these two items includes a few types of investment that are not covered by the plant and equipment surveys, chiefly farm construction and equipment purchases, construction by nonprofit institutions, and capital expenditures that firms charge to current expense and that therefore

[28] In the following discussion, investment is treated as an "exogenous" variable; that is, it is not estimated on the basis of past relationships with other variables. Some writers have attempted to set up such "patterns of relationship" for some part or all of private investment. See, for example, the Klein-Goldberger model described on p. 392 and Roos, *op. cit.*, pp. 390–394.

[29] There is a considerable body of literature evaluating the accuracy of these surveys of anticipated expenditures on plant and equipment. See in particular the papers in Part IV of Universities–National Bureau Committee for Economic Research, *The Quality and Economic Significance of Anticipations Data,* 1960. These papers contain numerous references to the earlier literature.

do not get included in the survey figures. Separate estimates must be made for these omissions, either on the basis of such direct evidence as is available or by some rule of thumb.[30]

With some sort of adjustment for omitted items, the S.E.C.–Commerce and McGraw-Hill surveys provide the basis for a preliminary forecast of future expenditures on producers' durable goods and nonresidential construction. These estimates can be further checked in various ways. One useful partial check is provided by the surveys of capital appropriations—i.e., of advance authorizations to spend—which are regularly published by the National Industrial Conference Board.[31] Available data on new orders for durable goods and on construction contracts provide some further evidence, and so does the scanty information available on capacity. In any event, the preliminary forecast must be checked again at a later stage and made consistent with our final forecasts regarding supply conditions, price changes, estimated profits, availability of capital funds, and so on.

Estimates of the prospective volume of residential building are available in various government and private publications.[32] A forecast can also be made directly. A direct forecast needs to proceed in at least two stages. First, the "underlying" demand for housing should be investigated. This involves estimates of prospective new family formation, possible undoubling of families in existing dwellings, the effects of migration (particularly from farm to city), and probable reductions in the housing stock due to accidental losses, demolitions, and so on. Secondly, account must be taken of various short-run influences that affect the current demand for new housing and the willingness of builders to undertake the construction of new dwellings. These include the existing stock of unsold new houses, the vacancy rate in rental properties, the prices of old houses, and, of particular importance as we saw in earlier chapters, prospective changes in credit conditions and in government regulations affecting mortgage financing.

A forecast of residential building usually first takes the form of

[30] The omitted items have remained a fairly constant percentage of the total of nonresidential construction and producers' durables.

[31] Published quarterly in *The Conference Board Business Record*. The sample includes manufacturing firms only.

[32] For example, toward the end of each year the Department of Commerce publishes a forecast for the next year in *Construction Review*.

an estimate of the number of housing starts, which must then be converted into an estimate of dollar expenditures. This conversion requires taking into account prospective building costs and possible changes in the average size and quality of dwelling units, as well as the short average lag of actual construction expenditures behind housing starts.[33] Here again, various sorts of information can be used as checks—for example, the behavior of residential building contracts and permits and applications for FHA and VA financing.

Of the different kinds of investment, the most difficult to forecast is change in inventories. Yet, as we have seen, net changes in inventories constitute a significant element in the short-run fluctuations in GNP.

What we are trying to forecast is the "change in business inventories" component of the GNP.[34] What we have to go on chiefly is the recent behavior of this item as recorded in the national income accounts, a body of data showing the total inventory holdings in manufacturing and trade, considerable information on new orders and sales, and computed sales-inventory ratios in various branches of manufacturing and in wholesale and retail trade.[35] Most of this information is regularly reported in the *Survey of Current Business*.[36] Inventories of durable goods are particularly subject to fluctuation, and it is important to pay particular attention to the inventory-sales ratio and to new and unfilled orders in this sector. For both durable and nondurable goods, and for both manufacturing and trade, present inventory-sales ratios should be compared with what might be considered to be "normal" ratios (given whatever long-run trends seem to be at work); and consideration should be given as to whether these ratios in recent months seem to have been moving toward or away from "normal." In this connection, it should be remembered that inventory-sales ratios tend to fall during

[33] Adjustment must also be made for the substantial amount of additions and alterations that are not included in the estimate of expenditures on new dwellings.

[34] In this sort of brief introductory treatment, we may ignore the "inventory valuation adjustment" which the Department of Commerce applies to the change in book value of inventories.

[35] For a careful evaluation of available inventory data, see *Statistics of Business Inventories: Report of Consultant Committee on Inventory Statistics,* Board of Governors of the Federal Reserve System, 1955.

[36] In addition, interpretive articles on the recent behavior of inventories appear from time to time in the *Survey*. See, for example, the issue for April, 1959, pp. 3–8.

at least the first part of a cyclical expansion and to rise during most or all of a contraction. As we saw in Chapter 10, total inventories tend to lag behind sales, although the lag since World War II has not been so pronounced as it was earlier.

On the basis of information of this sort, a rough forecast is made of the future behavior of total inventory holdings in manufacturing and trade, and from this one can derive a preliminary estimate of net investment in nonfarm inventories.[37] To this would have to be added an allowance for investment in farm inventories, which is ordinarily fairly small.

However the initial forecast of inventory investment is made, the estimate should be considered highly tentative and preliminary, to be revised upward or downward in light of the forecaster's final conclusions as to the future course of business and prices.

To these components of gross private domestic investment must be added a forecast of net exports of goods and services. This amounts to a prediction of the net surplus of exports over imports. Valuable surveys of the international transactions of the United States are presented periodically in the *Survey of Current Business,* and this and other current sources provide information on recent prospective changes in the foreign demand for American goods and on the current behavior of imports.

GOVERNMENT EXPENDITURES

Accurate data are available on the current expenditures of the federal government; and the government's budget provides a basis for projecting these expenditures into the future. This information is available in a number of current sources. When large new appropriations are made by Congress, as happened after the invasion of South Korea, it is necessary to predict when the resulting expenditures will be made. This is not always an easy task. In 1950–1951, military expenditures rose more slowly than was generally anticipated. As a result, some forecasts made in the fall of 1950 tended to overestimate the amount of government spending in the first half of 1951.

The basic source for projecting federal expenditures is the President's budget message, transmitted to Congress in January for the

[37] Nonfarm inventories outside of manufacturing and trade can safely be ignored since they account for less than 10 percent of the total.

fiscal year beginning the following July 1. The projections of expenditures in this document have to be modified to allow for Congressional action during the ensuing months. Here again, also, adjustments are needed to convert the data as originally presented into the form in which they appear in the GNP accounts. What we wish to estimate are government purchases of goods and services. To secure this figure we must adjust expenditures in the conventional or administrative budget to exclude transfers to trust accounts, grants-in-aid to state and local governments, transfer and interest payments to the public, and so on.[38]

Expenditures of state and local governments must be added to those of the federal government. Unfortunately, we do not have the authoritative, centralized projections with which to begin that we have for federal expenditures. However, since state and local government expenditures are fairly stable in the short run, they can be safely projected on the basis of the most recent available data. In recent years, this has meant projecting a fairly steady upward trend. On occasion, it may be desirable to adjust such extrapolations to allow for anticipated changes in tax revenues and in state and local bond issues.

It should be remembered that at this point we are estimating only expenditures on new goods and services. Government transfer payments are not included here but are dealt with at a later stage, when disposable income is related to gross national product.

FORECASTING CONSUMERS' EXPENDITURES

It is necessary to divide total consumers' spending into at least three categories—expenditures on durable goods, on nondurables, and on services. Of these three, as we know, the first is the most volatile and needs to be considered separately. The other two, nondurables and services, may be estimated separately or combined together.

Within the category of consumers' durables, it is expenditures on automobiles that create particular difficulties for the forecaster; and, if time and resources permit, it is wise to estimate this item by itself.

[38] See the detailed adjustments in U.S. Department of Commerce, *National Income,* 1954 ed., p. 146. See also pp. 595–596, below, where the distinctions among the administrative budget, the cash consolidated budget, and expenditures and receipts on income and product account are further elaborated.

If a single forecast is to be made for total expenditures on consumers' durables, then this estimate must be adjusted to allow for any special circumstances that are likely to make automobile purchases abnormally high or low.

The discussion in Chapter 5 of the relation between consumption and income suggests the general procedure that can be used to forecast consumers' expenditures. From past data, we can derive an equation relating expenditures on each of the three categories of consumption to disposable income. This will give us regression lines or "consumption functions" for each type of expenditure similar to the regression line for total consumption in Figure 15 on page 99.

There are good reasons for combining nondurables and services and computing one regression for all consumers' expenditures except those on durables.[39] Whether we treat nondurables and services separately or combine them, the estimated relationships to disposable income should be adjusted to whatever extent thought necessary. Relevant evidence would include the recent behavior of these expenditures in relation to disposable income, current data on the behavior of retail sales and prices, and so on. One factor in particular to take into account is that the ratio of these expenditures to disposable income is likely to be a bit higher in recession than during a boom period.

In the case of durables, let us assume that we make a single estimate for the total, including automobiles. Here again, we can start with a regression relating past expenditures on consumers' durables to disposable income. In this case particularly, it is necessary to adjust this relationship to allow for any special factors that we think are at work. If expenditures on durables and increases in installment debt were abnormally large in the recent past, this may suggest that the relation of expenditures on durables to disposable income will be lower than usual in the near-term future. Two types of general evidence are helpful in making these adjustments. One is the recent behavior of the relationship between spending on durables and disposable income. Do recent expenditures show a tendency to move toward or away from the regression line? The second type of evidence is contained in special studies of the current demand situation for various types of consumers' durable goods. Such studies are published from time to time in *Survey of Current Business* and are also available in other publications, including both

[39] Cf. Lewis, *op. cit.*, pp. 467–469.

trade journals and the general business and financial press. Also of some use in this connection are the annual Surveys of Consumer Finances conducted by the Survey Research Center at the University of Michigan as well as the surveys published in the Federal Reserve Bulletin.[40] These contain, among other things, reports on consumers' plans to buy automobiles and other durable goods.

We now have three adjusted equations (two equations, if nondurables and services are combined), each relating a particular type of consumers' expenditure to disposable income. The equations are then added to secure a new equation, which in effect represents a prediction of how *total* consumers' expenditures will be related to disposable income during the period for which the forecast is being made.[41]

RELATING DISPOSABLE INCOME TO GNP

So far we have related consumption only to disposable income. It is now necessary to link disposable income to GNP. This relationship is determined chiefly by the behavior of gross corporate savings, indirect and corporate income taxes, social security contributions, government transfer and interest payments, and personal taxes. Given information on present tax rates, dividend and depreciation policies, the probable course of government transfer payments, and so on, it is possible to construct an equation relating disposable income to GNP for the period for which the forecast is being made. This relationship must be revised from time to time to allow for changes in tax rates and other conditions that affect the way in which disposable income is tied to the gross national product.[42]

[40] A new periodic survey of consumer buying plans has recently been undertaken for the Federal Reserve Board by the Bureau of the Census. See also the surveys of consumer buying plans conducted by the National Industrial Conference Board, as described in the Conference Board's *Business Record,* vol. 15, November–December, 1958. The predictive value of surveys of consumers' plans and attitudes has been frequently debated. See, for example, *The Quality and Economic Significance of Anticipations Data, op. cit.,* and the other references cited there.

[41] An alternative is to treat automobile expenditures as autonomous and to make a dollar forecast for this type of spending directly, in the way that we estimate, for example, residential building. If we do this, our final adjusted equation for total consumers' expenditures will exclude automobile purchases, the estimated amount of which will be added to the total of investment and government spending.

[42] For a study of the relation between disposable income and GNP, see James Duesenberry, Otto Eckstein, and Gary Fromm, "A Simulation of the United States Economy in Recession," *Econometrica,* vol. 28, October, 1960, pp. 749–809.

We now have two equations of the following form:

$$C = a_1 + b_1 Y_D$$
$$Y_D = a_2 + b_2(GNP).$$

By combining these two, we can relate consumers' expenditures directly to GNP.

We now have a forecast of how consumers' expenditures will be related to GNP, and we also have direct estimates for investment and government spending. We can combine these estimates into a single equation in the way shown on page 526. Solving this equation provides our preliminary forecast of the GNP for the period in question. Given this estimate of the GNP, we can then use the other equations described to secure preliminary forecasts of consumers' expenditures and disposable income.

We can illustrate this procedure in the following hypothetical case. Let us assume that we have estimated the nonconsumption components of GNP to be as follows for some future calendar quarter.[43]

Producers' durables	30
Nonresidential construction	20
Residential building	20
Inventory investment	5
Net exports	3
Government expenditures:	
Federal	52
State and local	50
Total	180

After our various adjustments, let us assume that we believe that total consumers' expenditures will be 93 percent of disposable income and that disposable income will be 70 percent of GNP.[44] Then we can write:

$$GNP = 180 + .93(.70\ GNP).$$

[43] Annual rates, in billions of dollars.

[44] Putting these estimated relationships in average rather than marginal terms has some advantages besides simplicity of calculation. Thus, the average propensity to consume is more stable than the marginal propensity, and it is perhaps safer to forecast the future ratio of total consumption to total disposable income than to estimate the ratio of a change in consumption to a change in disposable income. For a more detailed discussion of this and other alternatives, particularly with respect to ways of relating disposable income to GNP, cf. Lewis, *op. cit.*, pp. 499–500.

Solving this, we arrive at a preliminary forecast of approximately 516 billion dollars for the GNP.

CHECKING THE PRELIMINARY FORECAST

As we emphasized at an earlier point, this sort of preliminary forecast must now be checked in a variety of ways to insure that estimates of the different components of the GNP are consistent with each other and with such other evidence as is available to the forecaster. The nature of some of these adjustments may be briefly indicated.

One useful check to make at an early stage is to note the amount of personal saving that is implied in the forecasts of disposable income and total consumers' expenditures. If the resulting volume of saving seems to be out of line with the behavior of personal saving in the recent past and with other evidence regarding the current spending propensities of consumers, adjustments in the consumption estimates may be necessary.

It is also necessary at this point to make a rough forecast of the probable behavior of prices. The preliminary forecast of the GNP represents an estimate of future aggregate demand. This can be compared with rough projections of supply to arrive at a provisional judgment as to whether the course of prices will be upward or downward. Other types of evidence should also be consulted—for example, the movement of sensitive wholesale prices, the current behavior of the money supply and its velocity, special conditions affecting agricultural prices, prospective wage negotiations, and so on. If all of this evidence points clearly to a significant change in prices, the preliminary forecasts of the components of the GNP will have to be revised. In particular, the forecast for investment in inventories may have to be altered in light of the predicted movement of prices.

One simple type of check is perhaps the most important of all. In what direction does the preliminary forecast of total GNP imply that business activity will move? Is this implied forecast of the direction of change consistent with all the evidence available regarding the current phase of the business cycle?[45] And are the forecasts for expenditures on plant and equipment and for investment in inven-

[45] The evidence provided by a selected group of statistical indicators such as those presented in Table 37 can be of value at this point.

tories consistent with this assumed rate of change in business activity?

A variety of other adjustments can be made in the preliminary forecasts. For example, the predicted value of the GNP implies a particular behavior of corporate profits. Are the estimates of investment consistent with this behavior of profits? The range of such adjustments and refinements is limited only by the facilities and time available to the forecaster. At some point, having made all his estimates as consistent with one another and with supplementary evidence as he can, the forecaster has arrived at his final prediction of the GNP and each of its major components.

The variety of ways in which it may be necessary to adjust the preliminary forecast suggests how far we are still from a completely objective and "scientific" method of forecasting. If we are prepared, as we must be, to alter our preliminary forecast because it seems to imply that business activity will move in a direction opposite to that suggested by other evidence we think is important, our forecasting technique is obviously highly subjective and open to a wide margin of error. Similarly, arbitrary adjustment of the various consumption functions represents the forecaster's subjective reaction to a range of evidence that cannot be compressed into stable functional relationships. The same is true of the other checks mentioned. This is at once the strength and the weakness of this sort of forecasting. As long as we do not know and cannot express quantitatively *all* the relevant relationships that determine economic behavior, we must use a flexible method that permits the factor of judgment to enter. But when we introduce the subjective element, new sources of error are created. There is no escape from this dilemma. As we said early in this chapter, forecasting will probably always remain something of an art.

ESTIMATING EMPLOYMENT AND UNEMPLOYMENT

It is desirable to translate forecasts of the gross national product into estimates of total employment and the probable volume of unemployment. This process calls for making estimates of man-hour output, length of the work week, and the civilian labor force. Current data on labor productivity and a projection of recent trends can provide a forecast of future man-hour output. This, multiplied by an estimate of the hours worked per year on a full-time basis, gives an estimate of output per man-year. The ratio of GNP to value

of output per man-year yields the desired forecast of employment. If we then estimate the civilian labor force (total labor force minus the number of men in the armed services), we secure a forecast of the probable volume of unemployment.

These calculations also provide the basis for making an estimate of potential aggregate supply to compare with our forecast of aggregate demand. If we forecast potential output per man-hour (i.e., labor productivity), the labor force, and hours worked per year per man, we can estimate the potential output that could be produced by the total labor force (minus frictional unemployment). A comparison of this measure of potential aggregate supply with our forecast of aggregate demand will suggest to what extent we expect demand to be pressing on supply in the economy as a whole. Or our estimate of unemployment can be used for the same purpose.

LONG-TERM PROJECTIONS

Thus far, we have been talking about short-run forecasting—say, for a year ahead. It is also fairly common to engage in a particular type of long-term forecasting, say, for 10 or 20 years into the future. The more useful of these longer-run forecasts are frequently called projections rather than forecasts. Actually what they amount to are forecasts of aggregate supply or potential (not actual) output, given a range of assumptions about the future growth of the labor force, changes in productivity, and so on.

Actually we have already seen in Chapter 8 how this type of long-term forecast of potential output would be made. On page 187, for example, we presented a simple equation showing how productive capacity depends on population, the proportion of the population in the labor force, the number of hours worked per year per man, and productivity per man-hour. Table 18 on page 188 provides an example of such a long-term projection, where the variables are expressed in terms of estimated average rates of growth rather than absolute figures.[46]

[46] A considerable number of these long-term projections of potential output have been made by various individuals and groups. For examples other than that given in Table 18, see Staff of the Joint Economic Committee, *Potential Economic Growth of the United States During the Next Decade* (83rd Congress, 2nd Session, 1954), and Gerhard Colm, *The American Economy in 1960,* 1952. A number of the problems involved in making the necessary estimates are discussed in Conference on Research in Income and Wealth, *Long-Range Economic Projection,* Studies in Income and Wealth, vol. 16, 1954.

These long-term projections of full-employment capacity serve a number of useful purposes. For one thing, they give us some idea of how fast aggregate demand needs to grow in order to maintain full employment. And they may suggest measures that need to be taken to insure that aggregate demand grows at a satisfactory rate —measures, for example, to stimulate consumption or particular types of investment. Such projections can also be a helpful guide to long-run planning by both private business and public agencies. If we have a forecast of potential capacity for, say, 20 years ahead, and if we assume that aggregate demand will expand enough to keep unemployment at a reasonably low level, then we have some idea as to what level of GNP to count on in our planning for the future.[47]

FORECASTING DEMAND FOR THE INDUSTRY AND THE FIRM

The businessman has to forecast not only the future course of business activity but also, what is of more direct concern to him, the future demand for the output of his industry and his firm. We shall therefore conclude this chapter with a brief discussion of one method that is used for forecasting the sales of an industry and a single company.[48]

The first step in this procedure is to try to discover some systematic relationship between the industry's sales of a product or group of related products and one or more national (or regional) aggregates—such as GNP, disposable income, industrial production, and so on. The choice of the variable or variables to use depends chiefly on two criteria. What national or regional series are logically related to sales of the product in question? And which variables give the best correlation with past production or sales of this product? The second step consists of making a company forecast on the basis of the predicted behavior of the whole industry. Sometimes these two

[47] Cf. Gerhard Colm, "Economic Projections: Tools of Economic Analysis and Decision Making," *American Economic Review*, vol. 48, May, 1958, pp. 178–187.

[48] More detailed discussions will be found in the following references: Leo Barnes, *Handbook for Business Forecasting*, Part 2; Wilson Wright, *op. cit.*, esp. chap. 15; National Industrial Conference Board, *Forecasting Sales*, Studies in Business Policy No. 25, 1947, and *Forecasting in Industry*, Studies in Business Policy No. 77, 1956; L. J. Paradiso, "How Can Business Analyze Its Markets?" *Survey of Current Business*, March, 1945, pp. 6–13; Clement Winston and Mabel A. Smith, "Analysis of Long-Term Markets," *Survey of Current Business*, November, 1957, pp. 17–22; C. A. Dauten, *Business Fluctuations and Forecasting*, 1954, chaps. 21–22; and the texts by Bassie and Bratt previously cited.

steps are combined, and the firm's sales are related directly to national or regional aggregates.

A firm or industry ordinarily produces several different products, and even the same product may be sold to several different types of buyers who respond to different sorts of cyclical influences. It is frequently necessary, therefore, to break down an industry's total sales into homogeneous product groups and to carry out a separate correlation analysis for each group.

The procedure in making an industry forecast can best be described by a simple illustration, for which we shall take the consumption of paperboard in the United States.[49] The demand for paperboard has been expanding rapidly, not only because the national income has been rising but also because of a succession of new uses that have been found for the product. Thus, the demand facing the industry has been expanding for two reasons: because the total output of the economy has been increasing and because special growth factors have been causing the consumption of paperboard to rise more rapidly than total output or income.

This is illustrated in the two panels of Figure 44. If we look at the scatter diagram in the upper panel, we see that there is a strong positive relation between paperboard consumption (measured vertically) and industrial production (measured horizontally).[50] But part of the growth in paperboard consumption was more or less independent of the rise in industrial production. If we use the technique of multiple correlation and relate paperboard consumption both to industrial production *and* the passage of time, we get the results shown by the regression lines in the two panels of Figure 44.

The line *AB* in the upper panel, in effect, provides an estimate of the net relationship between paperboard consumption and industrial production. But something else was at work to cause the consumption of paperboard to rise faster than can be accounted for by the increase in industrial production. If we take the ratio of actual consumption of paperboard to that which would be estimated from the line *AB* and plot these ratios as a time series, we get the result shown in the lower panel of Figure 44. On the average, the consumption of paperboard tended to increase about 2½ percent per

[49] This illustration is taken from Winston and Smith, *op. cit.*, pp. 20–21.

[50] Note that both scales are logarithmic, so that the regression line measures the percentage change in one variable in response to a percentage change in the other.

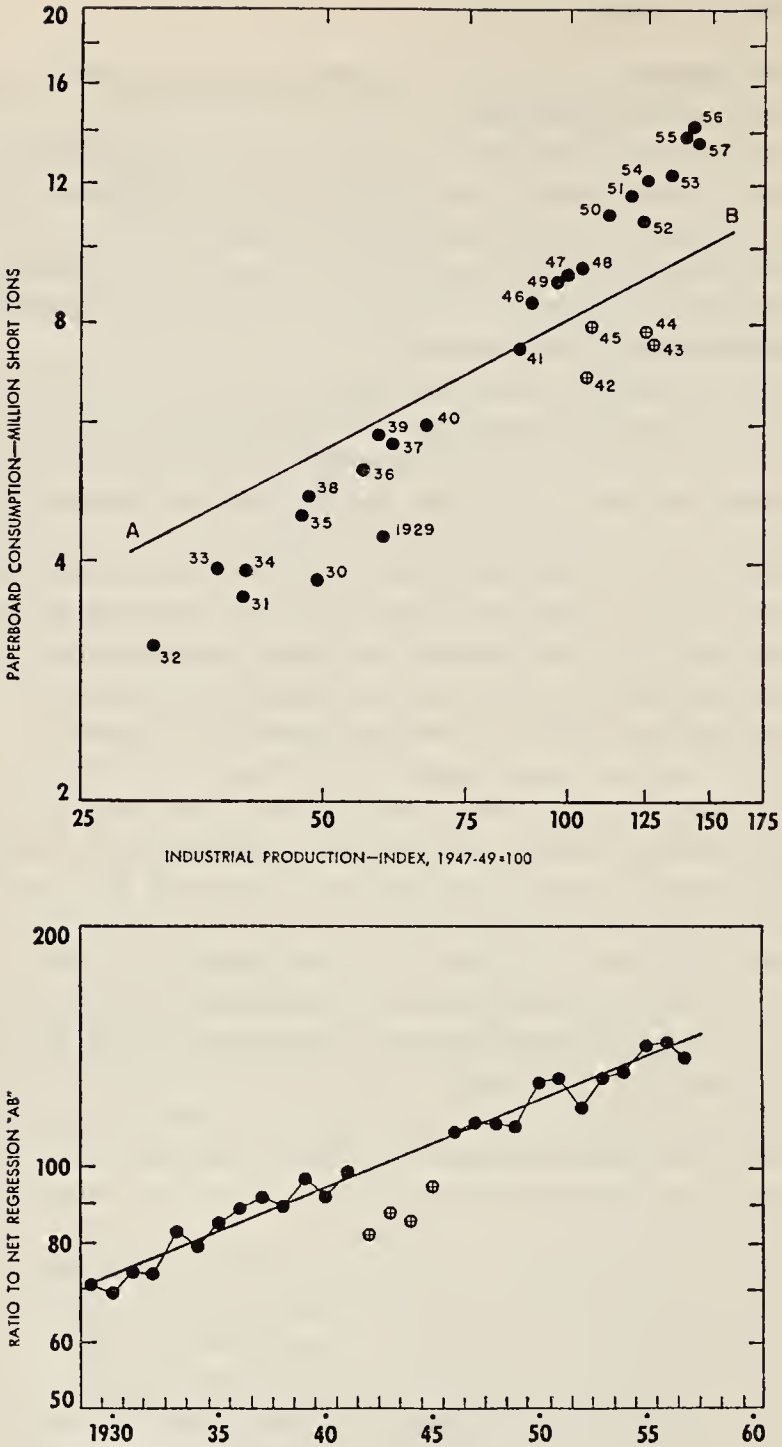


FIGURE 44. Relation of Paperboard Consumption to Industrial Production and Time, 1929–1957.

From *Survey of Current Business*, November, 1957, p. 21.

year over and above the changes that can be explained by the movement of industrial production.

Thus, a forecast of paperboard consumption for a moderate number of years ahead would have to take account of not only the change expected in total industrial production but also the special growth factors causing this industry to expand over and above the expansion in general business. For a *short-run* forecast, anything known about the particular cyclical behavior of the industry would also have to be taken into account. There may be, for example, special factors that cause paperboard consumption to fall below the long-run relationship in times of business recession. There is some evidence in the lower panel of Figure 44 that this may be the case.

If we can assume that the relations shown in Figure 44 will continue in the future, and if we have a forecast of the probable level of industrial production for a future period, we can then forecast the probable amount of paperboard consumption. Obviously, the first "if" in the preceding sentence is an extremely important one. The relations shown in Figure 44 need to be continuously checked as the future unfolds. In particular, it would be unwise to continue to use the net upward trend shown in the lower panel of Figure 44 as a basis for future forecasts without checking, on the basis of independent evidence, that such a trend is continuing to influence demand for the product. As we saw in Chapter 9, for example, the sharp upward trend in new industries eventually becomes subject to retardation. The situation here is parallel to that discussed on page 532, where we saw that it is frequently necessary to make adjustments in the past relation between consumers' expenditures and disposable income in estimating the future relation between these two variables.

The lower panel of Figure 44 suggests that one or more unknown variables, in addition to industrial production and the passage of time, may have had some effect on the demand for paperboard. The deviations around the trend seem to be too systematic to be due entirely to chance. More detailed analysis might indicate what these additional variables were. In many cases, there may be no need to look for such additional variables, and frequently there will be no discernible net time trend in the basic relationship used. In some cases, careful study may reveal a fairly complicated relationship, in which the industry's sales depend not only on such variables as in-

dustrial production or disposable income but also on such additional variables as the cost of living, the prices of competing commodities, consumers' stocks of the commodity in question (an important variable in the case of durable goods), and so on. Probably in most cases, particularly for industries producing consumers' nondurables and basic raw materials, a fairly simple relationship is sufficient. A simple relationship is desirable since it must be tested every time it is used in the light of independent evidence regarding special factors that may be at work. An obvious objection to complicated relations, in addition to the time and expense involved, is that the future values of each independent variable must be estimated before we can use the relationship to make an industry forecast.

Once an industry forecast is made, it can be used as a basis for estimating the future sales of a particular company. This can most readily be done in terms of the ratio of the company's sales to sales for the whole industry. The behavior of this ratio during past years may reveal a trend, indicating that the company has been growing more or less rapidly than its competitors. Extrapolation of this trend, together with any adjustments necessary to allow for special selling efforts or other unusual factors likely to affect the company's position in the industry, yields a percentage figure that can be applied to estimated industry sales to give a forecast of the sales that the company may expect to make.

Forecasts of company and industry sales can be made by other methods besides the general procedure described here. Many companies still prefer to rely on field surveys by their sales departments and on similar techniques. The best results are frequently obtained when two or more methods are used and the results reconciled.[51]

The general method described in this section has come into increasing favor since World War II. It obviously has much to commend it, and its use is likely to grow. If it is to yield satisfactory results, however, a warning must be kept in mind that we have already repeated several times in this chapter. No historical relationship, no matter how closely it has held in the past, should be extended into the future without careful checking on the basis of all the evi-

[51] The various methods that may be used are summarized in National Industrial Conference Board, *Forecasting in Industry*. See also Controllership Foundation, *Business Forecasting*, for a survey of the methods used in making sales forecasts by a sample of American corporations.

dence available. This rule holds for company and industry forecasts as well as for general business forecasts. It is also worth remembering that, under the procedure described here, the company forecast is tied to a forecast of some broad national variable. In an important sense, therefore, the accuracy of the company forecast depends on how well we can forecast the basic economic variables on which the industry's and the company's sales depend.

CHAPTER 18

THE GOAL OF ECONOMIC STABILITY

THE GOAL of economic stability can be resolved into the twin objectives of sustained full employment and the avoidance of large changes in the general level of prices. On this, virtually all economists are agreed. We want to keep the economy continuously at the full-employment level, and at the same time we want to avoid marked swings in prices in either direction. As it is sometimes put, we want to avoid both deflation and inflation. The question now before us, then, is this: Given the destabilizing forces that are continuously at work, how can we overcome or offset these forces so that full employment may be maintained without inflation?

Before World War II, the problem of economic stability was thought of as a problem of preventing depressions. The Great Depression, as well as earlier experience with cyclical downswings, focused our attention on the need to prevent mass unemployment. Since World War II, while not forgetting the danger of unemployment, we have also become concerned with the threat of continuing inflation. A policy of economic stabilization must aim at both objectives, full employment and price stability.

In the last decade or so, there has been much discussion of a third objective—maintaining a satisfactorily rapid rate of growth in total output. Thus, in 1959, the Joint Economic Committee of Congress undertook a large-scale study of the best ways of achieving simultaneously "a high and stable level of employment, a high rate of growth in our national output and productive capacity, and a high degree of stability in the general level of prices." To a very considerable extent, this goal of rapid growth overlaps that of full employment. If the labor force and labor productivity continue to

increase, then maintaining full employment means also growth in total output. Much of the recent debate about how to speed up the rate of growth has centered around two issues: how to accelerate the increase in labor productivity and how to keep unemployment at a minimum. Later in this chapter we shall consider further the goal of rapid growth. But now let us look at the other two objectives —full employment and price stability.

THE MEANING OF "FULL EMPLOYMENT" AND "PRICE STABILITY"

Neither of these objectives is easy to define. Full employment does not mean that unemployment is reduced to zero. Even at the peak of the war effort during World War II, the Bureau of the Census reported that some 500,000 to 1,000,000 persons were unemployed. Yet virtually everyone would agree that the American economy was then operating at forced draft—that output and employment were higher than the "full-employment" level. Some unemployment always exists—because people are in the process of moving from one job to another, because of sickness or seasonal layoffs, and so on. This is what we mean by "frictional unemployment."

What, then, does "full employment" mean, and how much unemployment is consistent with full employment? For our purpose, a fairly loose definition will do, and we shall adopt that offered by a committee of the American Economic Association: "Full employment means that qualified people who seek jobs at prevailing wage rates can find them in productive activities without considerable delay. It means full-time jobs for people who want to work full time. It does not mean that people like housewives and students are under pressure to take jobs when they do not want jobs, or that workers are under pressure to put in undesired overtime. It does not mean that unemployment is ever zero."[1]

Another definition may be cited to illustrate the need for careful formulation of the full-employment objective. A committee of experts for the United Nations defined full employment "as a situation in which employment cannot be increased by an increase in ef-

[1] Committee of the American Economic Association, "The Problem of Economic Instability," *American Economic Review*, vol. 40, September, 1950, p. 506. For a useful survey of different definitions of full employment and their policy implications, see the paper by Albert Rees in Universities–National Bureau Committee, *The Measurement and Behavior of Unemployment*, 1957, pp. 13–60.

fective demand."[2] This is clearly going too far. A country may be suffering the beginning pangs of inflation; unemployment may be less than the usual frictional amount; and yet it may be possible to induce a further increase in the labor force and in total employment by inflating aggregate demand still further. Obviously, such a policy, while it might increase employment somewhat, would also lead to an inflationary rise in prices.[3]

How much unemployment is consistent with our definition of full employment? The answer is not likely to be the same for all countries. It depends primarily on the degree of labor mobility and on the amplitude of the seasonal swings in employment.[4] So far as the United States is concerned, we shall probably not be far off if we say that full employment requires that not more than 3 to 4 percent of the labor force be unemployed.[5] With a civilian labor force in 1960 of approximately 70 millions, full employment would require that the unemployment figure not exceed, say, 2.8 million—and preferably be somewhat less than this. An unemployment figure substantially less than this—say, one million—would imply widespread labor shortages and a situation in which aggregate demand exceeded supply at current prices. Unemployment amounting to much more than this—say, four or five million—would indicate that aggregate demand was too low. At the end of 1960, unemployment in the

[2] United Nations, *National and International Measures for Full Employment*, 1949, p. 13.

[3] See also Sir William Beveridge, *Full Employment in a Free Society*, 1945, p. 18. Beveridge's definition contains the widely quoted statement that full employment means "having always more vacant jobs than unemployed men." This has been criticized on the ground that an excess of jobs is unnecessary and would lead to inflation.

[4] See *The Extent and Nature of Frictional Unemployment*, Study Paper No. 6 prepared for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959). In addition to short-term "frictional" unemployment, an economy may suffer from some degree of longer-run "structural" unemployment—for example, in distressed areas or industries or among minority groups or older workers. This type of unemployment has been a significant part of the total in the United States in recent years. Cf. Joint Economic Committee, *Staff Report on Study of Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959), pp. 170–187.

[5] This makes no allowance for involuntary part-time unemployment. Making allowance for this and certain other factors, T. K. Hitch estimates that at "full employment" the American economy utilizes about 96 percent of the total working time available from the labor force. "Meaning and Measurement of 'Full' or 'Maximum' Employment," *Review of Economics and Statistics*, vol. 33, February, 1951, pp. 1–11.

United States totaled about 4.9 million, out of a civilian labor force of 71 million. This implied a seasonally adjusted unemployment rate of about 6.8 percent. The American economy was then substantially short of the full-employment level.

The goal of price stability is somewhat easier to define. Again we may quote the committee of the American Economic Association: "Price-level stability means the absence of any marked trend or sharp, short-term movements in the general level of prices."[6] We want to avoid any pronounced secular trend in the price level, either upward or downward. We also want to avoid wide short-run fluctuations in either direction. At the same time, however, we do not want to interfere with the flexibility of *individual* prices. The relative prices of particular commodities should be free to shift widely in response to changes in consumers' tastes and particularly to changes in cost conditions.

Price-level stability should be defined in terms of both wholesale and retail prices. Price-level stability means that neither the wholesale nor the retail price level (as measured by appropriate indices) would be permitted to vary by more than predetermined amounts. Of course, we should not expect either sort of index to remain absolutely fixed. The best way to implement the goal of price-level stability is to establish a range within which the selected price index is expected to move. This range would be wider for wholesale than for retail prices, since the former are ordinarily more sensitive than the latter.

The goal of economic stability has a short-run and a long-run aspect. This is true of both objectives, full employment and price stability. The goal of short-run economic stability is essentially the goal of eliminating (or at least moderating) cyclical fluctuations in business activity. Short-run stability, however, is not enough. We also want to control the long-run level around which the permissible fluctuations are to occur. We do not want stability of output or employment at a level well below that of full employment. Nor do we want a "creeping" sort of inflation that may double the price level every 25 years or so. In seeking cyclical stability, then, we want also to avoid both *secular stagnation* and *secular inflation.*

No particular problem arises here with respect to the full-employment objective. This goal obviously means short-run stability *and*

[6] *Op. cit.,* p. 506.

full employment. Secularly, employment would rise with the growth of the labor force. Output would rise with the increase in employment and in labor productivity. Appropriate allowance would have to be made for any reduction in the number of hours worked per week; some output might be sacrificed for more leisure.

Although the desirable long-run objective is clear and unambiguous in the case of employment, this is not so with respect to the price level. We might seek short-run stability of prices on a horizontal secular trend or on one that was slowly rising or slowly falling. Arguments have been advanced for all three sorts of trend. Limitations of space do not permit a summary and evaluation of these arguments here, and we shall merely state our own position. In our opinion, the long-run objective should be a constant price level, although this has become a difficult goal to achieve since World War II.[7] This means that MV must rise secularly as total output grows. It means also that, with increasing labor productivity, money wage rates should rise with the secular increase in output per man-hour. This objective seems preferable to either (1) a falling price level with constant money wage rates or (2) a slowly rising price level with money wages rising still more rapidly.

MAINTAINING A SATISFACTORY GROWTH RATE

We saw in Chapter 8 that, over the past century, total output in the United States has grown at an average rate of about 3.5 percent per year. The rate was higher than this in the nineteenth century; it has averaged somewhat less since World War I. There have been numerous complaints in the last few years that the American economy has been growing too slowly, and it has been suggested that we should try to achieve a much higher rate of growth—perhaps as high as 5 percent per year.[8]

There is no single answer as to what is an appropriate or satisfactory rate of growth. As we saw earlier, if we maintain full employment, growth in total output will automatically occur as long as the labor force continues to expand and output per worker keeps on in-

[7] It is possible to have different trends in wholesale and in retail prices. Gains in productivity tend to reduce the former more than the latter. Thus, if we keep the trend in retail prices horizontal, it may be possible to permit a slight downward drift in the level of wholesale prices.

[8] See, for example, the testimony of Leon Keyserling, in *Employment, Growth, and Price Levels,* Hearings before the Joint Economic Committee, Part 1, p. 106.

creasing. Assuming aggregate demand is maintained at a level high enough to insure full employment, growth of output depends on the variables listed on page 187—population increase, proportion of the population in the labor force, the trend in hours worked per year, and changes in output per man-hour. All of these variables can change and have changed in the past. Before settling on some desirable rate of growth, we need to know something about the likely trends in these variables. If the rate of population growth declines, if children stay in school longer and fewer women enter the labor market, and if people choose more leisure and work fewer hours per year, then we shall have to be satisfied with a lower rate of growth than if these variables were more favorable to rapid expansion in output.

In a free society, the government does not try to influence the variables just mentioned (population, labor force participation, and number of hours worked) in order to achieve some particular rate of growth. But the government can affect the rate of growth by influencing the level of aggregate demand, in order to insure that labor and other resources are fully utilized, and it can in a variety of ways stimulate the growth in labor productivity.

Given prospective trends in the relevant variables, it has been suggested that a rate of growth in the neighborhood of 4 percent per year would be an appropriate goal for the United States in the 1960's.[9] This is significantly higher than the rate of growth after the Korean War, and higher also than the average rate of growth during the past century.

RECONCILING THE STABILITY AND GROWTH OBJECTIVES

It is widely believed that there has been an inflationary bias in the American economy since World War II. The reasons for this belief include: the strength of organized labor and the persistent pressure on trade union leaders to insist on wage increases larger than the gains in labor productivity, the greater willingness of employers to meet wage demands (based in part on the assumption that higher labor costs can be passed on in higher prices), the increased rigidity of prices in recessions, and the assumption—by workers, employers, and consumers alike—that the government is committed to

[9] See *Employment, Growth, and Price Levels,* Hearings, Part 9A, p. 2959.

maintaining a high level of employment and purchasing power.[10]

If this inflationary bias does exist—and the evidence does point in that direction—then some secular rise in prices cannot be avoided if we adhere strictly to the goals of full employment and rapid growth. There is fairly general agreement that a high level of employment is a more important objective than long-run constancy in the price level. I have suggested elsewhere that a reasonable compromise might be the following combination of objectives: an average unemployment rate over a series of years of 4 percent, a rate of growth in output of 4 percent, and a gradual increase in the Consumer Price Index of not more than 1 percent per year.[11] Even this compromise will not be easy to achieve.

THE IMPORTANCE OF OTHER GOALS

Important as they are, economic stability and rapid growth are not the only goals on which we must keep our eyes fixed. We assume that most people in democratic countries would emphasize also the objectives of peace, national security, and the maintenance of basic democratic institutions. It is obviously undesirable to stabilize employment at home in a way that will breed ill will abroad and foster unrest in other countries. It is also true, particularly in these troubled times, that the national defense places a heavy burden on the economy and that stabilization measures must be framed with this problem in mind. While striving for peace and security, as well as economic stability, democratic societies also seek to preserve all those "essential liberties which are more precious than full employment itself."[12] This means the maintenance of the political institutions that we associate with the word *democracy*. It means also a large degree of personal freedom in economic matters, particularly freedom within wide limits as to how a person spends his income and freedom in the choice of an occupation and a job. In the United States, it means also freedom to start and manage a business,

[10] This subject has generated a tremendous literature since the war. A useful starting point is C. L. Schultze, *Recent Inflation in the United States*, Study Paper No. 1 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels*. See also the other volumes in the Committee's study; also The American Assembly, *Wages, Prices, Profits, and Productivity*, 1959.

[11] See *Employment, Growth and Price Levels*, Hearings, Part 9A, pp. 2959–2960.

[12] Beveridge, *op. cit.*, p. 36.

subject to a considerable and growing amount of government intervention in the public interest.

Our list of social objectives includes also a steady rise in the standard of living. This is one reason why growth in output is one of our goals. Conditions should also be such that the composition of total output can change readily in response to the introduction of new products and to changes in consumers' tastes. Stability does not mean inflexibility.

Finally, we must set down the goal of continuing social and economic reform. Undoubtedly, we shall continue to argue among ourselves as to what reforms are necessary, what the effects of these changes will be, and what costs have to be incurred. Despite such arguments, we shall undoubtedly continue to move in the direction of what is sometimes called "the welfare state," toward a broadened social security program and toward better housing, health, and education, especially for low-income groups. We shall also want to take further steps to conserve and develop our natural resources, to improve our cities, our farms, and our highway system, and so on. Here we may run into an important conflict of objectives, particularly if the demands for national defense create serious inflationary pressures. The objective of economic stability may sometimes require that we go more slowly with some of these social programs than we should like. Or we may be able to achieve both stability and a rapid expansion of social expenditures only by resort to the kind of detailed economic controls that we should like to avoid. Where such conflicts between objectives occur, the economist can do no more than present the consequences of alternative lines of action. The actual choice that is made must depend on the value judgments of those who are responsible for policy—government officials, legislators, and ultimately the voting public.

OBSTACLES TO ACHIEVING STABILITY

It is sometimes said that we cannot eliminate cyclical fluctuations because we do not yet know enough about their causes. We can immediately dismiss this fear as groundless. Knowledge of the causes of changes in the weather is not needed to develop heating and air-conditioning systems that permit us to offset these changes. Even if we knew no more than that employment depends on aggregate de-

mand and that the components of such demand fluctuate, we would have the basis for constructing a stabilization program of some effectiveness. Actually, our knowledge goes much further than this. In particular, we know a good deal about the various ways in which government can influence business activity. We know, for example, that a restrictive monetary policy, increased taxes, and reduced government spending will tend to hold back an inflationary boom, even if we have not located the precise causes that started off the boom. Though the area of disagreement among economists on many matters is notorious, they are fairly well agreed on the kinds of measures that are most likely to foster economic stability.

Another and more serious difficulty is our inability to forecast accurately the future course of business, even a few months ahead. Coupled with this forecasting difficulty is our inability to diagnose fully elements of weakness in the current economic situation. We must take this into account in the kinds of policies we recommend. We cannot rely entirely on "preventive medicine," to be administered on the basis of continuous diagnosis and prediction. Since we will sometimes be late to recognize the need for action, we want measures that can be put in effect promptly and that can be withdrawn or reversed quickly. So far as possible, we should like to use measures that will operate more or less automatically in accordance with certain criteria, without the need for elaborate diagnosis and prediction before action is taken. Almost certainly, the limitations on our ability to diagnose and predict mean that we cannot hope to eliminate all mild recessions and booms, but we certainly know enough to diagnose serious trouble in time to do something about it and to prevent minor fluctuations from developing into something much more serious.

Our ability to act sufficiently promptly to prevent marked economic instability is limited more by administrative and political difficulties than by economic ignorance.[13] The administrative and legislative machine on which we must rely moves slowly and ponderously. The executive branch of the government must see the

[13] For a description of the administrative and political difficulties that interfere with an effective stabilization program in the United States, see Roy Blough, "Political and Administrative Requisites for Achieving Economic Stability," *American Economic Review*, vol. 40, May, 1950, suppt., esp. pp. 165–170, and S. K. Bailey, "Political Elements in Full Employment Policy," *American Economic Review*, vol. 45, May, 1955, pp. 341–350.

need for action, must then decide that it will act, then plan a line of action, and then perhaps go to the legislative branch for authority and funds.

Obviously we need to do something to streamline this elaborate and cumbersome procedure. The stabilization objective should be clearly stated, spelled out in some detail, and given the force of law. The authority of the executive branch to act promptly should be clearly established. So far as possible, there should be set up clear-cut criteria to guide government officials as to when to act. All this is not likely to be enough, if we leave the nature and timing of stabilization measures entirely to the discretion of the government agencies concerned. There are strong grounds for preferring measures that will operate *automatically* in the desired counter-cyclical way. "Effective stabilization policies require . . . that positive countermeasures, the nature of which is announced beforehand, be designed which automatically come into operation in certain clearly defined eventualities."[14] We have made only a start in this direction in the United States. Although automatic measures are to be preferred, a comprehensive stabilization program is likely to include discretionary as well as automatic policies.

Some of the most serious obstacles to a successful stabilization program are political rather than administrative. In this country, political difficulties are of several sorts. There is the traditional lack of confidence between Congress and the President. The failure of Congress to accept many of the President's recommendations under the Employment Act of 1946 is a good example. Perhaps of greater importance is the fact that the stabilization objective frequently conflicts with other objectives that have strong political support. Nearly every policy of government has some bearing on economic stability. Stabilization policies may involve tax changes, more or less government spending, control of credit, etc., all of which impinge on other government policies. Support for these other policies may seriously interfere with carrying through effective stabilization measures. Related is the problem of pressure groups—private pressure groups and also the vested interests of established government agencies and of legislators in particular programs that may be hurt by some stabilization measures. Farm groups do not want a policy that will prevent farm prices from rising in an inflationary boom; organ-

[14] United Nations, *op. cit.*, pp. 39–40.

ized labor may oppose inflation controls that threaten to prevent wage increases; government agencies and private groups concerned with welfare programs are not likely to acquiesce in a curtailment of government expenditures; and so on. These conflicts will be reflected in Congress itself, where we find many blocs seeking to support certain policies and to oppose others. "This fragmentation of Congress prevents adequate emphasis on economic stability as an objective of legislation; other considerations take precedence."[15]

These are serious and perhaps insuperable obstacles to the achievement of an effective stabilization program. Some of the things that need to be done are obvious, particularly the development of greater economic sophistication among both legislators and the public at large and education of the public to recognize more clearly the difference between group interests and the public welfare. But, to some extent, these difficulties will always be with us. They seem to be part and parcel of the democratic process. They help to explain why we have not progressed further than we have in the direction of economic stability. We must keep them in mind in evaluating the types of stabilization measures which are available.

THE TOOLS AVAILABLE

We have seen that income and employment fluctuate with changes in the level of aggregate demand. We know also that aggregate demand is merely the sum of consumers' expenditures, private domestic and foreign investment, and government spending. Of these, consumers' spending is the largest, but also the most passive; private investment is the most unstable; and government spending is the one that can be most readily manipulated through government action. If we are to achieve economic stability, then, the general lines along which we must proceed are clear. We must either stabilize the various components of aggregate demand or, if this cannot be done, offset the fluctuations in some components by counter movements in the others. The latter alternative immediately suggests the possibility of using changes in government expenditures to offset fluctuations in consumers' and business spending.

Since spending takes place through the use of money, changes in the supply of money will be one of the factors affecting the level

[15] Blough, *op. cit.*, p. 169.

of spending. This suggests that monetary policy is one of the methods available to control the level of aggregate demand. In particular, it is an indirect method of bringing about desired changes in the level of private investment.

These brief preliminary comments are enough to suggest the main types of measures available on which we must rely to make the economy more stable. In the next two chapters we shall discuss the effectiveness and feasibility of these measures primarily under the following headings:

1. Control of consumption and private investment.
2. Monetary policy.
3. Fiscal policy—i.e., using taxes and government expenditures to offset or influence variations in private spending.

To these three we should add a fourth, which has not been suggested by the preceding discussion. This is wage-price policy. In a world of strong trade unions, oligopolistic markets, and organized farm blocs, prices and wages do not play a purely passive role. There are forces making for changes in prices and wages that are to some extent independent of the other influences leading to movements in aggregate demand. When prices and wages change, consumption and investment are affected. We have become familiar with these semiautonomous price-making forces during the inflationary years following World War II. Rounds of wage increases were in part the result of a prior rise in prices and in turn led to a further rise in prices.

The importance of changes in wages and prices is particularly apparent in times of inflation. Indeed, as we have seen, there seem to have been forces at work since World War II that have imparted an inflationary trend to wages and prices—not only in the United States but also in other countries. We shall look at this problem further later on. But even in times of declining demand, we cannot afford to be indifferent to the behavior of wages and prices. For example, do inflexible wages and prices accentuate or moderate a downswing in business? Here we get into a range of issues on which there are sharp differences of opinion. Further discussion of this problem belongs in a later chapter, but we have already said enough to suggest that a full-fledged stabilization program may re-

quire some sort of supervision over the price and wage policies of business firms and trade unions.

We shall conclude this chapter with a few words about the tools available to stimulate the rate of growth in total output. To do this, we must once again make our distinction between aggregate demand and aggregate supply. So far as growth of output depends on expansion of aggregate demand, we have already mentioned the tools that are available—measures to stimulate directly consumption or private investment and monetary and fiscal policy. There are also steps that can be taken to stimulate aggregate supply—that is, to make the productive capacity of the economy grow more rapidly. These include a wide array of possible measures—for example, improvement of the health and education of the population, steps to increase labor mobility, encouragement of scientific research and technological development, better roads and urban transportation, and so on.

Since even in a book this large some pertinent topics must be omitted, we shall not have anything more to say about the steps that might be taken to increase the productivity of the labor force and thereby aggregate supply.[16] In the next two chapters we shall concentrate on what can be done to stabilize and expand aggregate demand. In addition, in Chapter 21, we shall consider briefly how wages and prices can be influenced other than through the level of aggregate demand.

[16] See, however, the references in footnote 10 on page 550; also the discussion of the forces affecting the growth of output in Chapter 8.

CHAPTER 19

METHODS OF ACHIEVING GREATER STABILITY

AT THE end of the last chapter we pointed out that we can try to stabilize the components of private demand (that is, consumption and investment) or we can try to offset fluctuations in private spending by appropriate changes in government spending. Let us first see what can be done to stabilize consumers' expenditures, which ordinarily account for about two thirds of the entire gross national product in the United States.

STABILIZING CONSUMERS' EXPENDITURES

The amount of consumers' expenditures in any period can be thought of as the result of two sets of influences: those determining the level of disposable income after personal taxes and those determining the ratio of consumers' expenditures to disposable income (the propensity to consume). The most important cause of cyclical changes in consumption is variation in the level of disposable income. This is what we mean when we say that consumption is ordinarily passive, rising or falling as income changes. If we want to stimulate or reduce consumption, we must do it primarily through changing the amount of income available to consumers, though something can also be done to influence the willingness or ability of consumers to spend out of a given income.

The importance of stabilizing disposable income is well illustrated by the behavior of the American economy since World War II. In each of the postwar recessions, as we saw in Chapters 15 and 16, the so-called automatic stabilizers operated strongly to hold

up disposable income and therefore consumption. This was one of the important reasons for the mildness of these business contractions.

MEASURES TO STABILIZE DISPOSABLE INCOME

There are several ways of moderating the cyclical instability in consumers' incomes. Particularly important in this connection are social security programs, especially unemployment insurance, and income taxes, especially when put on a "pay-as-you-go" basis. Social security contributions by employers and employees rise and fall with the amount of employment; social security payments by the government, particularly unemployment insurance payments, move inversely with the business cycle.[1] Thus, during a business downswing and declining employment, the government increases payments to the unemployed at the same time that its social security collections are declining because fewer people have full-time jobs. A widespread and liberal unemployment insurance program can be an important factor tending to hold up personal incomes (and therefore consumption) during periods of declining employment.[2]

This is true of any program that automatically increases government transfer payments to the private sector of the economy during business declines and automatically reduces them during business upswings. Not only do unemployment insurance and relief payments operate in this way. A similar stabilizing influence is exerted by the American farm-support program. When agricultural prices are falling, the government makes large payments to farmers; these payments are reduced when farm prices rise. As a result,

[1] Old-age retirement and some other types of social security payments do not vary significantly with the business cycle, but the very stability of these types of expenditures helps to moderate the severity of depressions. They represent a type of income that does not fall when business activity declines. For a study of the role of social security programs in business cycles, see Ida C. Merriam, "Social Security Programs and Economic Stability," in Universities–National Bureau Committee for Economic Research, *Policies to Combat Depression,* 1956, pp. 205–235.

[2] Although provisions differ among the various states, unemployment insurance in the United States cannot be considered overly liberal in terms of the maximum payments available or the maximum period of unemployment for which payments are made. Hence the stabilizing effects of the program, particularly in the event of a serious depression, are limited. In the 1958 recession, emergency federal legislation was necessary to lengthen temporarily the period for which unemployed workers could get benefits. Cf. *Economic Report of the President,* January, 1959, p. 40.

consumption by the agricultural population is more stable than it would be in the absence of such a program.

A similar automatic stabilizing influence is achieved through a "pay-as-you-go" income tax. This is particularly true when tax rates are highly progressive and personal exemptions are well above zero. In this case, income tax payments fall relatively faster than total personal incomes in depression and rise relatively faster during business booms. This effect is felt immediately if a large part of the tax liability is met by deductions from current income, which is what happens under a "pay-as-you-go" system. As a result, *disposable* incomes (incomes after taxes) vary less than do total incomes before taxes; and this, of course, makes consumption more stable than it would otherwise be.

Corporation income taxes also perform this stabilizing role. Indeed, their stabilizing influence has been much more important in the mild recessions since World War II than has been that of the personal income tax. Corporate profits before tax fluctuate very widely over the cycle. At present tax rates, roughly half of any decline in corporate net income before tax is offset by a decline in corporate income-tax liabilities. This is one reason why dividends have been relatively well maintained in recent recessions. Thus, during the 1957–1958 recession, corporate profits declined by about 11 billion dollars, and corporate profits tax accruals fell by 5.5 billions. In the same period, personal income taxes fell only 1.4 billions, because personal incomes before taxes were well maintained.[3]

It is to be emphasized that these stabilizing results, whether we are considering the corporation or the personal income tax, are achieved without any change in tax *rates*. A constant rate structure causes the total of tax payments to vary relatively more than total personal income. The effect we have been describing is entirely automatic. In addition to relying on this automatic effect, the government might go further and raise tax *rates* (and reduce exemptions) during booms and follow the opposite policy during depressions. This would moderate still more the cyclical fluctuations in disposable income. We shall discuss this possibility further in the next chapter.

Where the results are purely automatic—as they are with unem-

[3] See the discussion of the 1957–1958 recession in Chapter 16.

ployment insurance, our present farm-support program, and a given structure of tax rates—we can say that we have introduced automatic stabilizers into the economy. These tend to reduce cyclical fluctuations in consumption by reducing the fluctuations in the disposable income out of which consumers' expenditures are made.[4] Since these stabilizers introduce an automatic cyclical-response pattern into the government budget—causing expenditures to rise relative to revenues during depressions and the reverse during business upswings—we can say that a certain amount of flexibility has been built into the government budget. The phrase *built-in flexibility* has become popular in the literature on fiscal policy in recent years, and we shall return to it in our discussion of fiscal policy in the next chapter.

INFLUENCING THE PROPENSITY TO CONSUME

So far we have been concerned only with stabilizing consumption indirectly through reducing fluctuations in the incomes that consumers have available for spending.[5] We can also influence con-

[4] A recent econometric study suggests that, under the conditions prevailing in 1958, the government automatic stabilizers would have offset about half the decline in GNP in a short and mild recession. This allows for the effect of changes in corporation and personal tax receipts (offsetting about 30 percent of the decline in GNP), transfer payments (offset of 18 percent), and social security contributions (offset of nearly 5 percent). The decline in business saving would have provided a further substantial offset. The amount of each offset would differ depending on the length of the recession. Cf. James Duesenberry, Otto Eckstein, and Gary Fromm, "A Simulation of the United States Economy in Recession," *Econometrica,* vol. 28, October, 1960. For other studies of the quantitative effect of the automatic stabilizers, see D. W. Lusher, "The Stabilizing Effectiveness of Budget Flexibility," in *Policies to Combat Depression, op. cit.,* pp. 77–89; and M. O. Clement, "The Quantitative Effect of Automatic Stabilizers," *Review of Economics and Statistics,* vol. 42, February, 1960, pp. 56–61.

[5] The last step in the stabilized-income approach to controlling the level of consumption has been proposed by J. H. G. Pierson, who advocates that a given level of consumption be "underwritten" by the government. The level of consumption to be guaranteed would be that level which, with the average amount of private investment and government expenditures expected in the future, would insure a level of aggregate demand sufficient to insure full employment. Disposable incomes and thus consumption would be stabilized at the desired level by such means as changes in tax rates, consumers' subsidies, and so on. See his *Full Employment and Free Enterprise,* 1947, and "The Underwriting Approach to Full Employment: A Further Explanation," *Review of Economics and Statistics,* vol. 31, August, 1949, pp. 182–192. A more recent proposal for stabilizing personal incomes is that made by P. J. Strayer in the *American Economic Review,* vol. 40, December, 1950, pp. 827–844.

sumption more directly by changing the fraction of their disposable incomes that consumers spend. Or, to put it another way, we can try to influence the manner in which consumers divide their disposable income between spending and saving.

One possible way of doing this is through the redistribution of incomes. Low-income families consume a larger fraction of their incomes than do high-income families. Hence, measures that tend to equalize personal incomes will tend to increase somewhat the ratio of consumption to income in the economy at large. However, this is not a very convenient way of moderating short-run cyclical fluctuations in consumption. The most direct way of changing the distribution of incomes is through making the tax structure more or less progressive. But most economists would agree that we do not want to introduce radical changes in the tax *structure* every year or two, making it more progressive when business declines and undoing this action when business revives. Not only would the political and social repercussions be an obstacle to doing this frequently, but also constant tampering with the tax structure would increase the uncertainty within which business must operate and almost certainly would tend to depress the level of private investment in the long run.[6] There is also some doubt that moderate changes in the tax structure, which did not at the same time change the total yield of taxes, would have a significant effect on the level of consumption.[7]

Changes in the progressiveness of the tax structure are not a good device to use to combat cyclical fluctuations, but they may serve a useful purpose in dealing with long-run trends. This depends largely on how progressive the tax structure already is. If the economy is suffering from secular stagnation, tax rates might

[6] While cyclical and large-scale changes in the relations between different tax rates are not desirable, the same objections do not apply with equal force to changes in the general level of rates. See Chapter 20.

[7] It has been suggested that, given the tax rates in the United States after World War II, "feasible changes in tax structure, as distinct from changes in tax yield, are not likely to go far in dealing with a serious deficiency of consumption in a period of deflation or a serious excess of consumption in a period of inflation." R. A. Musgrave and Mary S. Painter, "The Impact of Alternative Tax Structures on Personal Consumption and Saving," *Quarterly Journal of Economics*, vol. 62, August, 1948, p. 495. For a more recent study, see M. Bronfenbrenner, T. Yamane, and C. H. Lee, "A Study in Redistribution and Consumption," *Review of Economics and Statistics*, vol. 37, May, 1955, pp. 149–159.

be made more progressive and left that way. This would permanently reduce the amount of savings that would be made at a given level of income. Unfortunately, it might also, depending on the circumstances, reduce further the level of private investment. In a country already having a high and very progressive set of tax rates, the opportunities here are likely to be limited.

There are other ways of affecting the distribution of income and thus the economy's propensity to consume. One such way is to alter the division of the national income between wages and profits. An increase in wages at the expense of profits will tend to increase consumption. Unfortunately, such a shift may also reduce investment incentives and thus the amount of private investment. We are probably safe in concluding, at least so far as cyclical changes are concerned, that measures aimed directly at the redistribution of incomes are not an effective way of reducing fluctuations in consumers' demand.

A more direct way of influencing the consumption-income ratio is to tax consumption and saving at different rates. One suggestion during World War II, for example, was that there be a special tax levied on consumers' expenditures, in addition to the regular income tax.[8] Another more recent proposal is that some part of a person's saving be treated as an exemption in computing his income tax, thus increasing the incentive to save.[9] Akin to these suggestions for curbing consumption during inflationary periods is the plan for compulsory saving, whereby, in addition to his income tax, the taxpayer would be required to turn over to the government a percentage of his income in exchange for government bonds to be redeemable at some time in the future.[10]

All of these proposals have the problem of inflation in mind and were not designed to deal with depressions as well as booms.[11] In

[8] See, for example, the articles by Milton Friedman and K. E. Poole, in the *American Economic Review,* vol. 33, March, 1943, pp. 50–73.

[9] S. H. Slichter, "The Problem of Inflation," *Review of Economics and Statistics,* vol. 30, February, 1948, p. 5.

[10] For references to and evaluation of these proposals, see James Tobin, "Taxes, Saving, and Inflation," *American Economic Review,* vol. 39, December, 1949, pp. 1223–1232.

[11] During the depressed 1930's the proposal was frequently made that consumption be stimulated by a special tax on current saving or on idle cash balances. One of the better-thought-out proposals of this sort was Arthur Dahlberg's plan for a tax on idle money. (For a summary of his views, see A. G. Hart, *Money,*

theory, of course, these plans could be reversed during deflationary periods; the special incentives to save could be withdrawn in order to stimulate consumption. However, there are good reasons for rejecting any plan that calls for frequent and important changes in the tax structure. In addition, most of these plans involve serious administrative difficulties, and there is also doubt that some of them would have the desired effectiveness. We conclude that plans for special taxes on consumption or for compulsory saving should be reserved for emergency situations when the threat of inflation is very great and when it is not politically feasible to raise income taxes sufficiently. In such cases, the government might well add a compulsory savings program to the highest level of taxation that proves to be politically possible.

One type of consumers' spending is particularly unstable—that part which goes into the purchase of automobiles, electric appliances, furniture, and other durable goods. A substantial part of these goods is bought on credit, a fact that accentuates the instability in this type of spending. Regulation of the terms on which such credit may be obtained is one way of exercising control over consumers' expenditures of this sort. Such "selective credit controls" were used in the United States during World War II (until 1947), briefly during 1948–1949, and again after the Korean crisis. The Board of Governors of the Federal Reserve System specified minimum down payments and maximum repayment periods for consumers' loans used to buy various kinds of durable goods. This is an effective means of reducing a particularly volatile kind of spending during inflationary periods, and it has also been used in other countries. Relaxation of such controls over consumers' credit can provide a stimulus during business downswings, although the stimulus will not be very effective if consumers do not want to buy even on easy credit terms or if lenders are unwilling to extend credit on these terms.[12]

Debt, and Economic Activity, 2nd ed., 1953, pp. 434–437.) Some of these proposals are reviewed by A. H. Hansen in *Economic Policy and Full Employment,* 1947, chap. 19. The plans were based on the assumption that aggregate demand would be more or less permanently deficient, whereas the proposals discussed in the text are concerned with the reverse problem—what to do about inflation.

[12] For an exhaustive analysis of the pros and cons of regulating consumer credit, see *Consumer Installment Credit,* Part II, *Conference on Regulation,* conducted by the National Bureau of Economic Research for the Board of Governors of the Federal Reserve System, 1957.

So far, we have been considering means of stabilizing consumption during the business cycle. Our conclusions may be summarized as follows: Automatic stabilizers should be used to reduce fluctuations in disposable income and consumption. The most important of these stabilizers available to the government are unemployment insurance and the corporation and personal income tax. Tax rates may also be changed, though we do not favor frequent or drastic changes in the structure of tax rates. When inflation is a serious threat, tax rates should be increased and perhaps supplemented by a compulsory savings program. In serious depressions, consumers' incomes will need to be supplemented beyond the aid provided by the automatic stabilizers. In addition to tax reduction, we recommend that some sort of consumers' subsidy be used on such occasions. For example, low-income groups generally, or the unemployed in particular, might be given stamps permitting them to obtain specified commodities free of charge or at reduced prices. The government would then redeem the stamps received by retailers. One great virtue of this arrangement is that it can be used to stimulate the demand for particular commodities that are notably in excess supply. The plan was effectively used in the 1930's in this country to move surplus food products.

THE LONG-RUN LEVEL OF CONSUMPTION

We must now consider the possibility that it may be necessary to control the long-run level of consumption, not merely its short-run fluctuations. The tools that are available are those already discussed. If the trouble is a chronic deficiency of demand, income taxes can be lowered, the tax structure can be made more progressive to reduce the saving of the rich, social security and social welfare benefits can be increased, etc. If the trouble is secular inflation, taxes can be raised and possibly made more regressive, and special inducements can be provided for saving. If necessary, a compulsory saving plan can be introduced.[13]

For two decades or more, the danger facing the nations of the Western world has been inflation, not depression. The control of inflation, if created by excess demand, requires that private spend-

[13] Nicholas Kaldor has also suggested that a tax on personal *spending* be substituted for the present system of progressive income taxation. See his *An Expenditure Tax*, 1955.

ing, including consumers' spending, be controlled.[14] The important means of curbing consumption during an inflationary period have been mentioned—all but one. That is rationing, with which direct price control is usually associated. We save this for last because we look upon it as an emergency device. Rationing violates a condition that we should like to observe—that a stabilization program should interfere as little as possible with the way an individual chooses to spend his income. Rationing means specifying the maximum amounts of various types of goods a consumer may buy.[15] Since it represents a kind of detailed control we want to avoid, rationing should be used only in emergency situations. In time of war it is unavoidable. But a country dedicated to the principles of personal freedom should think long before inaugurating a large-scale rationing system as a more or less permanent means of combating chronic inflation.

CONTROL OF PRIVATE INVESTMENT

Of the possible paths to economic stability, the most difficult is through the stabilization of private investment. As we have repeatedly seen, this is the most unruly of the components of aggregate demand. The very nature of technological progress causes private investment to occur in spurts, and the volatility of business expectations adds a further element of instability. Nonetheless, something can be done. Greater stability and steady growth in the remaining components of aggregate demand would tend to make investment somewhat more stable. And some direct steps can be taken to moderate the extreme fluctuations in private investment.

Since we are fairly certain that the level of investment reached at the top of a marked boom could in no case be maintained for very long, stabilizing investment is likely to mean holding back investment during boom periods as well as raising it during depressions. Thus it is possible that, with the existing propensity to con-

[14] The case of a "cost-push" type of inflation is discussed in Chapter 21.

[15] This is not true of "expenditure rationing," whereby the government specifies the maximum amount of money a household may spend on consumption goods during a given period. Within his spending ration, an individual may spend his income as he chooses. E. S. Shaw and Lorie Tarshis have proposed that the United States adopt expenditure rationing if it is again forced into an all-out economic mobilization. See their "A Program for Economic Mobilization," *American Economic Review*, vol. 41, March, 1951, pp. 30–50.

sume, the controlled level of investment during booms (reduced to conform to the estimated long-run trend) would not be high enough to maintain full employment.[16] If that should be the case, steps would have to be taken to raise permanently consumers' and government spending.

Private investment is not a homogeneous total. In particular, we have to distinguish (1) business expenditures on plant and equipment, (2) residential construction, and (3) investment in inventories.[17] Each of these presents problems of its own and is best treated separately from the others.

First, however, mention should be made of one type of stabilization policy that can have some effect on all the types of investment mentioned. This is monetary policy. Credit can be made more readily available when it is desired to stimulate investment, and credit can be tightened in order to hold back an investment boom. Because of the widely ramifying effects of monetary policy, we shall reserve detailed treatment for a separate section.

PLANT AND EQUIPMENT

Several things can be done to stimulate business long-term investment during periods of low activity. Among these, tax incentives are probably most important. One suggestion is that firms be allowed to deduct from their taxable income all or a fraction of any amounts invested in new plant and equipment.[18] The government might also provide that new plant and equipment acquired in depression periods could be depreciated at an accelerated rate, which would have the effect of reducing the firm's taxes during subsequent years. Both of these concessions could be withdrawn in boom periods, when it is no longer necessary to stimulate investment. Another tax measure that has been generally recommended provides that business firms be permitted to carry forward losses for a number of years in computing their tax liability. This would permit de-

[16] Cf. United Nations, *National and International Measures for Full Employment,* 1949, p. 36.

[17] The problem of foreign investment will be dealt with in a later chapter.

[18] Cf. M. Kalecki in Oxford University Institute of Statistics, *The Economics of Full Employment,* 1947, pp. 45–46, 48; Hansen, *op. cit.,* p. 143. The United Kingdom has used such a system of "investment allowances," and several other countries, notably Sweden, have made use of tax policy to stimulate or restrain private investment.

pression losses, and also the losses normally to be expected in the early years of a new firm, to be counted against the profits of later years and would clearly have a stimulating effect on the long-term level of investment, since the incentive would exist in good years as well as bad. To the extent that the incentive operated more strongly in depression than in prosperity, it would also have a stabilizing effect.[19]

In inflationary periods the problem is to restrict private investment. Monetary control can serve a useful function here. Some countries have restrained investment by some form of an investment tax, and some have directly controlled the flotation of new securities. If the inflationary stimulus is very strong, and particularly if private investment is competing with the government for strategic materials needed for military or other vital purposes, it may become necessary to use direct controls, as is done in wartime. In such cases, a firm may have to have a permit, license, or priority rating in order to acquire new plant and equipment. In a private-enterprise economy, such direct controls would be used only in times of national emergency.

If the need is to stimulate the long-run level of investment (as well as to stabilize short-period fluctuations), some further steps can be taken. The tax incentives to investment already mentioned could be broadened and made permanent. The capital-gains tax could be liberalized. Taxes on business profits could be reduced.[20] Special institutions could be created to provide long-term funds for small business and for other types of concerns not having ready access to the capital markets. And so on. The government could also subsidize scientific and industrial research more heavily in an effort to create new investment opportunities. Reform of the patent laws in order to prevent monopolistic control of new productive meth-

[19] For more detailed treatment of the effect of tax incentives on investment, see E. Cary Brown, "Business-Income Taxation and Investment Incentives," in *Income, Employment and Public Policy: Essays in Honor of Alvin H. Hansen,* 1948, and the references there cited; also Joint Committee on the Economic Report, *Federal Tax Policy for Economic Stability and Growth: Papers Submitted by Panelists . . .* (84th Congress, 1st Session, 1955), esp. chaps. 3 and 10, and Richard Goode, "Accelerated Depreciation Allowances as a Stimulus to Investment," *Quarterly Journal of Economics,* vol. 49, May, 1955, pp. 191–220.

[20] To the extent that reduced tax rates on profits or on high incomes generally increased saving more than investment, the result would be deflationary rather than stimulating.

ods would be another beneficial step, as would be the elimination of any monopolistic situations that tended to restrict expansion of capacity or to hold up the prices of particular capital goods.[21]

RESIDENTIAL CONSTRUCTION

In 1933, residential building was less than 10 percent of the peak rate reached in the 1920's. After World War II, another vast housing boom got under way. Obviously, we need to do something to eliminate the wild gyrations in this essential industry, which in recent years has accounted for 25 percent or more of gross private domestic investment.

We have learned a great deal about how to *stimulate* residential building in recent years. The record indicates that we have been less successful in *stabilizing* this type of activity, although so far we have been able to hold postwar fluctuations in house building within moderate limits. Some success has also been achieved, at least during the first three postwar cycles, in imparting to this type of construction a partially countercyclical character. Residential building has been restrained in the late stages of a cyclical boom, as we saw in Chapter 10, and it has tended to expand during the latter part of cyclical contraction. The chief steps that have been taken center around the terms on which mortgage credit is made available to borrowers and the protection given to lenders against default. The regulation of the terms of mortgage lending has been strongly reinforced by a flexible monetary policy. The supply of funds available for guaranteed and insured mortgages tends to be reduced during the later stages of a boom, when high interest rates make other forms of investment more attractive to lenders; and the reverse tends to happen during a recession.[22]

[21] For further discussion of possible ways of stabilizing and stimulating business investment, see Subcommittee on Investment, Joint Committee on the Economic Report, *Factors Affecting Volume and Stability of Private Investment: Materials on the Investment Problem Assembled by the Staff of the Subcommittee* (81st Congress, 1st Session, 1949). Policies to stabilize investment in various European countries are summarized in Angus Maddison, "The Postwar Business Cycle in Western Europe and the Role of Government Policy," *Banca Nazionale del Lavoro Quarterly Review,* June, 1960, pp. 99–148.

[22] Thus the sharp upsurge in residential building in the 1954 recession seems to have been due more to monetary ease than to legislative or administrative action liberalizing the terms for FHA or VA mortgages. See Leo Grebler, *Housing Issues in Economic Stabilization Policy,* National Bureau of Economic Research Occasional Paper 72, 1960, pp. 21–31.

The federal government now has an extensive system of mortgage insurance and guarantees. Mortgages are eligible for underwriting by the government against default only if they meet specified conditions as to interest rate, maturity, regular amortization of the principal, and ratio of the mortgage to the appraised value of the property. All of these conditions can be made more or less stringent. If building activity is low, the interest rate and down payment on insured mortgages can be reduced and the repayment period can be lengthened. If building activity is thought to be too high, the terms can be made more restrictive. This type of control can be extended to noninsured mortgages also. In the fall of 1950 the federal government did precisely this. As a step toward meeting the inflationary threat created by the Korean crisis, the terms for insured mortgages were substantially tightened, and similarly stringent conditions were imposed on the issuance of noninsured mortgages.

Another important variable involved in the cyclical behavior of building activity is construction costs. Building costs—both for labor and materials—are notoriously inflexible in depressions. If there has not been serious overbuilding during the preceding boom, a substantial reduction in building costs may provide an important stimulus to building during a depression.[23] However, it may not be easy to reduce building costs. Trade unions will resist wage reductions; monopolistic restrictions in the industry may be difficult to overcome; and various types of cost-increasing practices have been incorporated into local building regulations and the customs of the industry.

One factor that accentuated the building depression of the early 1930's in the United States was the disorganization of the mortgage market. Widespread defaults on mortgages, which had frequently been unwisely made, jeopardized the solvency of many financial institutions and made lenders unwilling to assume new risks. The present system of mortgage insurance and the general practice of requiring regular amortization largely correct this difficulty. A further step in this direction is the provision of a "secondary mortgage

[23] Examples of this stimulus in the Netherlands and Sweden are cited in League of Nations, *Economic Stability in the Post-War World*, 1945, pp. 140–141. A decline in building costs seems to have had a similar effect in the United States in 1921.

market," in which the government offers, under certain conditions, to take over mortgages from lenders.[24]

At this point we must offer a sharp warning. *Stimulation* does not mean *stabilization.* Until the Korean crisis, the effect of government intervention in the mortgage market in the United States was to make it increasingly easy to finance the building and purchase of homes. By virtue of lower interest rates and long amortization periods, monthly carrying charges were substantially reduced, and so was the necessary down payment. The result was to accentuate the housing boom after World War II. Not until the Korean crisis, in the fall of 1950, was the government's control of the availability of mortgage credit used to reduce the volume of building. And it was not until 1955 that "a peacetime policy of flexible management of the federal housing credit programs in the interest of economic stability was clearly enunciated."[25] Before then, except in wartime emergencies, federal housing policy had been aimed primarily at stimulation rather than stabilization.

The control of mortgage credit should be used with some goal in mind as to the amount of private building the economy wishes to maintain. If a housing boom exceeds this level and overcapacity is likely to develop, credit should be tightened. If building activity falls off, mortgage terms should be made more lenient. Here again we may run into a conflict of objectives. If a housing shortage coincides with a general inflationary boom, the desire to provide adequate housing for the population will conflict with the need to restrict investment in the interest of stability.[26]

Although the measures we have suggested can help to iron out some of the extreme fluctuations in building activity, they cannot provide the whole answer. Building booms will continue to occur,

[24] In recent years the Federal National Mortgage Association (FNMA) has provided a secondary market for federally guaranteed home mortgages. In addition, there has existed for some years in the United States a system of Federal Home Loan Banks which, very roughly, bear the same relation to savings and loan associations that the Federal Reserve Banks bear to commercial banks.

[25] Grebler, *op. cit.,* p. 3. Grebler's entire study is recommended for its careful analysis of federal policy toward residential building.

[26] Building activity can be controlled *directly,* through a system of permits and licenses, as well as *indirectly* through general monetary policy or regulation of mortgage credit. For an evaluation of Sweden's experience with direct control of building activity through the issuance of permits, see Eric Lundberg, *Business Cycles and Economic Policy,* 1957, esp. pp. 278–283.

to be followed by periods of reduced construction activity. The level of building activity is a function of other variables besides the cost and availability of credit and building costs—for example, the rate of growth of population, population shifts, the level of incomes, and the size and condition of the existing stock of houses. Given the character of the building industry, these variables will continue to cause fluctuations in the volume of residential construction, although careful planning should be able to reduce the amplitude of these fluctuations substantially.[27]

In view of the need for more and better housing in all countries, permanent stimulation of this type of investment seems desirable. Some of the means available to stimulate private building have already been mentioned: provision of mortgage insurance, reducing the cost of borrowing, and forcing down building costs. Among other things that can be done, we might mention tax legislation favoring builders and home owners, government aid in making land available, and government subsidies for some types of construction.[28] But even if the long-run objective is a permanently higher level of house building, the government should always be prepared to withdraw its stimulus temporarily if an investment boom threatens to get out of hand or if an excess supply of dwellings is in danger of developing.

INVENTORY INVESTMENT

We have seen that changes in inventories are an important element in short-run instability. Since fluctuations in this form of investment are to a considerable extent the result of changing

[27] For further discussion of the problem of stabilizing residential construction, see, in addition to Grebler's study, League of Nations, *Economic Stability in the Post-War World,* 1945, pp. 137–143; M. L. Colean and R. Newcomb, *Stabilizing Construction,* 1952; M. L. Colean, "Contracyclical Aspects of Private Residential Building," in Committee for Economic Development, *Problems in Anti-Recession Policy,* multilithed, 1954; Leo Grebler, "Housing Policies to Combat Depression," in Universities-National Bureau Committee for Economic Research, *op. cit.,* pp. 241–255. For recent European experience, see Maddison, *op. cit.,* pp. 122–123.

[28] In addition, the government can institute its own public housing program, discussion of which we omit at this point because we are here concerned only with private investment. Because of the relatively high cost of satisfactory shelter, a government-subsidized program including public housing seems necessary to provide adequate dwellings for the lowest income groups, particularly if slum clearance and urban redevelopment are involved.

expectations as to prices and sales, any measures that succeed in making output and prices more stable are likely to have some effect in moderating inventory fluctuations. A vigorous, countercyclical monetary policy is likely to have a direct influence on business inventory policies in the right direction. In particular, it is important to keep the economy highly liquid if a contraction begins so that business firms will not feel compelled to liquidate inventories in order to obtain cash.[29] The supplying of detailed information regarding current and prospective business conditions, by both government agencies and trade associations, may help to prevent excessive accumulation of inventories and to keep production in various industries closely geared to supply. The government and private business groups should also encourage the practice of using inventories as a buffer in order to minimize fluctuations in production and employment. Particularly where the products are nonperishable staples, firms can produce for stock in slack times and then reduce their inventories when demand improves. This policy introduces a desirable countercyclical variability into inventories.

MONETARY POLICY

We pointed out earlier that one way of influencing the components of aggregate demand is through monetary policy. We must now consider in more detail how and to what extent monetary control can contribute to economic stability. This is an area in which there has been almost a full cycle of opinion in the last forty years. In the 1920's, monetary policy would have been given first place in any professional discussion of stabilization measures. Then came the Great Depression, the development of Keynesian analysis (with its corollary that fiscal policy was the most effective way of controlling the level of spending), and the rapid rise, beginning in the 1930's, in the size of government budgets in both the United States and other countries. As a result, stabilization policy in the 1930's and 1940's came to mean chiefly fiscal policy and, in some countries, various sorts of direct controls even in peacetime.[30] More

[29] In suggesting that monetary policy can influence the accumulation and liquidation of inventories, we are not accepting Hawtrey's thesis regarding the causal importance of variations in the short-term interest rate. In our view, it is the availability of short-term credit, not its cost, which is important. To this should be added the psychological effect on business sentiment of any important action taken by the monetary authorities.

[30] Compare the review of business-cycle developments in chaps. 14–16.

recently, since the late 1940's or early 1950's, there has been a renewed emphasis on monetary policy on both sides of the Atlantic —both in economists' thinking and in actual government policy— although it is probably true that most economists still assign to monetary measures a secondary place in their proposals for stabilizing the economy. The "rediscovery of money" that has occurred has come about particularly because of the problems that have been encountered in dealing with the inflationary forces that were unleashed by World War II.[31]

TECHNIQUES OF MONETARY CONTROL

The most important instruments of monetary policy should be familiar to every student of economics. They are: changes in the rediscount rate of the central bank, open-market operations, changes in member-bank reserve requirements, and selective credit controls.[32] The first three of these operate on the reserves of member banks and thus influence the ability and willingness of the banks to expand credit.

The rediscount rate is a relatively ineffective weapon. In the United States, the commercial banks have been substantially out of debt to the Federal Reserve Banks since the 1930's, and rediscounts constitute only a small fraction of total Federal Reserve credit outstanding.[33] For a decade following 1933, the commercial banking system held substantial excess reserves. Banks do not like to borrow from the Federal Reserve system if it can be avoided, and they can ordinarily strengthen their reserve position by selling government securities rather than by borrowing. As we saw in Chapter 10, when the demand by business for loans rises during cyclical expansions, commercial banks typically sell large amounts of securities.

[31] See, for example, H. S. Ellis, "The Rediscovery of Money," in *Money, Trade, and Economic Growth: In Honor of John Henry Williams,* 1951, pp. 253–269. For a useful survey of recent thinking about monetary policy, see The American Assembly, *United States Monetary Policy,* 1958. For a brief survey of the instruments of monetary policy used in various European countries in the postwar period, see Maddison, *op. cit.,* pp. 132 ff.

[32] The central bank can also exercise direct control over the commercial banks in order to influence the amount or kinds of loans the latter make. This has been done, for example, in the United Kingdom.

[33] Thus, of total "Reserve Bank credit outstanding" at the end of 1959 of 29.4 billion dollars, 27 billion represented U.S. government securities held by the Reserve banks and only 900 million represented discounts and advances.

Hence, raising the Federal Reserve rediscount rate does not make it significantly more difficult or expensive for most member banks to obtain additional reserves if they want them. However, changes in the rediscount rate do have some psychological value, particularly if the changes are large and are accompanied by considerable publicity. In this case, short-term interest rates in the money market are likely to be affected, and so may be the attitude of banks toward making new loans.

Open-market operations—that is, the purchase and sale of securities by the central bank—are the most effective instrument of control available to the monetary authorities. When the central bank sells securities, whether to commercial banks or to the nonbanking public, member-bank reserves are reduced by a corresponding amount.[34] If the central bank buys securities, member-bank reserves rise. The potential power that the ability to engage in large-scale open-market operations gives to the central bank may be illustrated by the situation in the United States at the end of 1959. At that time, the Federal Reserve Banks owned about 27 billion dollars in government securities. Total member-bank reserves were about 19 billions. The Federal Reserve Banks thus had the power, at least in theory, to wipe out the reserves of member banks to any extent they wished. And with reserve ratios averaging about 15 or 16 percent, member banks would have had to contract deposits by roughly six times any decline in their reserves. Conversely, the reserve banks could have increased bank reserves by adding further to their holdings of government securities. Actually, as we shall see shortly, the ability of the Federal Reserve to act was considerably more limited than this suggests. Because of the need to maintain orderly conditions in the market for government bonds, the monetary authorities were not completely free to sell government securities when and in what amounts they wished in order to influence bank reserves. When a large government debt is outstanding, monetary policy and management of the public debt become inseparably intertwined.

The Treasury, in its fiscal operations, can also influence member-bank reserves. Whenever the Treasury shifts its deposits from

[34] If the nonbanking public buys the securities, demand deposits as well as reserves are reduced. If the sales are made directly to banks, only reserves are affected.

commercial banks to the Federal Reserve Banks, bank reserves decline. Collection of taxes from the public or the sale of bonds to nonbanking investors not only reduces private demand deposits but also, if the proceeds are deposited at the Federal Reserve, causes a corresponding decline in bank reserves. The reverse of these operations would cause bank reserves to increase.[35]

The Federal Reserve authorities have at their disposal another way of influencing the action of member banks. Under the Federal Reserve Act, as amended in 1935, member-bank reserve *requirements* can be varied between a minimum of about 10 percent and a maximum of about 20 percent.[36] Thus, without changing at all the total amount of reserves, the Federal Reserve can create excess reserves for member banks by reducing reserve requirements. Conversely, an increase in reserve requirements wipes out excess reserves and, if the increase is large enough, may force member banks to convert earning assets into additional reserves in order to meet the higher requirements. The effectiveness of this control device depends in large part on the range over which reserve requirements can be varied. Ever since 1937, reserve requirements have been at or close to the maximum permitted by law. Hence, the Reserve authorities were not able to increase them much further when inflationary forces threatened to get out of hand after World War II.[37]

Various proposals have been made to restrict further the ability of commercial banks to engage in multiple credit expansion on the basis of fractional reserves. The most extreme suggestion, which has been advanced by several well-known writers, is the proposal that banks be required to hold reserves equal to 100 percent of their demand deposits.[38] Since this would deprive banks of their

[35] If the Treasury sells new securities to the banks and deposits the proceeds in the Federal Reserve, member-bank reserves are reduced, but there is no reduction in the public's deposits. When the Treasury redeems securities held by the banks, bank reserves are increased if the Treasuty uses its balances at the Federal Reserve but not if it uses its deposits at the commercial banks.

[36] In the case of demand deposits, the minimum reserve ratios since July, 1959, have been 7, 10, and 10 percent for country, reserve city, and central reserve city banks, respectively. The maxima are 14, 22, and 22 percent.

[37] In the late 1950's, reserve requirements averaged a few percentage points below the legal maximum.

[38] See, for example, Irving Fisher, *100% Money*, 1935. For a brief discussion of the proposal, see Hart, *op. cit.*, pp. 437–439.

earning assets, deposit banking would have to be divorced from the business of making loans. Though the proposal has considerable merit, many economists "doubt that this measure would make a significant contribution to economic stability, and feel that the transitional difficulties of such a drastic change would be great and that less drastic measures can put the banking system on a satisfactory footing."[39]

The fact that commercial banks today hold large amounts of government bonds creates a special problem which orthodox monetary measures were not designed to handle. If the Federal Reserve is under obligation to buy these securities in order to support government bond prices, the banks can secure additional reserves whenever they need them. It would do no good to increase cash-reserve requirements with one hand and with the other to supply additional reserves to the banks by buying bonds from them. If the central bank stands ready to buy at fixed prices any securities the commercial banks want to sell, this means, in effect, that bank reserves have been "open-ended"; banks can convert bonds into cash reserves at will. This was the sort of situation that did prevail in the United States during the first few years after World War II. We shall have more to say about this problem later in this chapter.

To help meet this kind of problem, the Board of Governors of the Federal Reserve System and some economists have suggested that an additional reserve requirement, to be held in the form of government securities, be imposed on member banks.[40] This would immobilize a large part of the banks' holdings of government bonds (especially their short-term issues) so that they could not be converted into cash reserves. The effect would be the same as if cash-reserve ratios were raised, except that banks would be able to earn interest on that part of their reserves held in the form of government securities. This measure has considerable merit and might

[39] Committee of the American Economic Association, "The Problem of Economic Instability," *American Economic Review*, vol. 40, September, 1950, p. 528.

[40] For details of the proposal made by the Federal Reserve to Congress, which the latter never accepted, see *Federal Reserve Bulletin*, January, 1948, pp. 14–23. See also E. C. Simmons, "Secondary Reserve Requirements for Commercial Banks," *American Economic Review*, vol. 41, March, 1951, pp. 123–138, and Ervin Miller, "Monetary Policy in a Changing World," *Quarterly Journal of Economics*, vol. 70, February, 1956, pp. 38–42. Security reserve plans have been used in a number of other countries.

make a significant contribution to a resolution of the conflict that has existed between the stabilization objective (which calls for a tightening of monetary controls in times of inflation) and public-debt management aimed at supporting the prices of government bonds and keeping interest rates low.

In addition to the measures so far mentioned, the monetary authorities may also make use of "selective credit controls," which regulate the extension of credit for particular purposes. We have already referred to this type of regulation in discussing ways of controlling consumers' expenditures and residential building. Thus, the Federal Reserve had the power during most of the period from 1941 to 1952 to regulate the terms on which credit was granted to consumers. It has also had the power since 1934 to regulate margin trading in securities. Control of the terms on which insured mortgages may be granted has been exercised by the housing-finance authorities since the middle 1930's, and during the Korean conflict the Federal Reserve issued similar regulations with respect to non-insured mortgages.

Other types of credit may be regulated in similar fashion. Tightening up of selective credit controls tends to restrict particular types of spending, and relaxation of these controls acts as a stimulant. At best, unless their use is greatly extended, controls of this sort can be no more than a secondary weapon. They have the advantage of aiming directly at particular types of spending which, for one reason or another, may be especially in need of regulation. However, they have the undesirable feature of being discriminatory in character, since their aim is to control the extension of credit for particular purposes.[41]

EFFECTIVENESS OF MONETARY CONTROL

Insofar as monetary policy can make a contribution toward economic stability, it does so through influencing the components of private demand, particularly business investment. This influence is exerted through the supply of money, especially the volume of bank deposits, and through the terms on which loanable funds are

[41] For more extended discussion of selective credit controls, see the paper by Arthur Smithies in The American Assembly, *op. cit.*, pp. 73–89; also the references cited in footnotes 12 and 22.

made available (including the level and structure of interest rates).[42]

Monetary policy is much more effective in curbing a boom than in helping to bring the economy out of a depression. Given the necessary weapons, the central bank can always contract the money supply sufficiently to end any boom. Through open-market operations particularly, bank reserves can be reduced to the point where the banks can no longer make new loans and are forced to call outstanding loans. Thus, the volume of bank deposits will decline and, unless there is an offsetting rise in velocity, so will total spending. Since the supply of loanable funds will have been reduced, credit will be less readily available and interest rates will rise; and this will have a further discouraging effect on business spending. There will be a general scramble for liquidity; firms will try to sell more than they produce; *ex ante* investment will become less than *ex ante* saving.

If the medicine is administered in large enough doses, monetary policy can stop a boom. The trouble is that it is likely to do more than merely stop it. As the preceding paragraph suggests, monetary contraction can easily precipitate a downswing. This is the chief danger in using monetary policy alone to control an inflationary boom: monetary action strong enough to stop the boom is almost certain to be too strong to keep the economy at the full-employment level.

Though monetary policy should not be relied on exclusively to control a boom, it obviously has an important role to play during periods of rapid expansion. If the economy's propensity to spend is rising, the banks will be under pressure to expand their loans and investments and thus to add to the supply of money in the hands of businessmen and consumers. If the monetary authorities can keep the supply of money from expanding, or from expanding

[42] The debate on the effectiveness of monetary policy still goes on, and it has spawned a very large literature. For recent examples, see the papers by H. S. Ellis and A. G. Hart in The American Assembly, *op. cit.;* W. L. Smith, "On the Effectiveness of Monetary Policy," *American Economic Review,* vol. 46, September, 1956, pp. 588–606; Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st Session, 1959), chap. 9; and the numerous other references cited in these sources. For a British view, see *Report of the Committee on the Working of the Monetary System* (Cmd. 827 [1959]), known as the "Radcliffe Report."

any faster than output rises, the increase in spending will to that extent be held in check. And if borrowers find that loans are available only on more stringent terms, the expansion in spending will be at a slower pace than if credit were readily available at low interest rates. The monetary pressure should be moderate; and other weapons, particularly fiscal policy, should be used at the same time. But the monetary pressure should be applied.

Monetary policy is less effective in depressions than in booms. The old saying that "you can lead a horse to water but you can't make him drink" expresses the difficulty. During depressions, the central bank can feed excess reserves into the banking system, but it cannot force the banks to grant new loans. Even if the banks are willing to extend credit on favorable terms, businessmen may not be willing to borrow if they are pessimistic as to the future. We know that low interest rates alone are insufficient to stimulate investment when business sentiment is depressed.[43]

When a cumulative business contraction has got under way, monetary policy *alone* is not enough to reverse the decline and bring about an expansion in aggregate demand. Other ways must be found to increase total spending. Even in this case, however, monetary policy has an important role to play. We have noted before the scramble for liquidity that takes place during business downswings. It should be the job of the monetary authorities to offset to the maximum extent possible this rise in liquidity preference. The central bank should pump reserves into the banking system in order to induce banks not to liquidate their earning assets; through large open-market purchases, it can supply cash to the nonbanking public as well as to the banks; by lowering interest rates and holding up bond prices, it can help to maintain confidence in the financial and business community and thus reduce the general desire to hold cash rather than securities and goods. By prompt and vigorous action, the central bank can reduce the strength of the cumulative deflationary forces at work—though it cannot eliminate them entirely. If the central bank does this much, then the job to be done by other types of measures—aimed directly at stimulating the vari-

[43] An easy money policy may stimulate some types of long-term investment during a business recession if underlying opportunities are still favorable. This has been particularly true of residential building in the recessions since World War II. See p. 568 and footnote 22, above.

ous components of total demand—will have a much better chance of success.

MONETARY POLICY IN THE POSTWAR SETTING

Discussion of the effectiveness of monetary policy cannot be divorced from consideration of the changed environment within which the monetary authorities now operate. We have already touched on the nature of some of these changes, particularly those associated with the way government spending was financed during World War II. What we need to look at in particular are: (1) the great increase in the public debt since the early 1930's, (2) the excess liquidity with which the American economy emerged from World War II, and (3) the increasingly important role being played by nonbank financial intermediaries (such as insurance companies and savings and loan associations).[44]

THE PUBLIC DEBT AND EXCESS LIQUIDITY

At the end of World War II, the federal debt was nearly 280 billion dollars, and in the 15 years following it showed a further net increase of about 10 billions. In 1939, the figure was less than 50 billions; it was 16 billion in 1929. Of the total of 291 billions at the end of 1959, about 60 billions was owned by the commercial banks; at the end of World War II, the banks owned more than a third of the federal debt outstanding. For a considerable period, from 1941 to 1952, the banks' holdings of government securities exceeded their outstanding loans. (See Figure 45.) Figure 45 brings out also the tremendous increase in bank loans that has occurred since World War II. During the first half dozen of the postwar years, the expansion in bank loans was to a considerable degree made possible by the ability of the banks to dispose of a substantial part of their inflated holdings of government securities. During this period, the Federal Reserve system was under an obligation to support the prices of government bonds, an obligation that was finally removed by the Treasury—Federal Reserve Accord of 1951.

The size of the federal debt means that monetary policy cannot be dissociated from problems of debt management.[45] The easing and

[44] For a useful survey of postwar monetary policy and of the problems faced by the monetary authorities, see Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels, op. cit.*, chap. 9.

[45] See, for example, W. L. Smith, *Debt Management in the United States,* Study Paper No. 19 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels, op. cit.*

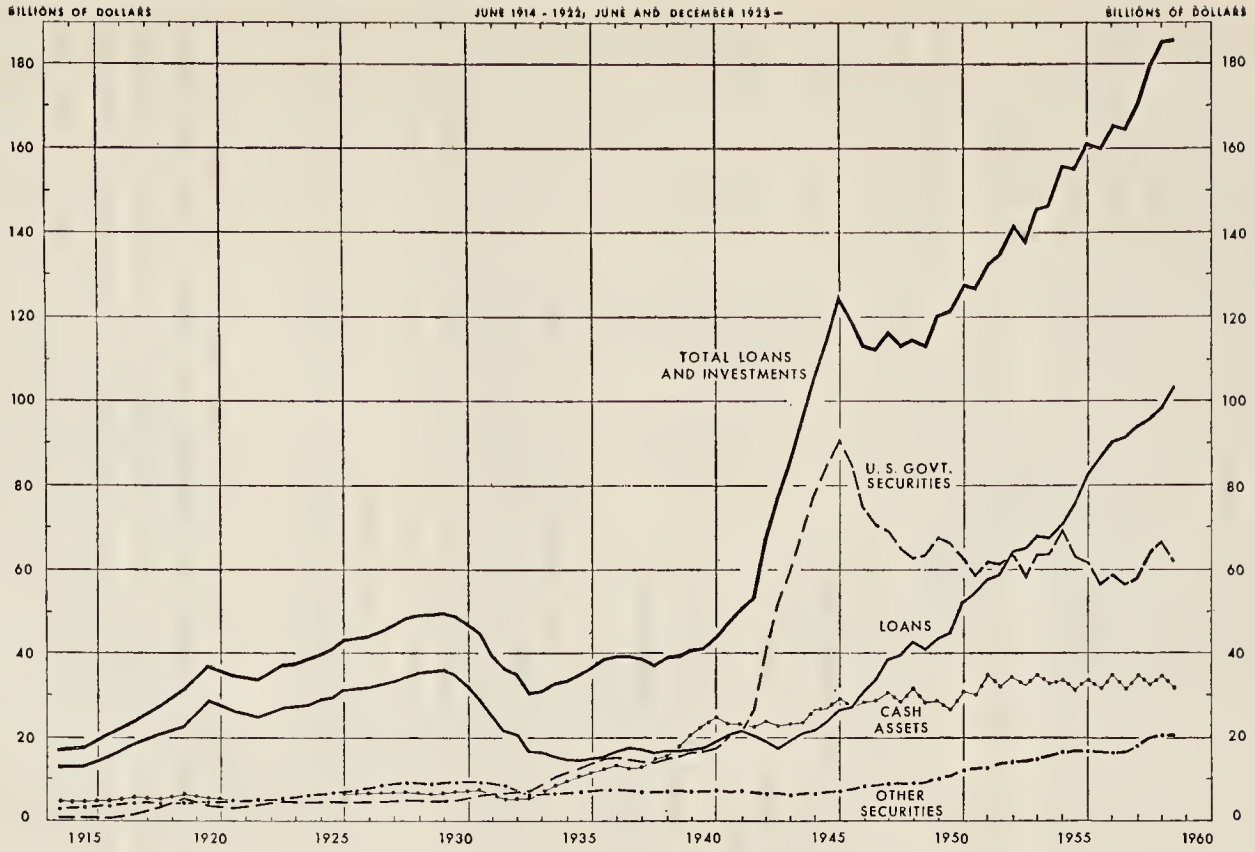


FIGURE 45. Principal Assets of All Commercial Banks, 1914–1959.

From *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics.*

tightening of credit leads to corresponding changes in interest rates and bond prices, and these changes affect the market value of the billions of dollars of government bonds held by individuals, business firms, banks, and other financial institutions. Further, a substantial fraction of the debt is represented by short-term securities that regularly have to be replaced by new Treasury offerings. The terms on which the Treasury can sell new securities depends in good part on the kind of monetary policy the Federal Reserve authori-

TABLE 38. Money Supply, Liquid Assets, and GNP, 1939–1958
(In billions of dollars) [a]

	1939	1945	1950	1958
1. Adjusted demand deposits plus currency. [b]	36.2	102.3	117.7	144.2
2. Liquid assets of nonfinancial sectors. [b,c]	81.0	250.1	286.2	394.5
3. GNP	91.1	213.6	284.6	441.7
4. Income velocity: (3) ÷ (1)	2.52	2.09	2.42	3.06
5. Ratio of GNP to liquid assets: (3) ÷ (2)	1.12	.85	.99	1.12

[a] Lines 1 and 3 from *Economic Report of the President*, January, 1960. Line 2 is from J. G. Gurley, *Liquidity and Financial Institutions in the Postwar Period*, p. 5, with the following modification: I have not included life insurance policy reserves, and I have added in marketable U.S. government securities held by all except financial institutions (which Gurley excludes). The latter were obtained from the *Federal Reserve Bulletin*, August, 1959, for the years from 1945 on. The figure for 1939 is based on ownership of the government debt as given in the *Economic Report of the President*.

[b] Taken for December 31 of the year indicated.

[c] Includes currency, demand and time deposits, mutual savings deposits, savings and loan and credit union shares, postal savings deposits, and U.S. government securities. Liquid assets of financial institutions are excluded.

ties are following, and the latter in turn cannot completely disregard the Treasury's needs in deciding on what course to follow.

The commercial banks swollen holdings of government securities constituted only a part of the excess liquidity with which the economy emerged from World War II. Total demand deposits and currency in the hands of the public approximately trebled between 1939 and 1945. The rise was more rapid than that in the national income, so that at the end of World War II the public held much larger cash balances relative to its income than was the case in 1939. (See Table 38.) In addition to its holdings of deposits and currency,

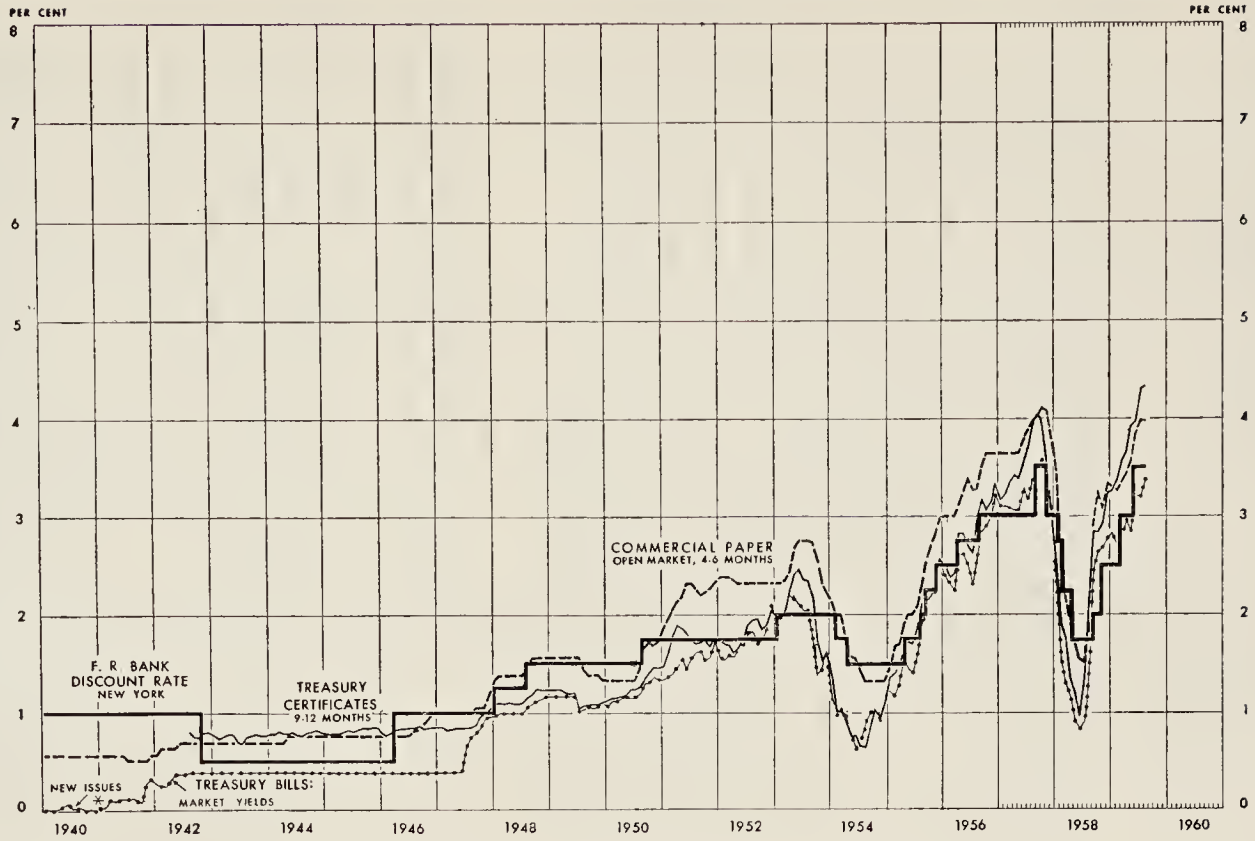


FIGURE 46. Short-Term Interest Rates, 1940–1959.

From *Historical Supplement to Federal Reserve Chart Book on Financial and Business Statistics.*

the public's holdings of other kinds of liquid assets, particularly government securities, had increased greatly. At the beginning of the postwar period, the ratio of nonbank holdings of liquid assets to GNP was considerably larger than it had been before the war, as can be seen from Table 38.

Given this excess liquidity, it was inevitable, once wartime controls were abolished, that private spending would expand rapidly —more rapidly than output could increase. This was the period of inflation, characterized by excess demand, that continued, with an interruption for the 1949 recession, until the latter part of 1951. As the GNP (in current prices) expanded more rapidly than the money supply, income velocity rose; and, from about 1950 on, interest rates increased. (See Figure 46.) It was not until some time in the 1950's that the excess liquidity in the economy was eliminated. Until it was, the ability of the monetary authorities to hold back private spending was severely restricted. In effect, households and firms were able in good part to finance an increase in spending by using excess cash they already had or by converting other liquid assets into cash, chiefly by transferring them to others who had excess cash to invest. The counterpart of this process was an increase in the velocity of the money supply.

FINANCIAL INTERMEDIARIES

The problems of the monetary authorities have been made more difficult by the growing importance of nonbank financial intermediaries—insurance companies, savings and loan associations, credit unions, and the like.[46] To an increasing degree, savings are being "institutionalized." Individual saving is going into these intermediaries, which in turn purchase the securities of or make direct loans to the ultimate borrowers—business firms, the federal and state and local governments, and individuals for the purchase of homes and consumer durable goods. The development of these nonbank inter-

[46] This has been emphasized in particular by J. G. Gurley and E. S. Shaw. See their "Financial Aspects of Economic Development," *American Economic Review,* vol. 45, September, 1955, pp. 515–538, and "Financial Intermediaries and the Saving-Investment Process," *Journal of Finance,* vol. 11, May, 1956, pp. 257–276; also J. G. Gurley, *Liquidity and Financial Institutions in the Postwar Period,* Study Paper 14 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels, op. cit.*

mediaries raises a number of problems for monetary policy, which we can do no more than mention here.[47]

1. The obligations of some of these intermediaries are highly liquid assets (savings deposits or the shares of savings and loan associations, for example). Thus, the process of saving in these forms can increase the liquidity of the economy, even if the money supply is not increasing.

2. If the monetary authorities tighten credit and interest rates rise, nonbank intermediaries may raise the interest rates that they offer, leading some holders of demand deposits to convert the latter into the interest-bearing obligations of the intermediaries; and the latter can then lend these funds to those who wish to increase their spending. Thus, idle deposits are converted into active ones, and velocity consequently rises.

3. The lending activities of the intermediaries may be to some extent beyond the control of the monetary authorities. Thus, pressure by the central bank on member-bank reserves (for example, through open-market operations) may restrict the lending of the commercial banks but not of nonbank institutional lenders. The latter may seek to attract additional funds by raising the interest rate they pay; and they may sell some of their existing (particularly government) securities in order to increase their lending. All this amounts to saying that financial intermediaries are able to some extent to mobilize idle funds that can then be spent by those to whom they lend, even in the face of a tight money policy.

FEDERAL RESERVE POLICY BEFORE AND AFTER THE "ACCORD"

In the light of these various developments, it is not surprising that the traditional instruments of monetary policy were difficult to apply in the early postwar years. A very stringent credit policy would have been needed at the end of the war to cope with the excess liquidity that then characterized the economy. But the size of the public debt and the Treasury's continuing need to refund

[47] See Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels,* pp. 351–356. For a skeptical view of the extent to which financial intermediaries have actually weakened monetary controls, see W. L. Smith, "Financial Intermediaries and Monetary Controls," *Quarterly Journal of Economics,* vol. 73, November, 1959, pp. 533–553. Other references will be found in these two sources.

maturing obligations led the Federal Reserve authorities to agree to maintain an easy money policy that would support the price of government securities. As a result, from 1945 to 1951, the monetary authorities were unable to follow even a moderately tight credit policy.

All this was changed in 1951 by the Accord, which restored to the Federal Reserve authorities the freedom to pursue an independent monetary policy, subject to the obligation to maintain orderly conditions in the government securities market.[48] The Accord was preceded by the report of a Congressional committee that received wide attention. Few economists would disagree with the committee's conclusion "that the advantages of avoiding inflation are so great and that a restrictive monetary policy can contribute so much to this end that the freedom of the Federal Reserve to restrict credit and raise interest rates for general stabilization purposes should be restored even if the cost should prove to be a significant increase in service charges on the Federal debt and a greater inconvenience to the Treasury in its sale of securities for new financing and refunding purposes."[49] This recommendation for a more flexible monetary policy than we had during the 1940's does not imply that monetary policy can or should be expected to do the job alone, nor does it mean that the problems of managing a large public debt should be ignored. What it does mean is that the monetary authorities should not permit the desire for low and stable yields on government securities to force them into an expansionary credit policy if the existence of inflationary forces calls for a moderately restrictive credit policy to supplement other anti-inflationary measures that are being taken. Although in today's setting monetary policy alone cannot control inflation, neither should monetary policy work at cross-purposes with other more effective anti-inflationary measures that seek to hold back the expansion in private spending.

Since 1951, the Federal Reserve system has followed a fairly vigorous countercyclical monetary policy—putting pressure on bank reserves and causing interest rates to rise during business expansions

[48] Cf. *Thirty-eighth Annual Report of the Board of Governors of the Federal Reserve System,* for 1951, pp. 3–8, 98–101.

[49] *Report of the Subcommittee on Monetary, Credit, and Fiscal Policies of the Joint Committee on the Economic Report,* (81st Congress, 2nd Session, 1950), p. 2.

and shifting quickly to a policy of credit ease when business declined and prices stopped rising rapidly.[50] (See Figure 46.) This was the procedure followed during boom periods in 1952–1953 and 1955–1957 and during the recessions of 1953–1954 and 1957–1958. However, a tight money policy was not enough to stop the inflationary rise in prices that occurred during 1955–1957, and Federal Reserve policy cannot be given much of the credit for the period of price stability from 1951 through 1954.[51]

Criticism of Federal Reserve policy since the 1951 Accord has been chiefly along the following lines.

1. Since 1953, the Federal Reserve authorities have followed the policy of confining open-market operations primarily to short dated government securities (i.e., Treasury bills). This has come to be known as the "bills only doctrine" and has been criticized by many economists on the ground that it unduly impairs the ability of the monetary authorities to affect long-term interest rates. In general, the Federal Reserve's position is that, in attempting to control the supply of money through its ability to affect member-bank reserves, it does not wish arbitrarily to impose a particular pattern on the structure of interest rates or unduly disturb the government securities market by buying and selling intermediate and long-dated securities.[52] Critics argue that one objective of monetary policy should be to raise or lower those particular interest rates that are most likely to achieve the desired stabilization goal. Thus, if long-term investment in buildings and equipment declines during a recession, special efforts might be made to force down long-term interest rates below the levels they would otherwise reach.

2. It has also been argued that, even since 1951, monetary policy has been unable to control the expansion in liquid assets brought on by the increasing activities of nonbanking financial intermediaries. Thus, according to one recent study, of an increase in total liquid assets of something more than 100 billion dollars between

[50] For a review of American monetary policy during 1951–1959, see *Staff Report on Employment, Growth, and Price Levels, op. cit.*, pp. 324–344. For a detailed review of British monetary experience during the same period, see the Radcliffe Report cited in footnote 42.

[51] See the review of these years in Chapter 16.

[52] See *Fortieth Annual Report of the Board of Governors of the Federal Reserve System*, for 1953, pp. 88–89. For a defense of this policy, see R. A. Young and C. A. Yager, "The Economics of 'Bills Preferably,'" *Quarterly Journal of Economics*, vol. 74, August, 1960, pp. 341–373.

1952 and 1959, only 10 billion came from an increase in the money supply.[53] The rest represents a rapid expansion in time deposits, savings and loan and credit union shares, deposits at mutual savings banks, and policy reserves of life insurance companies. Federal Reserve policy has not been able to prevent this expansion, although it has effectively held in check the increase in the money supply. The possible inflationary consequences of this expansion of liquid assets is still being debated.[54]

3. The third criticism comes from those who believe that the American economy could and should have grown more rapidly than it did during the 1950's. These critics believe that the tight money policy followed by the Federal Reserve authorities during, for example, 1955–1957 unnecessarily held back the rate at which total output rose and led to more unemployment that was consistent with a full-employment goal.[55]

4. Others have argued that, whatever the effects of the monetary policy followed on the rate of growth, the Federal Reserve authorities have put their faith in anti-inflationary tools that were in fact incapable of controlling inflation, given the forces that were actually causing prices to rise. These critics believe that the rise in prices from 1955 on was due primarily to such factors as the upward pressure on wages from strong trade unions, the market power of large firms, special factors affecting the prices of services, and so on—factors that could not be controlled by monetary policy. A tight money policy, it is argued, can effectively curb inflationary pressures (without generating excessive unemployment) only if these pressures arise from excess demand and if prices have some downward flexibility.[56] We can restate this in terms of the analysis of aggregate demand and supply in Chapter 4. If the aggregate supply function shifts upward and prices rise because, for example, autonomous influences increase costs, a restrictive monetary policy that restrains aggregate demand reduces output and generates unemployment without preventing the rise in prices.

[53] J. G. Gurley, *Liquidity and Financial Institutions in the Postwar Period*, p. 5.

[54] See the references in footnotes 46 and 47.

[55] See, for example, the testimony of L. H. Keyserling in *Employment, Growth, and Price Levels*, Hearings before the Joint Economic Committee, Part 1 (86th Congress, 1st Session, 1959), pp. 102–103.

[56] Cf. *Staff Report on Employment, Growth, and Price Levels, op. cit.*, pp. 402–403, 408.

5. It has also been argued that the Federal Reserve's tight money policy during boom periods has tended to discriminate against certain types of borrowers—small business, state and local governments, and those seeking to purchase homes. It is widely believed that general credit tightness discriminates against these groups, which other types of government policy seek to favor. That there actually has been discrimination against small business has not yet been convincingly documented. Whatever the facts, this criticism offers a good illustration of the kind of conflict that may arise between the stabilization goal and other objectives, a matter which we discussed at some length in Chapter 18.

A FINAL COMMENT

The debate on the effectiveness of monetary policy and on how and when monetary controls should be used is not likely soon to end.[57] But few today would take the position that the Federal Reserve authorities should not be free to follow a flexible monetary policy, although some would say that on occasions during the past 10 years monetary policy had been needlessly restrictive without accomplishing its primary objective.

While a flexible countercyclical monetary policy is certainly desirable, a number of cautions must be kept in mind. Monetary policy cannot be considered apart from fiscal policy and must be synchronized with the latter, in ways that we shall discuss in the next chapter. Further, monetary policy should not be the sole reliance for curbing inflation, particularly the kind of inflation that is due to more than a simple excess of aggregate demand. Thirdly, monetary policy cannot be divorced from debt management; and, as a result, the freedom of the monetary authorities will continue to be somewhat conditioned by the need to maintain a smoothly functioning market for government securities, including the steady flow of new securities that the Treasury must issue to replace those that mature.

Finally, as much of the preceding discussion suggests and as we have reiterated on a number of occasions in earlier chapters, total spending can change because of changes in velocity as well as in

[57] For a good example of this debate, see "Controversial Issues in Recent Monetary Policy: A Symposium," *Review of Economics and Statistics,* vol. 42, August, 1960, pp. 245–282.

the money supply. While monetary measures can clearly make some contribution to stabilization policy, they cannot by themselves stabilize velocity. Cyclical instability reflects changes in velocity more than in the quantity of money, and the upward movement of prices since World War II has been supported much more by the rise in velocity than by expansion of the money supply.[58]

[58] The implication of velocity changes for monetary policy has been stressed by a number of writers. For a recent survey that, on this score, reaches unfavorable conclusions regarding the effectiveness of conventional monetary policy, see S. W. Rousseas, "Velocity Changes and the Effectiveness of Monetary Policy, 1951–57," *Review of Economics and Statistics*, vol. 42, February, 1960, pp. 27–36.

CHAPTER 20

METHODS OF ACHIEVING GREATER STABILITY: FISCAL POLICY

SIGNIFICANCE OF FISCAL POLICY

EMPHASIS on the importance of government budget policy as a tool for achieving greater stability dates largely from the depressed years of the 1930's.[1] Several factors contributed to this: the limited effectiveness of monetary policy in stemming the tide of depression, the new emphasis placed on the concept of aggregate demand in economic analysis, and the growing size of the government budget relative to the national income.

Although there has been a strong revival of interest in monetary policy in recent years, fiscal policy continues to be given an essential role to play in programs for achieving economic stability and a satisfactory rate of growth. The sheer size of government budgets today makes this inevitable. Total government purchases of goods and services in 1959 constituted about 20 percent of the GNP. To finance these expenditures, plus transfer and interest payments and government subsidies, all levels of government in the United States collected more than 125 billions in taxes in the same year.

There is some justification for the view that the American gov-

[1] The reader will find more detailed treatment of the subject of fiscal policy in numerous volumes. See, for example, H. M. Somers, *Public Finance and National Income,* 1940; A. Smithies, "Federal Budgeting and Fiscal Policy," in American Economic Association, *A Survey of Contemporary Economics,* 1948; K. E. Poole, ed., *Fiscal Policies and the American Economy,* 1951; R. A. Musgrave, *The Theory of Public Finance,* 1959; American Economic Association, *Readings in Fiscal Policy,* 1955; and two collections of papers published by the Joint Economic Committee entitled *Federal Expenditure Policy for Economic Growth and Stability* (85th Congress, 1st session, 1957) and *Federal Tax Policy for Economic Growth and Stability* (84th Congress, 1st session, 1955).

ernment has not used fiscal policy as vigorously as it might have in the years since World War II and that, in the last decade, it has relied too heavily on monetary policy. Thus, it has been argued, if the federal government had relied less on restrictive monetary policy and more on larger budgetary surpluses during postwar booms, we might have had less inflation and more rapid growth than actually occurred.[2] While not all economists would agree with this particular view regarding the effects of American monetary and fiscal policy in the last decade or so, there would be no dissent from the proposition that, in today's world, government fiscal policy has a highly important role to play in any overall stabilization program.

QUANTITATIVE ASPECT OF FISCAL POLICY

Fiscal policy operates through the control of government expenditures and tax receipts. Government spending on new goods and services directly adds to aggregate demand and indirectly increases income further through the secondary spendings which take place (the multiplier effect). Government spending via transfer payments increases disposable income and thereby consumers' spending. Taxes operate to reduce the level of private spending, both consumption and investment. Thus fiscal policy can be used both to stabilize the private components of aggregate demand (through changes in tax revenues and in transfer payments) and to offset changes in the private components through compensatory changes in government expenditures.

For various reasons, state and local budgets do not have the flexibility needed for an effective stabilization program. In view of this, and because the federal government now collects and spends more than all state and local governments combined, we shall confine ourselves to a discussion of fiscal policy at the national level. This is not to deny, however, the importance of integrating tax and spending policies at all levels of government.

Government expenditures and receipts can be combined in various ways to achieve a given stimulating or deflationary effect on aggregate demand. Suppose that a certain amount of unemployment

[2] This position is taken strongly in Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st session, 1959), chap. 8.

exists and that fiscal policy is to be used to raise aggregate demand by the required amount. This can be done within the framework of a balanced budget by raising expenditures and taxes by the same amount. In this case, the increase in the budget would have to be relatively large—probably about as large as the change in aggregate demand desired. Or a deficit can be incurred—by raising expenditures, by reducing taxes, or by doing both together. The larger the deficit, other things being equal, the stronger will be the stimulating effect on private spending. The same reasoning applies if the aim is to reduce aggregate demand in inflationary periods.

Let us consider briefly how the government might stimulate private demand by increasing both spending and taxes equally, thus avoiding a budgetary deficit. The rise in government spending on new output would increase employment and private incomes, but the increased taxes would operate to hold down private spending. If these two influences on private spending were of equal strength, there would be no net increase in private spending, and the expansion in aggregate demand would be limited to the original increase in government spending. There would be no secondary multiplier effect. Thus, if $GNP = C + I + G$, an increase in G would raise GNP by the same amount even if the increase in taxes prevented any increase in C or I. If government expenditures were originally, say, 40 billion dollars, and if the objective were to raise the GNP by 20 billions—say, from 200 to 220—then the government's budget would have to rise by 20 billions or 50 percent in order to secure the desired 10 percent increase in aggregate demand.[3]

Though it is thus possible to influence the level of aggregate demand and employment by varying the size of a budget which is always kept balanced, fiscal policy would not be very effective as a stabilizing device if we were limited to varying the level and composition of a balanced budget. The stimulating effect of an incre-

[3] It is possible that increasing taxes and expenditures by equal amounts would have some net effect on private spending. For example, the deflationary effect of the additional tax receipts might be less than the stimulating effect of the new expenditure. In this case, a tax-financed increase in government expenditures would lead to some net increase in private spending. However, it is also possible that in some situations the deflationary effect of the higher taxes would more than offset the stimulating effect of the increase in government spending—say, because private investment was discouraged by the increase in taxes. In that case, the increase in government spending would be offset to some extent by a decline in private spending.

ment of government spending is reduced if it is offset by taxes, and the deflationary impact of a cut in spending is lessened if taxes are also reduced by the same amount. The primary impact of the government's fiscal operations comes from the size of the deficit or surplus that results, and the potential effectiveness of fiscal policy as a stabilizer lies chiefly in the government's ability to vary the difference between its expenditures and receipts.[4]

QUALITATIVE ASPECT

While we shall be concerned primarily with the *quantitative* aspect of fiscal policy—the amount of government expenditures and receipts and particularly the relation between these two magnitudes—we should not forget that the *qualitative* aspect of the government's fiscal program is also important. A given amount of revenue can be raised in many different ways—that is, by levying different kinds of taxes—and government expenditures can be made for a variety of purposes. Thus, a tax on corporate profits and a sales tax may both bring in the same amount of revenue, but the former may reduce private consumption less and private investment more than does the latter. A million dollars spent on public works of the conventional type in a depressed area is not likely to cause an offsetting decline in private investment; but the same amount spent in ways that compete with private industry or that raise costs to private producers may cause some offsetting decline in private spending.

We shall have to forgo any detailed discussion of the qualitative aspect of fiscal policy in this chapter. However, we have already had something to say on the subject in Chapter 19, where we considered the possibility of influencing the amount of consumers' and business spending through the use of particular kinds of taxes. On the whole, European countries have emphasized the selective (or qualitative) aspects of fiscal policy more than has the United States.

[4] Cf. R. Musgrave, "Alternative Budget Policies for Full Employment," *American Economic Review,* vol. 35, June, 1945, pp. 387–400, reprinted in *Readings in Fiscal Policy,* pp. 291–306. For further discussion of the stimulating effects of a balanced budget, see W. A. Salant, "Taxes, Income Determination, and the Balanced Budget Theorem," *Review of Economics and Statistics,* vol. 39, May, 1957, pp. 152–161. This article contains useful references to the earlier literature on this subject.

VARIANTS OF THE GOVERNMENT BUDGET

Business firms have more than one kind of budget, and the same is true of governments. First, we must distinguish between the administrative or bookkeeping budget, on the one hand, and the cash consolidated budget, on the other. In the fiscal year ending June 30, 1956, for example, the administrative budget of the federal government showed a small surplus of 1.7 billion dollars. Yet in the same year the government's cash receipts from the public exceeded its cash payments to the public by 4.5 billions. The cash surplus was more than twice as large as the bookkeeping surplus. The administrative budget records the income and expenses involved in the general operations of the government, whether cash payments to or from the public are involved or not. Thus, interest accrued on savings bonds is an expenditure in the administrative budget, though it does not involve a cash payment in that year. The chief discrepancy between the two types of budget results from the operations of the various trust accounts administered by the federal government, which are excluded from the administrative budget. (These are chiefly the social security trust funds, the retirement and disability trust funds for government employees, and the highway trust fund.) Thus, in the fiscal year 1956, when the consolidated cash surplus was 4.5 billions in contrast to a surplus in the administrative budget of only 1.7 billion, the trust funds received 2.3 billions more than they paid out.[5]

In the last few years, social security expenditures have about balanced receipts, so that there has not been much difference on this score between the cash and administrative budgets. For the fiscal year ending June 30, 1961, the government expected the administrative budget to show a surplus of 4.2 billions, or somehwat less than the estimated cash surplus of 5.9 billions. The difference was even less the year before.[6]

From the point of view of analyzing the effects of government operations on aggregate demand, there is at least one respect in which adjustment has to be made in the cash budget. This has to

[5] These figures are taken from Department of Commerce, *U.S. Income and Output*, 1958, pp. 178–179.

[6] Cf. the *Budget Message of the President and Summary Budget Statements*, for the fiscal year 1961, p. 899. These expectations of a large surplus were not fulfilled.

do with the matter of timing. Thus, taxes are recorded in the cash budget only when received, although business firms accrue their tax liabilities and charge them to expense even before they are paid. Another needed adjustment is to eliminate, from both receipts and expenditures, purely capital transactions such as purchase of land and existing assets. When these and certain other adjustments are made, we get *government receipts and expenditures on national income account.* This is the form in which the government budget enters into the national income accounts. Our discussion of budget policy in the following pages should be interpreted as applying to this last, national-income version of the government's budget, although no serious error is involved in considering the government's fiscal operations in terms of the cash consolidated budget.[7] The important thing is to take account of the total flow of income-generating funds between all branches of the government (including trust funds), on the one hand, and the business and consuming public, on the other.

LONG-RUN BUDGET PLANNING FOR STABILITY AND GROWTH

The government's budget can be planned on a long-run as well as on a short-run cyclical basis. Long-run fiscal policy is concerned with the long-run levels of (or trends in) expenditures and receipts and with the *average* surplus or deficit that results over a number of years. Here we are concerned with the long-run and growth aspects of fiscal policy. Within the framework of a long-range plan, the budget can be made to vary cyclically in order to moderate short-run fluctuations in business activity.[8]

For maximum effectiveness, fiscal policy should be planned on both a long-run and a short-run basis. The government should plan well ahead the amount and kinds of expenditures required to meet social objectives having a high priority—national defense, the normal peacetime functions of government, social welfare measures,

[7] For further discussion of these somewhat technical problems, see *U.S. Income and Output,* pp. 56–57 and 178–179, and J. P. Lewis, *Business Conditions Analysis,* 1959, pp. 398–403.

[8] For a more theoretical discussion of the relations between fiscal policy and growth, see Musgrave, *The Theory of Public Finance,* chap. 20. For suggestions as to how American fiscal policy might make a stronger contribution to growth, see Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels,* pp. 268–274.

and so on. The test should be the community's willingness to be taxed to have these functions performed by government.

Given a long-range program of essential government spending and a tentative tax program that will cover these expenditures, the question then becomes whether the expected trend in the rate of private spending (investment plus consumption) will tend on the average to keep the economy close to the level of full employment and provide a satisfactory rate of growth without inflation. It is here that forecasts, as unreliable as they may be, are necessary. Three possibilities must be considered: Aggregate demand may, on the average, tend to rise so rapidly as to exceed the full-employment level of output at present prices. Or there may be a long-run deficiency in aggregate demand so that a considerable amount of unemployment will exist, on the average; that is, our forecast might be that aggregate demand will rise too slowly to absorb the projected expansion in the labor force. Or, finally, the forecasts may suggest that the estimated long-run trend in private demand and essential government expenditures will just about balance the projected full-employment level of output with no net tendency toward either inflation or deflation.

Let us assume that these calculations lead to the conclusion that there will be a secular tendency toward inflation, induced by excess demand, during the next decade or so. In this case, long-run fiscal policy should aim at reducing the average level of aggregate demand over the period being considered.[9] This can be done by reducing the planned level of "essential" government expenditures, by increasing tax rates, or by combining these two methods. The aim would be to have the average level of tax revenues exceed the proposed level of government spending by an amount large enough to eliminate the "inflationary gap."[10] On the assumption that long-run inflationary tendencies are at work, the budget would show a surplus in most years, though a deficit might be in-

[9] It may also be possible to use fiscal policy to stimulate the growth of aggregate supply—for example, by altering tax rates or shifting expenditures in such a way as to accelerate technological change and otherwise increase labor productivity. Since we are dealing here with a problem of excess demand, the measures taken must be such as to increase aggregate supply more than aggregate demand.

[10] The "inflationary gap" may be defined as the amount by which aggregate demand would exceed the value of output (at a given level of prices) that would be produced at full employment.

curred in periods of cyclical decline. If the underlying inflationary tendencies were fairly strong and tax rates were set high enough to offset these tendencies, the budget would show a net surplus over the period as a whole, and this would permit a corresponding decline in the public debt.

If the prediction is that aggregate demand will tend to be deficient, the proposed level of government spending can be raised or tax rates lowered. Averaging the good years with the bad, the budget would show a net deficit over the period as a whole, with a corresponding rise in the public debt. If the deficiency in aggregate demand were substantial, a deficit might be necessary even in years of peak private demand in order to keep the economy close to full employment; in years of reduced private demand, the deficit would be even larger.

Suppose the long-run prediction is that there will be no tendency toward either an inflationary or a deflationary gap—that, on the average, aggregate demand will expand at about the desired rate, being somewhat excessive during boom years and somewhat deficient when private investment experiences a cyclical decline. In this case, the long-run objective would be a budget that is balanced *on the average*. If the budget is made to be responsive to cyclical changes, there will be a surplus in boom years, when private demand is too high, and a deficit in depressed years, when the level of private spending falls below the desired level.

If the economy is growing, a budget that is balanced on the average will mean that both government expenditures and government receipts are tending to expand at the same rate. A growing economy inevitably requires increasing government expenditures for a variety of essential purposes, and a steadily rising national income automatically yields higher tax revenues if tax rates remain unchanged. We can assume, with today's heavy reliance on income taxes, that tax revenues will rise at least as fast as the national income, and probably faster. It is possible that desired public expenditures will grow more slowly, depending on the country's feeling of need for public services and its willingness to pay for them. Thus, in a growing economy, it *may* be possible to maintain a balanced budget, have some expansion in public services, and still have some decline in tax rates.[11]

[11] In the United States, this is a more realistic prospect for the federal than for the state and local governments, assuming that defense spending does not

As pointed out earlier, fiscal policy can also operate through varying the size of a balanced budget and by changing the kinds of taxes and expenditures. We have already commented on the limitations of the balanced-budget approach. The *qualitative* aspects of long-run fiscal policy, however, should not be neglected. By changing the kinds of taxes imposed and by some changes in the direction of government expenditures, it may be possible to influence the long-run level of consumption or private investment in the direction desired and thus reduce the burden on quantitative fiscal policy. It may also be possible to influence the growth of productive capacity—for example, by stimulating scientific research and technological development. Some of the possibilities here were discussed in Chapter 19, where we dealt with the means of directly influencing consumption and investment.

This is a good point at which to raise a troublesome problem, to which we gave some attention in Chapter 18. This is the possible conflict of objectives which may arise, especially when quantitative fiscal policy is used as the primary instrument to achieve long-run stability. Thus, if there is a clear tendency toward secular deflation, a long-run policy of budget deficits may conflict with the objective of reducing, or at least preventing any further increase in, the public debt. In this case, the majority of economists today would argue that the stability objective should take precedence. If the long-run tendency is toward inflation, the stability objective may conflict with the desire for a higher level of social-welfare expenditures. To have both stability and the increased government spending may require a rise in tax rates that the community does not want to impose on itself.[12]

Conflicts of this sort were evident in American policy in the 1950's. The growth rate was less than many thought desirable; yet inflation was a serious problem not only during the Korean period but also during the latter half of the decade. Monetary policy was heavily relied on to restrain inflation, but in several years the federal budget showed substantial deficits. It has been argued that monetary policy

rise faster than the national income. Particularly when population growth is rapid, severe pressure is put on state and local governments to increase their spending—for schools, roads and streets, water resources, prevention of air and water pollution, etc. Yet their tax revenues are less sensitive to changes in national income than are those of the federal government.

[12] If higher taxes are not imposed to create a surplus and the inflationary pressures persist, direct controls on wages, prices, consumption, and investment can be used. In this case, another conflict arises—this time with the objective of maximum personal freedom.

should have been easier in order to stimulate growth, and that taxes should have been higher, particularly those taxes that would have had the most direct impact on consumers' expenditures. The increased government revenues could have been used to reduce the public debt, thus making funds available for private investment; or the government could have used the money for types of expenditures that might have increased productive capacity—for example, on education or on research and development.[13]

It is not easy to say whether such a policy would have been wise. A major difficulty was that, in the latter half of the 1950's, the rise in prices was due not to an overall excess of aggregate demand. There were strong "cost-push" forces at work, reinforced by excess demand in particular sectors of the economy. Higher tax rates would not necessarily have held back the rise in wage rates and might have slowed down the rise in output by retarding the rise in aggregate demand (unless the government had increased its expenditures to match the rise in tax receipts). It is an unfortunate fact that neither of our two main instruments of stabilization policy—neither fiscal nor monetary policy—is very well suited to deal with chronic inflation associated with strong trade unions and pricing practices in oligopolistic industries.[14]

Granted all the difficulties inherent in long-run fiscal planning for stability and growth—the inevitable errors in forecasting, ignorance as to the possible repercussions of different types of budgetary operations, potential conflicts in objectives, and so on—the attempt should nonetheless be made. Otherwise the government is likely to drift into a variety of policies—regarding the public debt, national defense, social welfare, etc.—which not only may conflict with one another but may also be inconsistent with the goals of full employment, rapid growth, and price stability.

COMPENSATORY FISCAL POLICY DURING THE BUSINESS CYCLE

Given the framework of a long-range budgetary policy, we can then proceed to secure the desired short-run, cyclical flexibility in the government's fiscal operations. Taxes and expenditures can be planned so that the year-to-year changes in the budgetary surplus or deficit will offset to some extent the cyclical instability in private

[13] Cf. *Staff Report on Employment, Growth, and Price Levels,* chaps. 8–9.

[14] We shall return to this problem in the next chapter.

spending, while over the average of several cycles the desired long-run results will be achieved.

There is widespread agreement among economists that we "cannot balance the budget annually and at the same time have a stabilizing fiscal policy."[15] Attempts to keep the budget balanced every year, during booms as well as depressions, would have an active destabilizing effect on business activity. During booms, when tax revenues automatically increase and a budgetary surplus tends to develop, keeping the budget balanced would call for lower tax rates or increased government spending. Either would accentuate the boom already under way. During business downswings, the balanced-budget principle would require that tax rates be raised and government expenditures reduced. Obviously, such a policy would only add to the deflationary forces already at work.

The kind of quantitative fiscal policy needed to counteract short-run or cyclical instability is clear. Tax revenues must rise relative to government expenditures during boom periods, and tax receipts should decline relative to outlays during periods of declining employment. This means, essentially, running a budgetary deficit during periods of low and declining activity, and reducing the deficit and perhaps generating a surplus during periods of rising and high-level activity. Over a succession of business cycles, tax revenues might, on the average, exceed or fall short of expenditures, in accordance with the government's long-range fiscal plan and the accuracy with which the necessary long-run predictions had been made. In other words, a compensatory fiscal policy aimed at moderating short-run instability can be consistent with any one of a variety of long-run fiscal objectives.

HOW TO ACHIEVE FLEXIBILITY

There are several ways of planning the desired flexibility in the relation between tax revenues and expenditures.[16] Basically, the following techniques are available: built-in or automatic flexibility, formula flexibility, and discretionary action. With respect to each of these, there is the further question as to the relative emphasis that

[15] Committee of the American Economic Association, "The Problem of Economic Instability," *American Economic Review,* vol. 40, September, 1950, p. 521. Cf. "Federal Expenditure and Revenue Policy for Economic Stability," *American Economic Review,* vol. 39, December, 1949, p. 1264. Both reports have been reprinted in *Readings in Fiscal Policy.*

[16] Cf. Musgrave, *The Theory of Public Finance,* chap. 21.

should be placed on the expenditure and the revenue sides of the budget; that is, should the desired offsetting effect be obtained by varying tax revenues or government outlays?

A compensatory fiscal policy that is entirely automatic (i.e., built-in budget flexibility) has many attractions. We have already noted the extent to which this sort of automatic flexibility has been built into the federal budget. With the present heavy reliance on income tax, tax revenues are very sensitive to changes in business conditions, even if tax *rates* remain unchanged. Social security contributions by employers and employees also vary in the desired manner. On the expenditure side, insurance payments to the unemployed, agricultural subsidies, and relief payments vary inversely with the business cycle. Thus, without any need for discretionary action by either Congress or the administration, expenditures rise relative to tax receipts during business downswings and fall relative to tax receipts during upswings. To this extent, automatic stabilizers have been built into the budget.

Some economists and businessmen take the position that a stabilizing fiscal policy should depend entirely on built-in flexibility, and that no attempt should be made to go further and change either tax rates or expenditure programs for the sole purpose of offsetting fluctuations in private spending. A leading proponent of this position is the Committee for Economic Development, an organization of businessmen that has done much constructive work in the field of economic policy.[17] A similar position is taken by a minority of economists.[18]

The advantages of a purely automatic program are fairly obvious.

[17] For a description and critical evaluation of CED's emphasis on automatic budget flexibility, see W. W. Heller, "CED's Stabilizing Budget Policy After Ten Years," *American Economic Review,* vol. 47, September, 1957, pp. 634–651. The CED program does not rule out discretionary changes in the event of a severe recession. Early in 1958, the CED recommended that personal income taxes be temporarily reduced by 20 percent if the business downswing then in progress pushed seasonally adjusted unemployment above five million. This was, in fact, just about the level reached at the low point of the recession. This recommendation by CED in 1958 suggests that it is more in favor of discretionary contracyclical tax policy than its earlier statements might have suggested. See *Anti-Recession Policy for 1958,* by the Program Committee of the Committee for Economic Development. For a general defense and elaboration of CED's recommendations regarding cyclical budget flexibility, see its *Problems in Anti-Recession Policy,* 1954.

[18] Cf. Milton Friedman, "A Monetary and Fiscal Framework for Economic Stability," *American Economic Review,* vol. 38, June, 1948, pp. 245–264.

It removes the need for making forecasts, which may be wrong, before action can be taken. Secondly, automatic stabilizers operate promptly; there is little of the administrative delay which inevitably accompanies discretionary changes in tax rates and expenditure programs. Further, such a program, by keeping tax rates unchanged and avoiding large and frequent changes in government spending programs, creates an environment favorable to private investment. Reliance on built-in flexibility is also the simplest way of integrating short-run and long-run fiscal policy.

Most economists favor the greatest possible use of automatic stabilizers; but it is also true that they "do not believe it prudent for policy to regard automatic flexibility as more than a first line of defense; more must be done to cope with serious economic fluctuations."[19] This means deliberately changing tax rates or expenditures, either according to a predetermined plan (formula flexibility) or on a completely discretionary basis.

"Formula flexibility" is a relatively new suggestion which has not yet been tried to any significant extent. Congress might enact rules under which income-tax rates would rise by stated amounts following some given rise in prices and would decline by stated amounts following a given fall in production or employment. Similarly, in accordance with rules laid down by previous legislation, unemployment compensation and other types of transfer payments could be made more liberal as the volume of unemployment increased. Or advance authorization could be given for varying the volume of public-works expenditures in accordance with changes in selected business indices. In these ways, the government's receipts and expenditures could be made to vary more widely over the cycle than is possible under the strict built-in flexibility principle. At the same time, since the changes would go into effect automatically in accordance with formulas laid down in advance, the danger of forecasting errors and administrative delays would be reduced to a minimum.[20]

Formula flexibility shades gradually into discretionary action. We

[19] "Federal Expenditure and Revenue Policy for Economic Stability," *American Economic Review,* vol. 39, December, 1949, p. 1268.

[20] Cf. Committee of the American Economic Association, *op. cit.,* pp. 524–525; "Federal Expenditure and Revenue Policy for Economic Stability," *American Economic Review,* vol. 39, December, 1949, p. 1268; E. E. Hagen, "The Problem of Timing Fiscal Policy," *American Economic Review,* vol. 38, May, 1948, suppt., pp. 417–429; and E. E. Hagen, in *Federal Tax Policy for Economic Growth and Stability, op. cit.,* pp. 62–66.

may illustrate this by considering the possibility of varying tax rates over the cycle. Congress might instruct the Treasury to change tax rates in a certain way in accordance with the movement of given business indices. This is formula flexibility with no discretionary leeway. Discretion would enter if Congress reserved to itself the right to veto each change before it went into effect, or if it planned the tax changes in advance but provided that they would go into effect only upon joint resolution of Congress when the latter believed that economic conditions warranted the change. As another possibility, Congress might authorize changes in tax rates within certain limits and leave it to the President's discretion as to when the changes should be put into effect. Congress could reserve to itself the right to veto the President's decisions. Finally, we have complete Congressional discretion with no advance planning, which is the present situation in the United States and most other countries.

From a purely economic point of view there is much to be said for some type of formula flexibility, or at least for giving to the executive branch wider discretionary authority than it now has to initiate changes in the timing or extent of the government's fiscal operations. Under the system of complete legislative discretion that now prevails, there is little advance planning and a maximum of delay in utilizing fiscal policy as a stabilizing device. Obviously, serious political and administrative problems arise in any attempt to broaden the discretionary authority of the Administration in the field of fiscal policy. The difficulties would be reduced if this discretion were narrowly circumscribed in accordance with a formula laid down by Congress in advance. Whatever the political difficulties, we need some system of achieving greater flexibility in the budget than can be provided by the automatic stabilizers that now exist. This means changes in tax rates and expenditure programs to offset fluctuations in private spending. In the following paragraphs, we shall consider what can be done in this direction without further reference to the methods that may be used to plan and put the changes into effect. To repeat, tax rates and expenditure programs can be altered in accordance with a predetermined formula or on the basis of purely discretionary action.

COMPENSATORY SPENDING AND TAX POLICY

The question now before us is this: If the relation between expenditures and receipts is to be deliberately changed in order to

counteract fluctuations in private spending, is it better to rely on changes in tax rates or changes in the volume of expenditures? If an inflationary rise in prices is developing, should the government concentrate on raising tax rates or on reducing the amount of its expenditures on transfer payments and on goods and services? If a depression threatens, should tax rates be lowered or expenditures increased, or should both be tried?

Economic thinking on this matter has shifted somewhat in recent years. At first, during the 1930's, emphasis was placed on the possibility of moderating cyclical fluctuations by varying the amount of government spending, which would be increased during depressions and reduced during booms. In particular, emphasis was placed on the stimulating effect of increased outlays on public works during depressions. More recently, attention has shifted to the results that can be achieved by changing tax rates.

Several issues are involved here. Whether, in attempting to moderate cyclical fluctuations, it is better to change tax rates or the amount of government expenditures depends particularly on (1) the multiplier effects of the proposed tax and expenditure changes, (2) questions of timing and flexibility, and (3) possible indirect repercussions on private spending.

Pure multiplier analysis suggests that a given increase in public expenditures on goods and services ought to have a larger impact on the GNP than a decline in tax revenues of the same dollar amount. The amount of the difference would depend on, among other factors, the kinds of taxes involved. This can be illustrated by the following simple example.[21]

Suppose the government proposes to increase its expenditures on new goods and services by a billion dollars in order to raise the level of employment. Tax rates are to remain unchanged. Assume that the marginal propensity to consume GNP is six-tenths. Then the increase in GNP from the additional government expenditures would be

$$\Delta \text{GNP} = \frac{1}{1 - .6} = 2.5 \text{ billion}$$ [22]

[21] For more detailed discussion, see Musgrave, *The Theory of Public Finance,* chap. 18; also Salant, *op. cit.*

[22] We could also write (in billions)

$$\Delta \text{GNP} = 1 + .6 + .6^2 + .6^3 \ldots = 2.5.$$

It should be noted that the net deficit at the new level of GNP will be less than the increase in government spending, since the increase in GNP will bring in some additional tax revenues.

Now assume that the billion-dollar stimulus is achieved through an appropriate reduction in the rate of the personal income tax. The initial effect is to increase disposable incomes. Only after this happens will there by any increase in spending and further multiplier effects on the GNP. If we assume that the marginal propensity to consume disposable income is eight-tenths, while the marginal propensity to consume GNP is six-tenths, the effect of the billion-dollar tax cut will be:

$$\Delta GNP = .8\left(\frac{1}{1 - .6}\right) = 2 \text{ billions}^{23}$$

While disposable incomes rise initially by one billion, 200 million of this is saved and only 800 millions are spent and are thus available to generate a multiplier effect.[24]

While theoretical multiplier analysis suggests that compensatory fiscal policy should emphasize changes in government spending, considerations of timing and flexibility point in the opposite direction. A tax cut, for example, can go into effect promptly, whereas it ordinarily takes considerable time to plan and carry through changes in expenditures. Also, a tax change can be more quickly reversed.[25] One weakness in trying to combat a short recession by increasing public expenditures, for example, is that the chief stimulating effect of such expenditures may be felt only after the next upswing is well under way. This is particularly true of the heavier types of public works. In addition, it must be remembered that government expenditures

[23] This can be written (in billions)

$$\begin{aligned}\Delta GNP &= .8 + (.8)(.6) + (.8)(.6)^2 + (.8)(.6)^3 \ldots \\ &= .8(1 + .6 + .6^2 + .6^3 \ldots) = .8(2.5) = 2\end{aligned}$$

Actually, a reduction in tax rates implies that the marginal propensity to consume GNP will be higher than before the tax cut, thus increasing the total multiplier effect of the tax-rate reduction. Allowing for this, and further allowing for the fact, noted in the preceding footnote, that increased government expenditures bring in some increase in tax revenues, it can be shown that, per dollar of deficit incurred, government expenditure still has a larger multiplier effect on income than does a tax reduction. Cf. Salant, *op. cit.*, p. 158.

[24] An increase in transfer payments works in the same way as a reduction in taxes. The first effect is to increase disposable incomes, and something less than the full amount of the increase will then be spent on goods and services.

[25] One qualification is necessary here. It might be politically difficult to raise taxes again in a boom after they had been reduced during the preceding depression. This difficulty can be met in part by specifying a time limit on the tax reduction, so that taxes will automatically increase again unless positive legislative action is taken to extend the reduction.

result in the provision of public services that are desired entirely apart from their effects on economic stability. If the community wants these services and is prepared to pay for them in the long run, then essential government expenditures should not be reduced simply to cope with inflation or increased merely to offset a decline in private spending. On this count, the argument strongly favors using taxes rather than government spending to achieve a countercyclical fiscal policy.

So far as indirect effects on private spending are concerned, the case is not clear-cut, but on balance probably favors tax rather than expenditure policy. Tax increases to curb a boom will tend to have some unfavorable effect on profit expectations, and thus will tend to hold back private investment; and tax reductions should have the opposite effect in depressions. The actual impact on private spending will depend on the kinds of taxes that are changed and on the general state of expectations.[26] Changes in government spending can also influence business expectations in the desired direction. Here again, the net result depends a great deal on the kinds of spending that are to be increased or reduced.

COMPENSATORY SPENDING

Let us now consider in more detail the possibility of varying government expenditures during the cycle—over and above whatever automatic flexibility exists in the spending side of the budget. It is now generally agreed that the possibilities here are, on the whole, rather limited. There is not much room in which to make year-to-year changes in the amount of spending on essential public services—on national defense, interest on the public debt, the basic social services, and so on. Some of these expenditures are completely inflexible, and most people would agree that it would be poor public policy to cause the other outlays in this category to change very much from year to year, even if it were possible to do so.

Although there is not much room for cyclical variation in the ordinary operating expenses of government, two other types of expenditures can be made to vary cyclically. These are transfer payments and outlays on public works. During depressions the volume of trans-

[26] As we noted earlier, frequent or radical changes in the *structure* of taxes may have undesirable effects on business expectations, but this objection does not apply to moderate cyclical changes in the general level of taxes.

fer payments can be deliberately increased—through larger relief payments, through more liberal benefits to veterans, the aged, and so on, or by subsidies to other groups. The result is to increase the disposable income and consumption of the persons thus benefited. Since the marginal propensity to consume among these groups would be high, the initial increase in private spending would be nearly as large as the rise in transfer payments, and private spending would be stimulated still further through the operation of the multiplier.

While the volume of transfer payments should be varied as part of a compensatory fiscal policy, the opportunities here are probably limited. Social security programs should be planned on a long-run basis, and the provisions of these programs should not change drastically with the ups and downs of some business index.[27] The chief opportunity for deliberately varying the amount of transfer payments is in the field of consumers' subsidies—to provide for the needy who are not covered by unemployment insurance and to supplement the benefits being received by those who are covered. As we mentioned in Chapter 19, a useful type of program in depressions is a plan under which specified groups are permitted to secure at no cost or at reduced prices foods and other products that are in excess supply.

In contrast to transfer payments, public-works expenditures have the virtue that they give rise to useful assets, the services from which add to the real income of the community. Though public-works programs should be planned on a long-run basis and in accordance with the community's willingness to pay for them, the actual construction of these public works can to some extent be accelerated or delayed in order to have a stabilizing effect on the business cycle. Twenty-five years ago, many economists felt that wide countercyclical changes in public-works expenditures could do much to stabilize the level of business activity. More recently, however, the opinion has been growing that public-works planning by itself cannot go very far in stabilizing the level of output and employment. This is particularly true with respect to "heavy" public works—dams, public buildings, large-scale river and harbor developments, public housing, and so on. Even if such projects are planned well in advance, in some cases it may take a year or more to secure the necessary appropriations, to

[27] Unemployment insurance benefits might be made to have greater cyclical flexibility. In particular, the period for which unemployed workers could secure compensation might be lengthened as the volume of unemployment increased.

let the contracts, and to get construction actually going. Two or more years may elapse between the initial decision to act and the completion of the project. Thus, it may be impossible to increase expenditures on large-scale public works rapidly enough to prevent a serious decline from developing. And once the projects are under way, it may not always be possible to reduce expenditures promptly when a new business upswing gets started.[28]

Obviously, what can be done to improve the timing of public-works expenditures should be done. As a minimum, government agencies should avoid planning public works so as to accentuate short-run instability. This has all too often been the case, particularly among state and local governments. It is during boom periods that these governmental bodies can most readily obtain funds, so that their expenditures on public works tend to be high when business is already prosperous and low when private spending is also at a low level.

In addition to what can be done in the field of "heavy" public works, "light" public works offer an opportunity for some cyclical variation in government expenditures, especially if a "shelf" of such projects is kept ready for use when needed. "Light" public works include road building and maintenance, some types of soil conservation and flood control, airport improvements, and other similar projects which can be started promptly and carried through to relatively quick completion.

Different types of expenditures have different indirect effects on private investment. In general, public-works expenditures are likely to be more stimulating than an equal amount of transfer payments. It is true that both raise personal incomes and consumption and set the multiplier process in operation.[29] In addition, however, public-works expenditures stimulate the construction industries, which fluctuate widely over the cycle. In depression, increased spending on

[28] See the interesting article by S. J. Maisel, "Timing and Flexibility of a Public Works Program," *Review of Economics and Statistics,* vol. 31, May, 1949, pp. 147–152; also his paper in *Federal Expenditure Policy for Growth and Stability, op. cit.,* pp. 382–397. For a more optimistic view regarding the use of public works as a stabilizing device, see M. L. Colean and Robinson Newcomb, *Stabilizing Construction: The Record and Potential,* 1952, app. Z.

[29] Transfer payments are in effect negative taxes. They first increase disposable income, and some leakage results before there is an increase in spending on new goods and services. Thus the multiplier for transfer payments, like the tax multiplier, is less than the expenditure multiplier.

public works means substantially larger orders for building materials and construction equipment; these may induce firms to replace worn-out equipment and to accumulate inventories; and so on. An increase in transfer payments, even though completely spent by the initial recipients, is likely to mean only a relatively small increase in total consumption; this may be met by a reduction in inventories by retailers and manufacturers; and in any event, it is not likely to provide a strong stimulus to investment expenditures by firms in the consumers'-goods industries. The same argument suggests that a given deficit will be more stimulating in its indirect effects if it takes the form of increased expenditures than if it takes the form of reduced taxes.[30]

Another point to be made here is that the "leverage" effect is not the same for all types of public-works expenditures.[31] Indeed, some types of spending may cause a contraction in private investment—for example, if the government begins to build facilities in an industry to compete with existing private concerns, or if the expenditures are of a sort that lead to an increase in costs for private firms. Thus, increased government expenditures in depression should be in fields normally reserved for government activity, in geographical areas where unemployment is substantial, and for projects that do not require materials or types of labor that are in inelastic supply.

COMPENSATORY CHANGES IN TAX RATES

Our conclusion thus far is that there is some room in which to adjust the volume of government expenditures in order to offset cyclical variations in private demand, and governments would be well advised to follow such a compensatory spending policy to the extent that it is possible. However, if the forces making for instability continue to be as strong in the future as in the past, some sort of compensatory tax policy will be necessary also. Such a policy involves

[30] Public works, however, do not necessarily require the same materials used in private construction. An expanded program of public works may offer very little stimulation to some building-materials industries and may perhaps overstimulate others. Cf. Julius Margolis, "Public Works and Economic Stability," *Journal of Political Economy,* vol. 57, August, 1949, pp. 293–303. See also A. M. Strout, "Primary Employment Effects of Alternative Spending Programs," *Review of Economics and Statistics,* vol. 40, November, 1958, pp. 319–328.

[31] Hansen uses the term *leverage* to include the total effect on both consumption and private investment of a given change in public expenditures. *Fiscal Policy and Business Cycles,* pp. 264–265.

raising tax rates when inflationary pressures are strong and reducing them when deflationary forces are at work.

If a compensatory tax policy is to be followed, the government must decide which kinds of taxes are to be changed, and by what amounts, in response to changes in private spending. Where, as in the United States, primary reliance is put on the income tax, the government would adjust the level of income-tax rates upward and downward in response to given changes in selected business indicators. With the income tax on a pay-as-you-go basis, changes in tax rates would be reflected promptly in the disposable income of individuals. It has also been suggested that the change might be limited to the first bracket (the "basic rate") of the income tax.[32] In the United States, this bracket accounts for half or more of the revenues collected under the personal income tax. This is the only rate paid by the great majority of taxpayers, whose marginal propensity to consume is higher than that of wealthier taxpayers subject to the higher brackets. Hence, a dollar's change in tax revenues caused by varying the basic tax rate will have a larger impact on consumption than a dollar's variation in tax receipts obtained through altering the higher tax rates. This proposal also has the advantage that it does not involve frequent and large changes in tax rates on business incomes, which might have an unsettling effect on business confidence and tend to reduce the long-run level of private investment.

A program of compensatory changes in tax rates must be planned in advance if it is to be effective. Some type of formula flexibility, along the lines previously discussed, is needed. If the planning is not done in advance, it will be impossible to put the changes into effect promptly enough to have the desired effect.

In addition to this type of quantitative tax policy, the government can change particular types of business taxes in order to stimulate or hold back the volume of private investment. We discussed the use of such tax incentives in Chapter 19. The purpose of qualitative tax policy, such as allowing investment made during depressions to be deducted from taxable income, is to induce a change in private spending even though the volume of tax revenues does not change very much. The aim is to influence the *propensity* to spend. In the case of quantitative tax policy, the aim is to influence the *ability* to

[32] Cf. C.E.D., *Jobs and Markets*, 1946, pp. 74–75; A. H. Hansen, *Economic Policy and Full Employment*, 1947, p. 141.

spend by varying the fraction of their incomes that taxpayers must turn over to the government.[33]

SUMMARY

We may summarize our conclusions as to compensatory fiscal policy as follows: Taxes should rise relative to expenditures during business upswings and fall relative to expenditures during downswings. To the extent possible, this effect should be achieved through built-in flexibility. Since such automatic flexibility is not likely to be enough, tax rates and expenditure programs need to vary with the business cycle. Some, though limited, flexibility in outlays can be achieved through planned variations in transfer payments and through the proper timing of public-works expenditures, especially of the lighter sort. A full-fledged compensatory policy will also need to include some changes in tax rates—both in the basic income-tax rate in order to affect consumers' spending and in various "incentive taxes" that influence business decisions to invest. To be fully effective, the details of such compensatory fiscal policy must be planned so that the desired changes can go into effect promptly in accordance with a formula that is worked out in advance.[34]

American fiscal policy since World War II has relied primarily on the built-in flexibility in the federal budget. Discretionary changes in tax rates as a stabilization device have, in general, been avoided.[35] There has been some slight increase in spending, particularly

[33] For further material dealing with this range of problems, see *Federal Tax Policy for Economic Growth and Stability, op. cit.*

[34] There is an important political reason why a formula for cyclical changes in tax rates should be worked out and announced in advance. Legislators find it easy to reduce taxes but are understandably reluctant to increase them. If we rely on *ad hoc* discretionary action, tax reductions made during a business downswing may not be promptly withdrawn during the following upswing.

[35] We have had one important case since World War II in which taxes were deliberately raised to control inflationary tendencies. This was during the Korean war. Taxes were reduced in 1948 (over a presidential veto) during the late stages of the boom, but not because Congress anticipated the 1949 recession. Taxes were cut at the beginning of 1954, after a recession had been under way for several months, but this reduction was automatic in response to Congressional action taken in 1951. For a detailed and critical review of federal fiscal policy during the first three postwar business cycles, see *Staff Report on Employment, Growth, and Price Levels,* pp. 215–255. The use of fiscal policy by various European countries during the same period is described in A. Maddison, "The Postwar Business Cycle in Western Europe and the Role of Government Policy," *Banca Nazionale del Lavoro Quarterly Review,* June, 1960.

through a speeding up of expenditures already authorized, during recessions. But, on the whole, the federal government has not put much emphasis on *discretionary* fiscal policy as a means of achieving greater economic stability. Actually, the government's fiscal operations have on occasion turned out to have had a destabilizing effect. Defense expenditures have been quite volatile, and the timing of changes in these expenditures has sometimes been such as to accentuate, rather than ameliorate, booms or recessions.

USE OF A "DUAL BUDGET"

In private business, a fundamental distinction is made between *current expense,* which is charged against current income, and *capital expenditures,* which are not. Sound financial and accounting practice does not require a business firm to meet all of its capital expenditures out of current sales receipts. This is in contrast to the usual practice of governments. The budget is said to be unbalanced —i.e., there is a budgetary deficit—if *all* expenditures are not covered by tax receipts, even if some of the outlays are for capital improvements which will have a useful life extending well into the future.

It has been frequently suggested that governments should use a double budget, both in the interest of sound accounting practice and in order to achieve a more effective stabilization program. While there are a number of variants of this proposal, the following indicates the essential idea.[36] There would be two budgets: the operating budget, which would include the ordinary operating expenses of government and current tax receipts, and the capital budget, which would include only outlays on durable capital goods (i.e., public works) . Whereas the operating budget would be kept balanced, at least over a succession of good and bad years taken together, there would be no obligation to "balance" the capital budget, which would be financed through government loans. The capital outlays would generate a series of current expenditures in future years—for maintenance, depreciation, and interest on the public debt—and these expenditures would be charged against each year's operating budget.

[36] For more detailed discussion, see Hansen, *Fiscal Policy and Business Cycles,* chap. 10, and Musgrave, *The Theory of Public Finance,* pp. 558 ff. The latter contains a number of useful references to the literature on capital budgets.

It is easy to see how such a dual or capital budget plan could be adapted to the needs of a stabilization program. With built-in flexibility, the operating budget would show a deficit in depressions and a surplus in prosperity. If tax rates were set high enough, the operating budget would be balanced over the cycle as a whole. Tax receipts, on the average, would be sufficient to cover all operating expenses, including current charges for interest and depreciation on capital outlays already made. The capital budget would be planned on a long-run basis. All capital expenditures would have to meet this test: that the community was willing to be taxed to meet the future interest and depreciation charges that would be included in the operating budget. So far as possible, capital expenditures would be made to vary with the cycle. During boom periods, postponable outlays would be deferred; during business downswings, such expenditures would be accelerated.

A dual-budget system has been used by Denmark since the 1920's.[37] A similar plan, but with provision for cyclical flexibility, was developed by Sweden during the 1930's and attracted widespread attention. What particularly excited comment about the Swedish program was the decision not to attempt to balance the operating budget every year. This was probably the first case in which the idea of a cyclically variable budget was officially adopted by a national government.[38]

A stabilizing fiscal policy does not require the use of a double budget, nor does such a budget provide an automatic guide as to how government expenditures should be financed in order to minimize fluctuations in aggregate demand. The argument in favor of this sort of budget is primarily twofold. It provides a rational basis for the planning and financing of public works; the desirability of such projects is weighed against the community's willingness to bear the cost not in one year but over the useful life of the asset. Secondly, the dual budget facilitates the integration of the stabilization and other objectives of the government's budgetary operations. There is less

[37] Some other countries have utilized capital and other extraordinary budgets, usually in a rather unsystematic way, for even longer periods.

[38] Actually Swedish economists in the 1930's approached the use of fiscal policy as an antidepression measure with some caution, more so than they did later. Swedish stabilization policy, both prewar and postwar, is discussed extensively in Erik Lundberg, *Business Cycles and Economic Policy*, 1957. See also B. Ohlin, *The Problem of Employment Stabilization*, 1949, chap. 4.

resistance to a compensatory fiscal policy if there are reasonable guarantees that the operating budget will be kept balanced over the period of a relatively few years and that the cost of durable assets will be financed out of taxes during the lifetime of these assets.

The federal government in this country continues to use the single-budget procedure, although various sorts of capital budgets are used by some state and local governments. During the 1930's, the federal government made some use of the distinction between "ordinary" and "extraordinary" expenditures, the latter being primarily relief and recovery expenditures resulting from the government's attempts to cope with the Great Depression. It has sometimes been suggested that such special depression-born outlays be segregated and financed out of taxes not immediately but only during subsequent boom periods. This is essentially an accounting device that helps to implement a countercyclical fiscal policy.

INTERRELATIONS OF FISCAL AND MONETARY POLICY

Since fiscal policy is concerned with *money* flows into and out of the Treasury, the government's budgetary operations have a monetary side which cannot be neglected. Fiscal policy creates two types of monetary problems. There is first the question of how to finance deficits and what to do with the funds resulting from a budgetary surplus. Secondly, the mere existence of a large public debt, even if it is no longer growing, raises important monetary issues, some of which we briefly considered in Chapter 19. Hence, we must pay some attention to the interrelations between monetary policy, on the one hand, and fiscal policy and debt management, on the other.[39]

FINANCING A DEFICIT

The stimulating effect of a budgetary deficit depends in part on the way in which it is financed. There are a number of possibilities. The government may simply print paper money to finance the excess

[39] For more detailed treatment of this range of issues, cf. K. E. Poole, ed., *Fiscal Policies and the American Economy,* chaps. 2 and 4; A. H. Hansen, *Monetary Theory and Fiscal Policy,* 1949, chaps. 11–14; R. A. Musgrave, "Money, Liquidity, and the Valuation of Assets," *Money, Trade, and Economic Growth: In Honor of John Henry Williams,* 1951; American Economic Association, *Readings in Fiscal Policy,* chaps. 13–18; and W. L. Smith, *Debt Management in the United States,* Study Paper no. 19 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels* (86th Congress, 2nd session, 1960). See also the additional references cited in these sources.

of expenditures over tax receipts. We can rule out this possibility for countries having well-developed banking systems. The other alternatives all involve the sale of securities—that is, an increase in the interest-bearing public debt.[40] The government can borrow from three types of lenders: the nonbanking public (individuals, corporations, and institutional investors such as life insurance companies), the commercial banks, and the central bank. It also has the choice as to the kinds of securities it will sell—long-term vs. short-term, marketable versus nonmarketable, and so on.

The stimulating effect of the deficit is greatest if the government borrows from the central bank; it is least if securities are sold to the nonbanking public. Let us see why this is so.

When the government sells securities to the central bank, the latter creates a deposit to the Treasury's account. Spending of this deposit puts funds in the hands of the public which are deposited in the commercial banks. The banks present the Treasury's checks to the central bank, which debits the Treasury's account and credits the reserve account of the member banks. The end result, then, is that the public's demand deposits rise and *member-bank reserves expand by an equal amount.* With our present system of fractional reserves, this means that the banks have excess reserves on the basis of which they can create additional deposits, either by lending to business or by buying securities. By financing the deficit in this way, the Treasury injects a double monetary stimulus into the economy. There is an immediate expansion in the money supply, measured by the new demand deposits received by the public; and the liquidity of the banking system is increased through the creation of excess reserves. In this case, the government's borrowing operations make it easier rather than more difficult for private borrowers to secure funds.

Assume now that the Treasury sells securities to the commercial banks rather than to the central bank. Demand deposits rise by a corresponding amount and pass into the hands of the public as the government spends the money. But there is no change in bank reserves. So far as the balance sheet of the banking system is concerned, investments in government securities (on the asset side) and de-

[40] If the Treasury has accumulated excess cash balances from past surpluses, the deficit can be financed in part by using these funds. To this extent, no increase in the debt is involved. Ordinarily, however, deficits and surpluses are matched fairly promptly by corresponding changes in the public debt.

mand deposits (on the liability side) rise by the same amount. *Excess* reserves of the banking system are now lower than before; total reserves are the same but deposit liabilities are larger. Thus the banks' ability to expand their loans has, for the moment, been impaired. This method of financing the deficit is, therefore, less stimulating than direct borrowing from the central bank, although both methods result in an increase in the money supply.

The enlarged bondholdings of the banks now raise another problem. The banks may do one of three things with their government securities: hold them to maturity, sell them to the central bank, or sell them to the public. If they hold them, the situation remains as described above. However, since government securities are highly liquid and can readily be turned into cash, the banks have larger secondary reserves than before. As a result, their willingness to lend is likely to grow; and, if they do expand their loans, they may try to convert some of their securities into additional cash reserves. Thus, new problems—problems of debt management—are created for the Treasury and the central bank. The banks can sell their securities to the central bank, thereby secure new cash reserves, and then expand their loans by several times this amount. They may choose to do this not when business is slack and such action would be welcomed, but later when business is booming and credit expansion adds to the inflationary forces then at work. If the central bank refuses to purchase the securities, the banks may sell them to the nonbanking public. The public gives up demand deposits in exchange. If the public is reluctant to buy and the banks persist in selling, bond prices will fall sharply and interest rates will rise. If the break in bond prices is severe, some degree of financial disorganization may result, and the solvency of some financial institutions may be put in jeopardy.

Thus, we see that deficit financing raises not one but two major monetary problems. One is how to secure the *immediate* monetary effects that are desired. This involves the decision as to whether to bring about an expansion in demand deposits, and whether bank reserves should be increased. The other monetary problem applies to the future. What will be the *future* monetary repercussions of the present increase in the debt? Will the present buyers of the government securities be content to hold them, and how will the enlarged supply of liquid assets affect the economy's propensity to spend in the future as business conditions change?

Let us suppose now that the government finances the deficit by selling securities to nonbank investors. This results in a transfer of deposits from the public to the Treasury. When the government spends the money, these deposits come again into the possession of the public. In this case, there is no net increase in the money supply and no change in bank reserves. Of the three ways of financing the deficit, this is obviously the least inflationary.

If the government's deficit spending is to increase aggregate demand, MV must rise. We have just seen that this last way of financing the deficit does not involve an increase in M. Hence, if the GNP is to rise, there must be a change in the income velocity of money. If the new government securities are exchanged for money that otherwise would have remained idle, velocity will increase. If, however, the public reduces its own spending in order to buy the securities, there is no change in velocity, and the increase in government spending is offset by a decline in private spending. During periods of declining business activity, the odds are that the funds used to buy the government securities would have otherwise remained idle and that the required rise in velocity will therefore take place. Further, it is safe to assume that, during such periods, the government can sell large amounts of securities to the public without a significant increase in interest rates, so that the government borrowing will not make it more difficult for private borrowers to secure funds.[41] In addition, the monetary authorities can take steps to insure that interest rates remain low.

As in the case of bond sales to the banks, the monetary situation is affected by the fact that the rise in the public debt has increased the economy's supply of liquid assets, and the fact that the larger supply of liquid assets will continue into the future. This, in itself, tends to have a stimulating effect on private spending. Although the supply of money has not increased, the total stock of liquid assets has risen. Since highly marketable bonds, especially short-term securities, help to satisfy the desire for liquidity, the public's willingness

[41] If the government has to incur a deficit when the economy is at full employment, as happens during major wars, selling bonds to the nonbanking public is clearly the least inflationary way of financing the deficit. An increase in the money supply is prevented. Since, under these conditions, the funds used to buy the bonds would probably have gone into either consumption or private investment, this method of financing the deficit helps to prevent a rise in private spending.

to spend its present cash holdings is likely to increase. When business is depressed, this is a net gain. It adds to the stimulating effect of the government's deficit spending. Unfortunately, this effect will be felt most strongly after business conditions improve and a boom has got under way. It is then that the large supply of liquid assets is most likely to increase the public's willingness to spend. The increased supply of liquid assets can stimulate spending in two ways. First, firms and individuals may be willing to spend more, relative to their incomes, if a substantial fraction of their assets are in liquid form. Second, the liquid assets are a means of financing an increase in spending. Holders of government securities can convert these assets into cash by selling them to the banks (which increases the money supply) or by selling them to nonbank investors who own idle deposits. The latter procedure leads to an increase in velocity of the existing money supply. As we saw in Chapter 19, an increase in spending financed in this way is difficult for the monetary authorities to control.

Thus, large-scale deficit financing, whether the securities are sold to banks or to the nonbanking public, raises problems of debt management and monetary policy for the future. Selling securities to the nonbanking public is the least inflationary method of financing a deficit; but even this method, by increasing the public's supply of liquid assets, stores up problems of monetary control for the future. These problems are illustrated in extreme form by the character of the inflationary boom after World War II, which was supported by the tremendous increase in the supply of money and liquid assets resulting from financing wartime deficits. During the 1930's, however, the economy absorbed a large increase in the supply of money and liquid assets without generating an inflationary boom. How inflationary are the ultimate effects of deficit financing depends on how much the supply of money and liquid assets increases relative to the national income, on the government's ability to reduce the debt during boom periods, and on the underlying nonmonetary forces that determine the volume of private spending, particularly the cyclical and secular influences operating on business expectations.

HANDLING A BUDGETARY SURPLUS

As we should expect from the preceding discussion, a budgetary surplus raises monetary problems also. The deflationary impact of

the surplus depends in large part on what the Treasury does with the excess receipts. A number of alternatives are available. The Treasury may simply accumulate idle deposits at either the commercial banks or the central bank. Or, what is more likely, it may reduce the public debt by retiring securities held by the central bank, by the commercial banks, or by the nonbanking public. The greatest deflationary effect will be felt if it accumulates idle deposits at the central bank or redeems securities held by the central bank. The deflationary effect is somewhat less strong if it accumulates the idle deposits at the commercial banks or retires securities held by the latter. The deflationary impact is least if the excess tax receipts are used to retire debt held by the nonbanking public.

When the government's tax receipts exceed its expenditures, the public's deposits decline and the Treasury's increase. Suppose these deposits are transferred to the central bank and left idle there. The net result is twofold. There is less money in the public's hands, and member-bank reserves are reduced by the same amount. The effect is the same if the excess receipts are used to retire bonds held by the central bank. Both demand deposits and bank reserves fall. As a result, the commercial banks have smaller excess reserves. If they were fully loaned up before, they must now liquidate some earning assets. In this case the Treasury's handling of the surplus has led to a tightening of the money market, the effect of which is added to that resulting from the initial decline in the public's cash holdings.

If the Treasury uses the surplus to retire bonds held by the commercial banks, a corresponding amount of deposits is wiped out; but there is no change in total member-bank reserves. With smaller deposit liabilities and the same reserves, the banks have higher reserve ratios, so that they can expand their earning assets if they wish. However, their secondary reserves are smaller by the amount of bonds retired. If the Treasury simply accumulates idle deposits at the banks but does not retire any bonds, the banks' reserve position is unchanged. The only effect is to convert some of the public's active deposits into idle Treasury deposits; the velocity of the total money supply declines.

Suppose that the Treasury retires debt held by the nonbanking public. In this case, there is no change in either the money supply or bank reserves. What the government takes away in taxes it returns to the public in exchange for some of the latter's holdings of securities.

The former owners of government bonds now have additional cash, which they may leave idle, spend on goods, or put into other securities. Probably only a small amount will go into consumption. For the most part, investors receiving cash for their government bonds will use these funds to buy corporation securities or other government bonds that are still outstanding. What has happened is that the supply of loanable funds has increased.[42] Interest rates will tend to fall, and security prices will tend to rise. On the whole, this will be stimulating to private investment. Here we have an inflationary influence that may partially offset the original deflationary pressure exerted by the budgetary surplus. The offset may be large or small, depending on how anxious investors are to convert cash into securities and on how responsive private investment is to the greater ease in the money market.

THE INTEGRATION OF MONETARY AND FISCAL POLICY

It is clear that monetary policy and fiscal policy must work hand in hand. The fiscal authorities make a monetary decision when they decide how to finance a deficit or what to do with a surplus.[43] In a full-fledged compensatory policy, deficits should be financed so as to have the maximum *immediate* stimulating effect, but with due regard to possible later repercussions from a large increase in the public's or the banks' holding of liquid assets. Surpluses should be handled so as to maximuze the immediate deflationary effect, without at the same time disorganizing the financial and money markets. In the case of both deficits and surpluses, the monetary authorities should pursue a policy that strengthens the effect being obtained by budgetary action. In depressions, the central bank should keep the commercial banks well supplied with reserves so that the latter can freely buy the securities being offered by the Treasury. When the govern-

[42] The funds that the former bondholders receive came from taxpayers. If the latter would have spent these funds on consumption, then there is an increase in the supply of loanable funds equal to the amount withdrawn from consumption. If the taxpayers would have saved all of the tax receipts used for debt redemption, then there is no net increase in the supply of loanable funds. Actually, the taxes would probably come out of both consumption and saving. Hence, there would be some net increase in the supply of loanable funds.

[43] The Treasury also makes a monetary decision when it decides what kinds of obligations to issue in order to refund debt that is maturing. Thus, management of the existing debt has monetary effects even if the current government budget is balanced.

ment is trying to control a boom through a budgetary surplus, the central bank should put pressure on bank reserves so that the commercial banks cannot readily replace the earning assets which they lose through the government's debt-retirement program. If the public debt is very large, and both the banks and the public are therefore well supplied with liquid assets, the monetary authorities should take steps to insure that these securities remain in stable hands and are not converted into cash and spent. In particular, this means that the commercial banks should not be permitted to unload government securities on the central bank in exchange for additional reserves. At this point we run into the potential conflict between sound monetary policy and the Treasury's desire for low interest rates and stable bond prices, an issue we considered at some length in Chapter 19.

The great increase in the size of the public debt in the last three decades has created a range of monetary problems to which there is no simple solution. As we have seen, debt management cannot be divorced from monetary policy. Changes in the size of the debt and the kinds of securities issued have monetary effects that react on the volume of private spending. The "problem of the debt" is not the burden on the taxpayer or the danger of repudiation. The serious problem is that the debt creates a huge volume of liquid assets which the banks can convert into reserves and which the nonbanking public can convert into cash. From the point of view of economic stability, the danger is that the existence of these liquid assets will cause private spending to rise too rapidly during periods of prosperity and that strong measures to control such inflationary tendencies may lead to financial disorganization. Some of the measures that have been proposed to deal with this situation were briefly discussed in Chapter 19.

The public debt inherited from World War II injected into the American economy an inflationary bias that lasted into the 1950's. And even now, serious problems of debt management remain. These problems fall chiefly under two heads, both of which have been suggested in our earlier discussion. First, the Treasury is faced with the continuing need to refund large amounts of new securities. Each refunding raises questions as to the kinds of new securities that should be issued—at what interest rates, for what maturities, etc. The decisions reached affect the money and capital markets and thus, indi-

rectly, the level of private spending. Second, the large debt outstanding enhances the economy's supply of liquid assets, and this effect is the more pronounced the greater is the fraction of the debt in the form of short-dated securities. A large volume of liquid assets in the hands of the banks and the nonbanking public weakens the relation between the money supply and the volume of private spending and makes the problems of the monetary authorities that much more difficult. This is one reason why a larger fraction of the outstanding federal debt needs to be in the form of long-term securities. The longer maturities are less liquid than the very short-dated securities and can less readily be converted into cash without some decline in market value.[44]

[44] For a review of the problems that have arisen in management of the public debt in the United States since World War II, see W. L. Smith, *op. cit.*; also Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels*, chap. 9.

CHAPTER 21

WAGE AND PRICE POLICIES AND THE PROBLEM OF INFLATION

IN THE search for economic stability, we cannot afford to neglect the behavior of prices and wages. Both may move in response to changes in spending, particularly when the economy is close to full employment. Further, as we saw in Chapter 4, the prices of both goods and labor can vary for reasons other than changes in aggregate demand. Business firms, labor unions, and government can, by the actions they choose to follow, deliberately influence prices and wage rates; and the resulting changes in prices and wages cannot be explained merely by reference to concurrent changes in total demand. These autonomous movements in wages and prices can induce changes in spending, which, in turn, affect output and employment and may lead to further changes in wages and prices, in a cumulative spiral. In the terms used in Chapter 4, aggregate supply as well as aggregate demand can shift upward (or downward); a shift in aggregate supply (implying a new level of wages and prices) will cause aggregate demand to change, which may lead to a further shift in the aggregate supply function, and so on.[1]

Whether a stabilization program needs to include action regarding prices and wages depends largely on the answers to three questions. First, how are prices, including the price of labor, likely to behave as total spending rises and falls? Second, are there autonomous forces operating to cause wages and prices to move independently of changes in aggregate demand? Finally, what effect does the

[1] See the discussion of aggregate supply in Chapter 4, where, among other things, we briefly considered the possible interaction between aggregate demand and aggregate supply.

movement of wages and prices have on the subsequent course of production and employment?

These are difficult questions, and economists are by no means agreed as to the answers. Before World War II, these questions were discussed primarily in terms of wage-price behavior in depressions. Since the war, the emphasis has, of course, shifted. The problem has been that of rising, not falling, prices, and the big policy question has become: how can we restrain the upward push of prices without sacrificing either full employment or a satisfactory rate of growth? In the next section we shall consider briefly wage-price policy in business contractions. The rest of the chapter will be concerned with ways of restraining the sort of inflationary pressures that we had in the latter half of the 1950's.[2]

WAGE-PRICE POLICY IN DEPRESSIONS

Let us consider first what is likely to happen during a business downswing. The decline in aggregate demand will show itself immediately in a decline in output. Flexible prices also will fall, especially in highly competitive industries and where commodities are bought and sold in organized markets. But in oligopolistic industries, where firms have some control over the prices they charge, prices will, at first, move down slowly or not at all. Autonomous price-making forces are at work here, stemming from the policies that businessmen choose to pursue. They have considerable discretion as to the price policies they adopt to meet the decline in demand.[3]

One result is that business downswings frequently generate distortions in the price structure. Two examples that have been common in past depressions may be mentioned. Farm prices have fallen much more than industrial prices. This was particularly true in the Great Depression of the 1930's. Costs and prices in the construction industry (and also in other capital-goods industries) tend to be sticky and to remain high when the demand for building falls off. A more balanced price decline, and particularly lower prices for investment goods, would undoubtedly tend to moderate the severity of a serious

[2] From 1946 to 1955, the rise of prices in the United States was more obviously induced by changes in aggregate demand than was the case afterward. This is not to deny, however, that autonomous forces did not operate to some extent in the earlier period also. Cf. footnote 13, below.

[3] See Chapter 10 for a description of typical price behavior during cyclical downswings.

downswing. Here is room for government intervention in the field of prices—to prevent a demoralizing decline in prices in competitive industries (especially agriculture) and to force down costs and prices where monopolistic forces are at work, especially in the construction and other investment-goods industries.

Prices, obviously, are strongly influenced by wages. A vigorously debated issue among economists has to do with the influence of wage cuts on the volume of employment and output during business depressions.[4] Wages are notably inflexible in a downward direction. Even before the recent rapid growth in the strength of organized labor, wages tended to lag behind prices in cyclical downswings. Today, with a large part of the labor force organized in strong unions, this downward inflexibility of wages is greater than ever before.

Does the maintenance of wages during business downswings tend to make depressions more or less severe? This question can be argued either way, depending on the assumptions that one makes. Wages have both a *cost* and an *income* effect. If wages remain high as prices sag, the cost effect reduces profit margins and will lead businessmen to reduce their demand for labor. If wages were lower and profit margins higher, employment might be enough greater to bring about an increase in payrolls, despite lower wage rates. But if employment does not rise enough to offset the wage reduction, consumers' spending will fall in response to the decline in payrolls, and this may accentuate the downward spiral.[5] On the other hand, the high wage rates and low profit margins shift the distribution of income in favor of wage-earners, reduce the ability of profit receivers to save, and thus push up the economy's propensity to consume. But the low profit margins also reduce businessmen's willingness to invest. Hence, it is difficult to say what the net effect will be on total spending. Perhaps the most crucial issue concerns the state of business expectations. If businessmen expect demand and prices to continue to fall, they are not likely to increase employment significantly in

[4] Cf. G. Haberler, *Prosperity and Depression*, 4th ed., 1958, pp. 239 ff., 395 ff., 491 ff.; A. H. Hansen, *Monetary Theory and Fiscal Policy*, 1949, chap. 8; Rendigs Fels, "The Effects of Price and Wage Flexibility on Cyclical Contraction," *Quarterly Journal of Economics*, vol. 64, November, 1950, pp. 596–610.

[5] Another question that needs to be examined here is how prices will behave if wages are reduced. We have assumed in the text that a wage cut will not automatically bring about a fully corresponding fall in prices immediately, so that profit margins will be affected to some extent.

response to a wage reduction. Indeed, the wage reduction may lead businessmen and consumers to expect further price and wage reductions. In this case, the wage cut will make matters worse rather than better.

One argument does point in favor of wage reductions fully matched by price declines during a business downswing. That is the monetary argument. If all costs and prices decline, the real or goods value of the community's stock of money and liquid assets rises correspondingly. Eventually, this rise in the real value of liquid assets, including money, is likely to lead to an increased propensity to spend by consumers and business firms. But this effect may be long delayed if business expectations are very unfavorable and are made even more so by continuously falling prices.

On the whole, wage reductions do not look like a promising way of stopping business downswings. In particular, most economists would now agree that *continuous* downward pressure on wages during a business decline would do more harm than good. If wages continue to sag, businessmen will wait for prices and wages to go still lower. Probably the best policy is for the government not to try to bring about a general reduction in wages if unemployment develops and to rely on the measures to stimulate aggregate demand discussed in the preceding chapters. If these are not fully successful, then a once-for-all wage reduction might be tried, in such a way that no further reductions will be anticipated.

Experience since World War II reinforces the conclusion that downward pressure should not be exerted on wages in recession, at least under the conditions that have prevailed since the war. If underlying investment opportunities remain favorable in a contraction (so that the decline in investment is chiefly in inventories), if the automatic stabilizers are working strongly, and if an appropriate monetary-fiscal policy is being followed, then minor contractions are likely to be milder and end sooner if wages are maintained than if they are cut. Indeed, under these circumstances, a significant decline in wages will generate fears of falling prices and will tend to make matters worse rather than better.

The question of depression wage policy must be evaluated in the context of the overall stabilization program being followed. If a government is committed to the full-employment objective and is prepared to utilize intelligently and vigorously the monetary, fiscal,

and other measures previously described, it should be possible to stop a business downswing without important reductions in the general level of wage rates. But the attempt should be made to improve the wage and price *structure*—particularly by removing monopolistic controls in fields where lower costs and prices would very probably lead to an increase in the volume of investment. This would be especially important if there were evidence that the long-run level of investment was deficient.[6]

FULL EMPLOYMENT AND INFLATION

For more than a decade, the threat of inflation has been much more in economists' minds than the danger of serious depressions. The inflationary side of the wage-price problem raises two critical and related questions: When aggregate demand rises, whether during the natural course of a business upswing or in response to governmental measures, is it likely that the increase in spending will push up wages and prices? Secondly, is there danger, if full employment is continuously maintained (and perhaps even if it is not), that the demands of strong labor unions and other organized groups will cause wages and prices to rise steadily, so that the objective of price stability becomes unattainable? Or to put it more bluntly, may not full employment and price stability be incompatible as objectives of policy?[7]

Let us see how wages and prices are likely to behave when demand is expanding. As long as there are excess capacity and considerable unemployment, increased spending is likely to be reflected chiefly in rising production and employment, though some price increases will

[6] Cf. H. S. Ellis, "Monetary Policy and Investment," in *Readings in Business Cycle Theory*. See also his essay in the Twentieth Century Fund volume, *Financing American Prosperity*, 1945.

[7] The literature on the causes and control of inflation, even that written in the last decade, would fill a good many library shelves. One useful collection is the final *Staff Report* and the *Study Papers* prepared for the Joint Economic Committee study of *Employment, Growth, and Price Levels* (86th Congress, 1st and 2nd sessions, 1959–1960). The *Staff Report* and Study Paper no. 1 by C. L. Schultze, *Recent Inflation in the United States,* are particularly useful. The monograph by Schultze is also helpful for its bibliographical references. Another good source is The American Assembly, *Wages, Prices, Profits, and Productivity*, 1959. For an introductory treatment, see W. L. Thorp and R. E. Quandt, *The New Inflation*, 1959. For more sophisticated theoretical treatment, see the references in Schultze's Study Paper to the work of Bent Hansen, Duesenberry, Dow, and others.

also occur. As full employment is approached, but even before it is reached, bottlenecks will begin to appear—in the form of material, capacity, or labor shortages in particular parts of the economy. Prices and wages in these areas will probably rise sharply. As demand expands still further and output reaches substantially the full-employment level, wage and price increases become more general. Beyond the point of normal full capacity, cost curves rise rapidly; the supply of raw materials cannot be rapidly expanded; and unions, spurred on by the rising cost of living and large business profits, press vigorously for wage increases.

If the underlying investment boom is a strong one, or if the government continues to support the boom after private investment begins to fall off, supply-generated price and wage increases may create an inflationary spiral. Higher wages lead firms to raise prices; the anticipation of higher prices leads to increased spending which supports the higher level of prices; rising prices lead to new rounds of wage increases, which cause prices to rise still further; and so on. In addition, with labor fully employed and no excess capacity, there are numerous bottlenecks; employers bid against each other for labor and materials; and it is difficult to adjust supply to shifts in demand. The inflation may be a gradual one and go on for a long time if the government steps in to offset any tendency toward a decline in spending. In this case, autonomous wage and price influences, backed by the government's full-employment policy, lead to a persistent and creeping inflation. In some cases, particularly right after wars, the inflationary expansion may be very rapid.

Even without the stimulus of wars, continuously maintained full employment can lead to a steady rise in wages and prices. As one writer has put it, "Full employment is loaded with inflationary danger."[8] If unions do not have to fear unemployment, if they are assured that the government will step in to stimulate total demand if employment should fall off at all, they are likely to press steadily for higher wages. If, as is likely, the resulting wage increases exceed the normal improvement in productivity, labor costs will rise, and businessmen will raise their prices. With higher industrial prices, farm groups will demand government help in raising agricultural prices. And the general rise in prices will lead to a new round of

[8] John T. Dunlop, "Wage-Price Relations at High Level Employment," *American Economic Review*, vol. 37, May, 1947, suppt., p. 243.

wage increases. If, at any time, business expectations deteriorate and some unemployment develops, the government, if it is committed to a full-employment policy, will step in to stimulate demand, and thus the inflationary process can go on. If the government tries to stop the rise in wages and prices—say, by embarking on a deflationary fiscal policy—it must then be prepared to put up with some unemployment, if only for a short period.[9]

A good many economists have become convinced that a consistently and successfully maintained full-employment policy inevitably leads to one of two results. Either the goal of price stability must be abandoned, or else full employment can be maintained without open inflation only by the government's interfering with traditional economic liberties—by depriving labor of its right to bargain freely with employers, by imposing wage and price controls, and by direct control of consumer and business spending through rationing, allocations, and similar devices. In this latter case, "repressed" inflation is substituted for "open" inflation.

Other economists are more hopeful and believe that organized labor and other groups can be educated to keep their demands within bounds—so that full employment can be maintained without a steady rise in the price level. Some observers, while agreeing that a full-employment policy leads to inflation in the absence of controls, prefer to limit the power of labor and other groups if this is necessary to achieve both price stability and full employment.[10]

Given the evidence thus far available, there does seem to have been a long-run upward bias in the price level since World War II —not only in the United States but also in the rest of the world.[11]

[9] The preceding paragraphs make no attempt to provide a rigorous demonstration of how inflation may occur in a fully employed economy. For this, see the references cited in footnote 7, particularly Schultze, *op. cit.* For our present purpose, it is sufficient to assume that "demand-pull" and "cost-push" factors interact to generate an inflationary process. The argument in the remainder of this chapter assumes only that, while both of these two sets of factors have been at work since World War II, it is the existence of the "cost-push" elements (what we have called the autonomous forces making for wage and price increases) that create special problems that cannot be solved by the use of conventional stabilization (i.e., monetary-fiscal) policy.

[10] A good deal of the remainder of this chapter is taken from a statement by the author in Joint Economic Committee, *Study of Employment, Growth, and Price Levels,* Hearings, Part 9-A, pp. 2955 ff.

[11] For a comparison of the movement of prices and wages in the United States and a number of other countries after World War II, see the papers by L. G. Reynolds and S. H. Slichter in American Assembly, *op. cit.,* pp. 110, 169; also M. W. Leiserson, *A Brief Interpretive Survey of Wage-Price Problems in Europe,*

The performance of both the American economy and those of other industrialized nations suggests that the more successful we are in achieving rapid growth and substantially full employment, the more likely is it that there will be a significant upward trend in the price level—more of an upward trend than most of us would like to have. This basic incompatibility is likely to continue unless there are some fundamental changes in the structure of the economy (changes which, on the whole, most Americans would not be willing to accept), or unless we can make more effective use of the policy tools now available or develop some new tools to add to those we now have.

Some of the reasons for this persistent inflationary pressure have already been suggested, but it may be worthwhile to elaborate on them at this point.

1. The strength of organized labor, obviously, has increased greatly since the 1930's. Given this strength and what we may call the internal dynamics of the trade union movement, there has been a persistent pressure to raise money wages at a rate that drives up unit labor costs. This pressure has been self-reinforcing. The succession of past wage increases has created a set of expectations such that union leaders feel compelled to insist on significant wage increases every year and employers feel obligated to grant a substantial part of what is asked.

2. In this connection, it is probably fair to say that employers have come to be willing to pay a high price for industrial peace. More accurately, employers want industrial peace in order to be able to take advantage of the sustained high level of demand, and developments since the war suggest to them that the price that has to be paid can be shifted to the consumer. This attitude might change in the face of any considerable period of excess capacity.

3. While the extent of "administered pricing" (by sellers with some control over the prices they charge) is probably no greater now than before the war, the prevalence of "mark-up pricing"—which extends through a large part of the economy and is not confined to the so-called "concentrated" industries—enhances the effectiveness of union pressure for higher wages and increases the likelihood that cost advances will be passed on as price increases. "Administered

---

Study Paper no. 11 for the Joint Economic Committee, *Study of Employment, Growth, and Price Levels,* 1959. The rise in prices has been less in the United States than in most other countries.

pricing" has always been widespread. Since the war, it has interacted with demand influences and union pressure to raise prices as well as wages. This is related to the previous point that employers have come to feel that industrial peace can be bought by wage increases that can be passed on as price increases.

4. While administered pricing has always been with us, it is probable that its character has changed somewhat in the last generation. Not only has there been a weakening of employer resistance to wage demands, but there has also been a growing resistance to reducing prices when demand declines in particular sectors. Not only wages but also prices have become more rigid in a downward direction. As a result, the economy has shown an increasing tendency toward a certain asymmetry in its price behavior.[12] Excessive demand in particular sectors pushes up wages and prices in those sectors; these wage and price increases are transmitted to other sectors, leading to a general rise in costs. If aggregate demand is not excessive, excess demand in some sectors means deficient demand in others. Yet prices tend not to fall in the latter. Indeed, they may rise because of the general rise in costs initiated in those industries facing an exuberant demand.

Another aspect of this asymmetry is that nonagricultural prices generally have been showing increasing resistance to decline in business recessions. Thus we are faced with what has been called the "ratchet effect." The price level moves up during cyclical expansions; it falls little or not at all and may even show a slight rise during business contractions.

5. Overhead costs per unit of output have risen significantly in the postwar period. (Industry has been using relatively more clerical and administrative workers and relatively less direct labor.) This has probably accentuated the tendency in oligopolistic industries toward full-cost pricing. It has probably also led some labor unions to exaggerate the increase in productivity of direct labor, giving rise to wage increases that have forced up unit labor costs and prices.

6. For a variety of reasons that we have already examined, the economy during the postwar period has been more resistant to cyclical contractions than before the war. This, of course, is as we want it

[12] Cf. Schultze, *op. cit.;* also J. C. R. Dow, "Analysis of the Generation of Price Inflation," *Oxford Economic Papers,* vol. 8, October, 1956, pp. 252–301.

to be. But the weaker are cyclical contractions in aggregate demand, the less likely are price increases in boom periods to be offset by price declines in recessions. In addition, both business and labor have apparently come to count heavily on the guarantees offered in the Employment Act. The effect on expectations is such as to increase the resistance to both wage and price reductions, and these expectations may also tend to make the recessions milder than they otherwise would be.

7. It is probable that, under the conditions that have existed in the postwar period, different rates of productivity increase in different industries contribute to the inflationary trend. At least, this is likely to be true if there are strong trade unions, particularly in industries with relatively high rates of productivity increase, and if aggregate demand is high enough to maintain a fairly tight labor market. Under these circumstances, the wage increases that set the pattern are likely to reflect above-average increases in productivity. Wages rise more or less correspondingly in industries with less-than-average increases in labor productivity. Prices must therefore rise in the latter sectors. Thus, we may get a situation in which relatively few prices decline and many rise, those that rise most sharply being in industries characterized by strong demand and little increase in productivity. A process somewhat of this sort has probably been responsible for part of the particularly rapid rise in the prices of many services since the war and also for the extent of the rise in some types of construction costs.

This is only a partial list of the factors that have helped to push up prices in recent years. This list suggests that there have been autonomous forces at work leading to a rise in wages and prices even in the absence of any large degree of excess demand. Some of the factors emphasized, however, would have had much less inflationary consequences had not the level of aggregate demand (and of the demand for labor) been, on the whole, high and rising during the postwar years. Neither a simple "demand-pull" nor an unqualified "cost-push" explanation can account for the movement of cost and price levels during the period 1945–1960.[13] But the important point

[13] The upward movement in prices since World War II has come in three waves—1946–1948, during the Korean War, and after 1955 (particularly during 1956–1957). Demand factors predominated in the first two periods much more than in the third. See the history of the postwar period in Chapters 15–16.

is this: *If* we do succeed in keeping aggregate demand high enough to permit something close to full employment and a reasonably rapid rate of growth, then the institutional arrangements which today impinge on the price-making process are virtually certain to impart a significant upward tilt to the trend in the price level. In short, under today's conditions, high-level employment and rapid growth are probably not consistent with a secularly unchanging price level.

SOME SUGGESTIONS AS TO POLICY

The preceding discussion suggests that, in the kind of world in which we now live, the conventional instruments of monetary and fiscal policy described in Chapters 19 and 20 are not enough to give us the combination of goals that we want—namely, not only full employment and rapid growth but also price stability. Monetary and fiscal policy have their effect primarily on aggregate demand, but the preceding section suggests that merely controlling aggregate demand may not be sufficient to prevent a steady upward trend in the price level. There is one way to hold down the price level by measures that operate only on aggregate demand. This is to exert enough deflationary pressure so that a substantial amount of unemployment restrains trade unions from asking for large wage increases and so that widespread excess capacity makes general price increases impossible. For most Americans, this is a clearly unacceptable alternative. If a choice has to be made, most of us would put the goal of full employment ahead of that of price stability.[14]

What then can be done? The uncomfortable fact is that in our kind of economic system, which emphasizes freedom of choice by individuals and private groups, we have no effective way of control-

[14] It is conceivable but not probable that, even with the pressures described in the preceding section, prices can be prevented from rising by a policy that leads to an amount of unemployment the public is quite willing to tolerate for an indefinite period. Our assumption is that this is not likely in today's setting. But we do assume in the following discussion (1) that monetary-fiscal policy is used to prevent the development of an inflationary gap and a situation of chronic overfull employment and (2) that mild recessions with moderate amounts of unemployment do occur and are allowed to bring about needed price adjustments, even if the overall price level does not fall very much. For a consideration of this range of issues in the context of postwar Swedish policy, see Erik Lundberg, *Business Cycles and Economic Policy*, 1957, chap. 10.

ling the autonomous forces making for wage and price increases previously described. At least, we have no effective way without resorting to the kinds of government intervention that we are unwilling to tolerate except in wartime.

If these inflationary pressures continue, we need to develop some new techniques so that, while remaining within the framework of existing institutions, the government can exert some influence over the autonomous price-making forces that we have described. Admittedly this is difficult to do. None of the proposals thus far made commands general acceptance, and this is true of the suggestions that we offer in the following paragraphs. Our suggestions, which have also been made by others, seek to influence the environment within which private wage and price decisions are made but avoid the use of direct controls and prohibitions. These suggestions, in effect, propose that the federal government take a more active role in inducing labor and business leaders to follow policies that are consistent with our stabilization objectives.

1. First of all, we suggest that, at least once a year, probably soon after submission of the President's Economic Report, a representative group of labor leaders and businessmen should be brought together for intensive meetings with government officials.[15] The latter would presumably be drawn from the President's Council of Economic Advisers, other departments of the federal government, and from the Board of Governors of the Federal Reserve System. Exhortation should be kept at a minimum; but the current and prospective economic situation should be explored thoroughly, current government policies should be discussed, and there should be consideration of what private policies (including but not confined to wage and price policies) are most consistent with the goals of high employment, rapid growth, and a workable degree of price stability.[16] Through such meetings as these, or through some other form of regular conference, the attempt should be made to bring about better coordination between monetary-fiscal policy and private wage and price policies.

[15] The President submits his Economic Report to Congress in January of each year. Submission of this report is required by the Employment Act of 1946.

[16] See the similar suggestion made by John Dunlop in The American Assembly, *Wages, Prices, Profits, and Productivity*, pp. 148–149. The same proposal is included in the final report of the Assembly at which Dunlop's paper was presented.

2. In this connection, there needs to be a closer working relationship between the Federal Reserve System and organized labor. The Board of Governors apparently does not have any systematic procedure for regular conferences with labor leaders. No representative of labor sits on the board. The Federal Reserve Act states that, in his appointments to the board, the "President shall have due regard to a fair representation of the financial, agricultural, industrial, and commercial interests. . . ," but no mention is made of labor. A somewhat similar situation prevails with respect to the boards of the individual Federal Reserve Banks. It might be wise to have labor representatives on the boards of at least the twelve Federal Reserve banks, and both the Board of Governors and the individual Reserve Banks should arrange to meet regularly with representatives of labor. The monetary authorities need to be sensitive to the interest of workers in the level of employment and the rate of growth, and these suggestions might help labor to recognize the interrelations between its wage demands and the stabilization objectives of monetary policy.

3. The President's Economic Report should be made into a more effective instrument for bringing public pressure to bear on private wage and price decisions. The Economic Report, virtually from the beginning, has tended to run in terms of broad generalities, well-meaning exhortations, and innocuous platitudes.[17] The Economic Report should be more explicit about the objectives to be sought during the coming year and the combination of public and private policies needed to attain these objectives. In this connection, the Administration should not be afraid to state frankly what range of wage increases it believes to be compatible with price stability, what needs to be done to accelerate increases in productivity, what policy measures it is prepared to take under various contingencies, and so on. There are obvious risks in having the government engage in such frankness. But we cannot get away from the fact that, given the economic power now residing in private groups and our desire to maintain existing free institutions, we must find ways of bringing public opinion to bear on the parties at interest in order to supplement the conventional instruments of economic policy. As one economist with much practical experience in labor negotiations has put

[17] There has not been much difference between Republican and Democratic administrations in this respect.

it, "the full potential of the leadership of the federal government has never been used persistently and imaginatively to shape decisions by private parties on wages and prices or to influence the climate of ideas within which such decisions are made."[18]

4. It would probably strengthen the government's hand somewhat to have the goal of price stability written into the Employment Act. But this should be done in such a way as to permit a somewhat flexible interpretation of price stability and so as to recognize the prior status of high employment and satisfactory growth.[19]

5. More vigorous enforcement of the antitrust laws, as well as a more liberal tariff policy, would tend to increase the degree of competition in American industry and weaken the market power of large firms. Such a program is desirable on a number of grounds. By itself, however, it cannot be expected to eliminate the kinds of inflationary pressures that we have been dealing with in this chapter.

6. It is possible for the government to intervene in the determination of wages and prices without going so far as to impose mandatory controls on specific wages and prices. The following comments of the staff of the Joint Economic Committee are worth repeating in this connection:[20]

> We believe there should be a presumption against Government intervention in wage and price determination, unless the circumstances involved make it necessary. If this approach were to be utilized, several alternatives are available, reflecting increasing degrees of intervention. These would include establishment of a study group to advise the President on important price and wage changes; the use of factfinding procedures, with or without the issuance of a report and recommendations; the requirements of prior notification to the Government of proposed price or wage increases in certain key industries; the power to suspend such increases; and, finally, direct price and wage controls.
>
> At this stage of our knowledge and experience, we believe that if such an approach were to be utilized, it should be limited to the establishment

[18] Dunlop, *op. cit.*, p. 148.

[19] The Employment Act imposes on the federal government the obligation "to promote maximum employment, production, and purchasing power," but it makes no explicit reference to price stability as a goal. For a range of views on this subject, see the "Compendium of Papers" submitted to the Joint Economic Committee, *The Relationship of Prices to Economic Stability and Growth* (85th Congress, 2nd session, 1958).

[20] Quoted from Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels,* pp. 439–440.

of factfinding procedures to be invoked at the discretion of the President and to result in the issuance of a report and recommendations regarding the justification and desirability of price or wage increases. . . .

Such recommendations would not be binding on the unions and business firms involved. As in the first three of our recommendations, the purpose would be to affect the environment of opinion within which price and wage decisions are made.[21]

7. A variety of governmental measures can be used to accelerate the upward trend in labor productivity—for example, larger expenditures on education, greater support of scientific research, steps to reduce restrictive labor practices and improve labor mobility, and so on. The faster the rise in labor productivity, the less the rise in labor costs per unit of output that results from a given rate of increase in wages, and the weaker will be the pressure on business firms to raise their prices.

CONCLUSION

We suggested in Chapter 18 that a feasible combination of goals to strive for was an average rate of unemployment of not more than 4 percent, a 4 percent rate of growth, and a rise in the price level that does not exceed 1 percent per year on the average. While this combination of objectives is possible, it will not be easy to achieve. If there are strong inflationary pressures at work from the side of aggregate supply, the measures suggested in this chapter would make some contribution to holding back the rise in prices without depressing aggregate demand, and this would make it easier to achieve the employment and growth objectives.

This chapter might properly end with a few warnings.

1. Heavy reliance on monetary-fiscal policy to restrain inflation may involve significant costs in the form of unemployment or impeded growth. At the other extreme, highly expansionist monetary-fiscal policy aimed at stimulating growth may have undesirable consequences on the trend in prices. A corollary is that selective monetary and fiscal measures, which aim at controlling the level of

[21] It is only fair to say that a good many economists, legislators, and government officials—not to mention businessmen and labor leaders—have serious misgivings about this sort of government intervention, even if no mandatory controls are involved. For a review of the issues involved, see E. S. Redford, *Potential Public Policies to Deal with Inflation Caused by Market Power,* Study Paper no. 10 for the *Study of Employment, Growth, and Price Levels,* 1959.

spending in particular sectors, may frequently be more useful than overall measures that operate on aggregate demand as a whole.

2. The stabilization objectives should be interpreted flexibly. We cannot expect to eliminate all fluctuations in employment, output, and prices. A goal of not more than 4 percent unemployment, on the average, is compatible with a higher rate of unemployment during brief recession periods, provided the unemployment rate falls below 4 percent during boom periods. Similarly, prices during cyclical booms are almost certain to rise at a faster rate than that which we take as our long-term goal. This should not lead to a strong dose of deflationary medicine if we can reasonably count on some decline in prices later on.

3. It is a safe assumption that the world will continue to change. During the 1950's, it seemed clear that we did have to worry about inflation, that the inflationary forces at work came partly from the supply side, and that the forces operating on aggregate demand were such as to keep the business recessions of the period brief and mild, without any large-scale government intervention. We cannot say in what respects the future will be different. One possibility—which is not a comfortable one to contemplate—is that we shall have to work harder to maintain full employment while at the same time the influences described in this chapter will continue to generate some upward pressure on prices. This is not a prediction. It is merely to suggest that we need to develop a flexible set of policy measures capable of dealing with a variety of possible developments. Policies that seemed appropriate in the recent past are not necessarily those that will always work best in the future.

CHAPTER 22

INTERNATIONAL ASPECTS OF THE PROBLEM OF INSTABILITY

THE THEME of this chapter was concisely stated in a report published by the League of Nations toward the close of World War II: "Cyclical fluctuations in business activity in an economically integrated world are not a national but an international phenomenon, and require not only national but international action."[1] Nations are bound together by a network of trading and financial relationships, although these ties have been weakened in various ways during the last thirty years. Once a boom or depression begins in one part of the world, it tends to spread to other countries. The mechanism through which this extension occurs is similar to that which causes a change in business activity in one region of a country to be transmitted promptly to other parts of the same national economy.

Countries are particularly vulnerable to economic disturbances originating in other parts of the world if they rely heavily on international trade. Thus, though booms and depressions usually begin in the more highly industrialized economies, they spread quickly to less developed countries whose economies are closely geared to the export of a few basic raw materials or agricultural products. This sensitivity to outside cyclical disturbances holds also for industrial-

[1] League of Nations, *Economic Stability in the Post-War World,* 1945, p. 17. Among other volumes on the international aspects of business fluctuations, see T. C. Chang, *Cyclical Movements in the Balance of Payments,* 1951; J. J. Polak, *An International Economic System,* 1953; Oskar Morgenstern, *International Financial Transactions and Business Cycles,* 1959; Ilse Mintz, *Trade Balances During Business Cycles: U.S. and Britain Since 1880,* National Bureau of Economic Research Occasional Paper 67, 1959.

ized countries that engage extensively in international trade. For example, in reply to a questionnaire from the United Nations regarding programs for maintaining full employment and economic stability, the Belgian government commented that "in a country like Belgium . . . the problem of unemployment is related above all to the problem of foreign trade." The Netherlands, in replying to the same questionnaire, stated that a policy of high employment would be followed "as far as international circumstances will not frustrate such a policy."[2]

This is why the entire Western world is so deeply concerned about the problem of instability in the United States. The American economy has shown itself to be particularly unstable. It is true that, relative to its total national income, the United States does not rely heavily on foreign trade. But, because of the size, wealth, and extent of industrialization of this country, the welfare of many other countries is bound up with their ability to sell goods in the American market and the terms on which they can secure goods and capital from the United States. Even before World War II, the United States exported more than any other country, and it was the largest importer of raw materials. In addition, it was the world's leading creditor nation.[3]

Since the early 1950's, the international economic position of other countries, particularly in western Europe, has been greatly strengthened. The "dollar shortage" of the first years after World War II has disappeared, and the minor American recessions during the 1950's were felt abroad much less than was generally predicted. Thus, while it is still true that the United States has a greater impact on the world economy than any other country, what happens in other countries—particularly in the larger European nations—also has an important effect on international economic conditions. It is not so true now as it once was that "when the United States sneezes, the rest of the world gets pneumonia."[4]

[2] These quotations are from United Nations, *Maintenance of Full Employment*, 1949, p. 10.

[3] Much the best study of America's role in the world economy during the interwar period is H. B. Lary and Associates, *The United States in the World Economy*, United States Department of Commerce, 1943.

[4] See, for example, E. M. Bernstein, *International Effects of U.S. Economic Policy*, Study Paper no. 16 for Joint Economic Committee, *Study of Employment, Growth, and Price Levels* (86th Congress, 2nd session, 1960).

THE INTERNATIONAL OCCURENCE OF BUSINESS CYCLES

PARALLELISM BEFORE WORLD WAR II

Before World War II, there was a clear-cut tendency toward parallelism in the cyclical fluctuations of different nations. There were, of course, numerous irregularities, and minor cycles occurred more frequently in the United States than in other economies. But the major cyclical turning points occurred at about the same time in

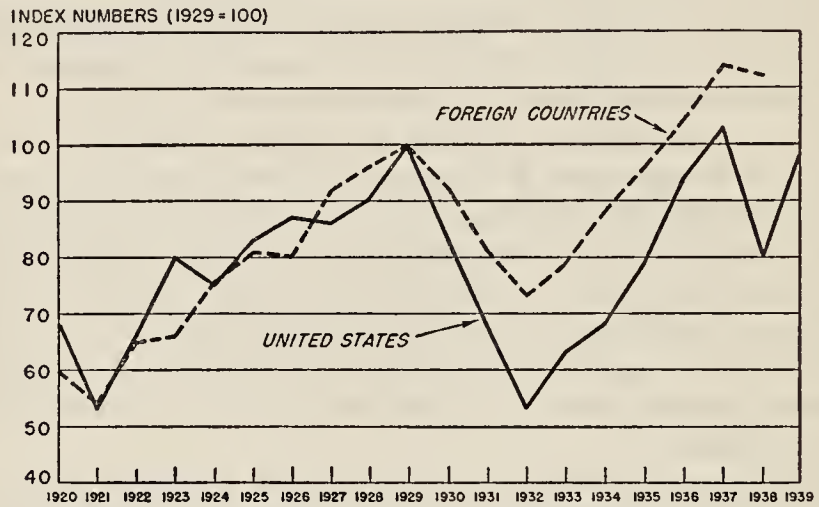


FIGURE 47. Industrial Production in the United States and in Foreign Countries, 1920–1938.

Russia is not included in the series for foreign countries. From H. B. Lary and Associates, *The United States in the World Economy* (U.S. Department of Commerce, 1943), p. 30.

most countries. Some evidence regarding the international synchronization of business cycles is presented in Figure 47 and Table 39.[5]

Figure 47 compares industrial production in the United States and in foreign countries during the interwar period. The boom of the 1920's was a world-wide phenomenon, and so was the depression of the 1930's. Thus, pretty much the whole world (outside Russia) shared in the major cycle of 1921–1932. Figure 47 also suggests, however, that there were some important differences in the cyclical behavior of the United States and the rest of the world. In particu-

[5] For further evidence on the relation between American business cycles and those in the rest of the world, see Mintz, *op. cit.*, and Morgenstern, *op. cit.*

lar, the cyclical swings in business activity were wider in the United States than in most other countries. Also, the United States went through a series of minor cycles which were not precisely duplicated in the economies of other nations.

Further details regarding the international synchronization of

TABLE 39. Cyclical Turning Points in Four Countries, 1879–1938 [a]

United States		Great Britain		France		Germany	
Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough
	3/79		6/79		9/79		2/79
3/82	5/85	12/82	6/86	12/81	8/87	1/82	8/86
3/87	4/88						
7/90	5/91	9/90		1/91		1/90	
1/93	6/94		2/95		1/95		2/95
12/95	6/97						
6/99	12/00	6/00	9/01	3/00	9/02	3/00	3/02
9/02	8/04	6/03	11/04	5/03	10/04	8/03	2/05
5/07	6/08	6/07	11/08	7/07	2/09	7/07	12/08
1/10	1/12						
1/13	12/14	12/12	9/14	6/13	8/14	4/13	8/14
8/18	3/19	10/18	4/19	6/18	4/19	6/18	6/19
1/20	7/21	3/20	6/21	9/20	7/21	5/22	11/23
5/23	7/24	11/24	7/26	10/24	6/25	3/25	3/26
10/26	11/27	3/27	9/28	10/26	6/27		
8/29	3/33	7/29	8/32	3/30	7/32	4/29	8/32
				7/33	4/35		
5/37	6/38	9/37	9/38	6/37	8/38		

[a] Taken from A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, 1946, pp. 78–79. A few of the American turning points have been changed slightly on the basis of more recent information.

business cycles are given in Table 39, which lists the National Bureau's dates for the cyclical turning points since 1879 in the United States, Great Britain, France, and Germany.[6] All four countries participated in the two major cycles between 1879 and the 1890's, but only the United States experienced minor cycles during this

[6] A country that is more closely tied to the United States than these European countries is, of course, Canada. For a careful study of the business-cycle connections between Canada and the United States, which suggests that there may have been some modest decline in Canadian sensitivity to American business conditions, see two articles by G. Rosenbluth in *Canadian Journal of Economics and Political Science,* vol. 23, November, 1957, pp. 480–503, and vol. 24, February, 1958, pp. 21–43.

period.[7] From 1900 on, minor cycles show up to some extent in the other countries also, though the synchronization with American cycles is by no means precise.

The wartime cycle of 1914–1918 and the postwar boom and decline of 1919–1921 were world-wide phenomena. The only exception revealed in Table 39 is explained by the hyperinflation in Germany after World War I. Although all four of the countries went through minor cycles in the 1920's, the sychronization seems to have been less close than during the 15 years before 1914. This is not surprising in view of the disordered state of international monetary and trading relationships during the decade following World War I. However, activity in all four countries began to decline in 1929–1930, and in none of the four did recovery from the Great Depression begin before the middle of 1932.

DEVELOPMENTS SINCE THE 1930's

The economic ties binding the nations of the world together were significantly weakened during the 1930's. The international gold standard disappeared; governments interfered with the free flow of trade through the introduction of exchange control, import quotas, bilateral agreements, and other quantitative trade restrictions; and various countries embarked on internal stabilization programs which, when supplemented by direct control of foreign trade and of capital movements, had the effect of at least partially insulating their economies from fluctuations appearing elsewhere in the world.[8] Thus, the sharp American recession of 1937–1938 had no effect on the war economies of Germany and Japan, where military expenditures combined with stringent foreign-trade controls permitted business activity to remain at a high level. In other countries, also, internal deflation was kept to a minimum by domestic expansionary policies combined with direct controls over foreign trade or a willingness to permit the value of their currencies to fluctuate.

This situation continued into the first years after World War II. Until the early 1950's, most of the world suffered from a dollar shortage; other countries needed to buy from the United States more

[7] Of the three European countries, at least Great Britain seems to have experienced minor cycles before 1860, though not in the later decades of the nineteenth century. See the interesting study of British cycles by W. W. Rostow in his *British Economy of the Nineteenth Century*, 1948, pp. 31 ff.

[8] Cf. League of Nations, *International Currency Experience: Lessons of the Inter-War Period*, 1944.

than they could sell to this country. At the same time, most other countries had to embark on heavy programs of economic reconstruction and development. Inflationary pressures were strong; balance-of-payments deficits were common; and governments intervened, as they had in the 1930's, to regulate the flow of funds across national boundaries. In addition, most governments today participate actively in domestic economic affairs to insure a high level of employment and a rapid rate of growth, and many have large-scale domestic programs (housing, social security, etc.) aimed at achieving a wider distribution of the fruits of economic progress.

Since the early 1950's, the economic position of particularly the European countries has improved markedly. Output has grown rapidly; the balance-of-payments position of the rest of the world vis-à-vis the United States has improved to the point that we can no longer talk of a dollar shortage; and restrictions on the international flow of goods and capital have been greatly liberalized. Thus "it has been possible for European countries to eschew most of the restrictive and discriminatory policy weapons which they acquired in the prewar and early postwar years."[9] This return to liberal trade policies has been made possible by a number of factors: American aid and a liberal American foreign economic policy, strong pressure for closer international cooperation, a highly buoyant aggregate demand, and the sophisticated use of domestic stabilization policies that have helped to maintain a high level of employment.[10]

Growth in most European countries was rapid during the 1950's, and only a very attentuated pattern of cyclical fluctuations is evident in the statistical record.[11] (See Figure 48.) Such recessions as occurred were considerably milder than the postwar American contractions. In the first 15 years after World War II, western Europe experienced three very mild recessions—in 1949, in 1951–1952, and in 1957–1958. The first and third represented to some extent a reaction to American recessions—although the first reflected chiefly balance-of-payments difficulties that led to the widespread devaluations of 1949, and the third resulted in good part from restrictive

[9] A. Maddison, "The Postwar Business Cycle in Western Europe and the Role of Government Policy," *Banca Nazionale del Lavoro Quarterly Review,* June, 1960, p. 143.

[10] Cf. Maddison, *op. cit.*

[11] During the 1950's, growth in industrial production was relatively slow in Belgium-Luxembourg, Sweden, and the United Kingdom. It was very rapid in Germany, and also quite rapid in France, Italy, and the Netherlands.

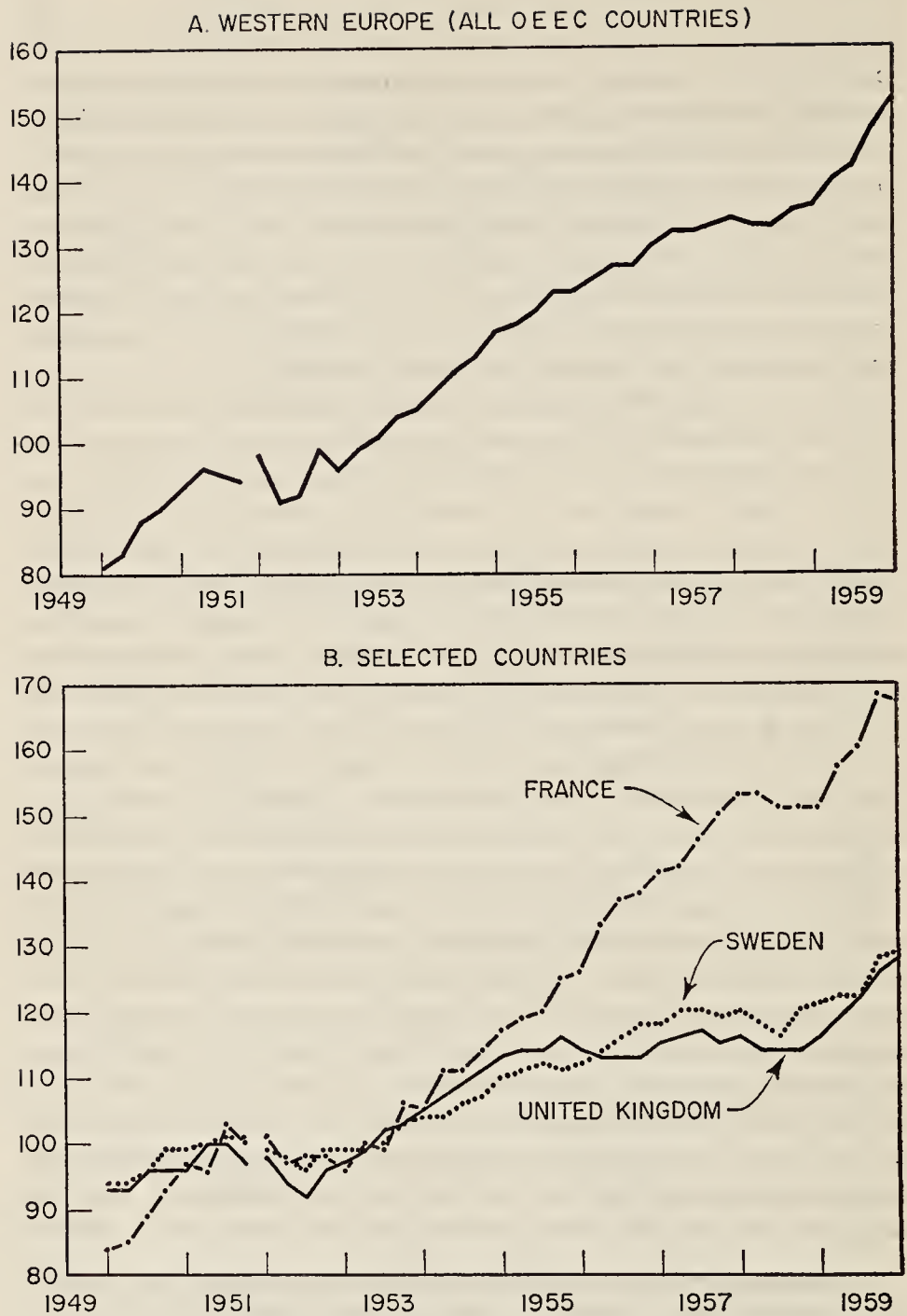


FIGURE 48. Industrial Production in Western Europe, 1950–1959.

OEEC refers to the Organization for European Economic Co-operation. Data are taken from various issues of OEEC, *Statistical Bulletins: General Statistics.*

domestic policies aimed at curbing inflationary tendencies. The 1951–1952 recession, which was the most widespread of the three, was in part a reaction to the inventory speculation in the early months of the Korean war; in addition, balance-of-payments difficulties and strong anti-inflationary measures in some countries were contributory factors.[12]

HOW BUSINESS FLUCTUATIONS ARE TRANSMITTED INTERNATIONALLY

TRANSMISSION THROUGH THE BALANCE OF PAYMENTS

As we have had occasion to note many times, fluctuations in economic activity occur because of changes in total spending. Booms and depressions are transmitted internationally through changes in *international* spending, which involves foreign-exchange transactions and thus the international balance of payments. "To observe the process by which business fluctuations spread internationally, we have only to look at the effects of a boom or depression in one country on the different elements in the balance of payments."[13]

How business fluctuations in one country operate through the balance of payments to induce similar fluctuations in other parts of the world is illustrated by the behavior of the American balance of payments during the Great Depression. The decline in production and incomes in the United States during 1929–1932 led to a sharp falling off in American purchases of goods and services from foreign countries. In addition, foreign investment by the United States virtually ceased. As a result, the number of dollars made available to the rest of the world declined by 68 percent between the end of the 1920's and the bottom of the depression.[14] Countries that sold to the United States found their exports drastically curtailed, and loanable funds for investment projects ceased to be available to countries that had been borrowing in the American market. Produc-

[12] A helpful survey of business-cycle developments in Europe is to be found in Maddison, *op. cit.* See also the annual surveys published by such international organizations as the Organization for European Economic Co-operation (OEEC), the Economic Commission for Europe, the Bank for International Settlements, the Department of Economic and Social Affairs of the United Nations, and the International Monetary Fund.

[13] League of Nations, *Economic Stability in the Post-War World*, p. 92.

[14] Cf. *The United States in the World Economy*, p. 173. The decline in business activity in the United States in 1937–38 was accompanied by a decline of nearly 50 percent in American imports. (*Ibid.*, p. 197.) American imports also reacted sensitively to the mild recession of 1949.

tion and incomes in these countries declined, and so did the prices of the goods they sold. With fewer dollars, they could buy less from the United States and other countries. Thus, the international balance of payments acted as a channel through which deflationary forces were transmitted throughout the trading world. The more a particular nation's economy depends on exports the demand for which is cyclically sensitive, or on the ebb and flow of foreign investment, the more vulnerable it is to deflationary (or inflationary) forces emanating from some other part of the world.

HOW A BOOM SPREADS TO OTHER COUNTRIES

Let us now examine in more detail how business fluctuations tend to spread internationally.[15] We shall begin by assuming that a vigorous boom has begun in one country—say, the United States. Expanding business activity means rising incomes, and some part of the rise in incomes will be spent on foreign goods and services. Thus, an expansion in domestic business activity brings about an increase in imports. How large will be the expansion in imports? This depends on the rise in domestic spending and on the marginal propensity to import, the latter being the proportion of an increase in incomes spent on imported goods. The larger the propensity to import in the country first experiencing the boom, the greater will be the stimulus felt by exporting industries in other countries. However, even in a country with a small marginal propensity to import, a boom may have an important impact on the export industries of other countries if on the average its imports comprise a large share of total world trade. This is clearly the case for the United States.[16]

Thus, the rise in incomes and spending in the country first experiencing the boom stimulates the exporting industries in other countries. The increase in employment and incomes in the export indus-

[15] For more detailed treatment of the theory of the international transmission of business fluctuations, see, for example: Gottfried Haberler, *Prosperity and Depression,* 4th ed., 1958, chap. 12 and pp. 461–473; Ragnar Nurkse, "Domestic and International Equilibrium," in S. E. Harris, ed., *The New Economics,* 1947, chap. 21; W. A. Salant, "Foreign Trade Policy in the Business Cycle," in American Economic Association, *Readings in the Theory of International Trade,* 1949, pp. 201–226; Fritz Machlup, *International Trade and the National Income Multiplier,* 1943; J. M. Letiche, *Balance of Payments and Economic Growth,* 1959; and the additional sources cited in footnote 1.

[16] The *marginal* propensity to import of the United States is low but not as low as the *average* propensity. As a result, the percentage change in imports going with a given percentage change in national income—i.e., the income elasticity of demand for imports—tends to be moderately high. Cf. Chang, *op. cit.,* chap. 2.

tries spills over into the domestic industries of these countries. Through the multiplier, the larger export surplus in each country affected leads to a general rise in incomes, employment, and spending. With the improvement in business activity and business expectations, domestic investment is also likely to rise.

Countries exporting agricultural staples and other raw materials (i.e., so-called "primary" commodities) are particularly likely to benefit from an expansion occurring in a highly industrialized country such as the United States. The increase in their raw-material exports raises domestic incomes, and this will lead to a rise in their imports of finished products from other countries, including the country that initiated the boom. The expansionary impact on the raw-material countries is twofold. Not only is the physical volume of their exports likely to react sensitively to a higher level of incomes in other countries, but also the prices of their major exports are likely to rise significantly. Thus the national income of countries exporting primary products is likely to respond vigorously to an increased demand for their exports, and this will lead to a substantial rise in their imports from other countries. At the same time, the rise in prices of international staples will lead to a worldwide improvement in expectations, may induce some anticipatory buying, and may make conditions more favorable for an expansion of international investment.

In this way, a rise in spending in one country can gradually spread throughout the trading world. In the country initiating the boom, imports first rise, in response to the increase in domestic activity; and later, exports also expand, as other economies begin to share in the boom. For the latter countries, the opposite sort of lag is evident. Their exports increase first, in response to the larger purchases by the first country; then, as the rise in exports stimulates employment and incomes, imports also increase.[17] These lags between imports and exports create "gaps" in the international balance of

[17] The concept of the "foreign trade multiplier" is sometimes used in explaining this process. If we ask how much the national income will rise in response to a given increase in exports, we must take into account not only that some fraction of each increment of income will be saved but also that some part will be spent on imported goods. The larger the propensity to import, all other things being equal, the smaller will be the multiplying effect on the national income of a given increase in the export surplus or in domestic investment. But the larger will be the increase in imports and the impact on other countries; and, where the initial stimulus came from a rise in exports, the more quickly will equilibrium be restored in the balance of payments.

payments that we shall have to look at carefully a little later in this chapter.

So far, we have emphasized the direct effect of changing incomes and expenditures on the volume of imports and exports as the chief means through which business fluctuations are transmitted internationally. While this is undoubtedly the most important factor, changes in prices and in the volume of international investment also play a role in the international transmission of business cycles.

In the country first experiencing a boom, prices are likely to rise, particularly as full employment is approached. In general, if prices and costs tend to rise more in one country than in the rest of the world, imports into that country tend to increase and its exports to decline. This price-cost effect works in the same direction as the effect of rising incomes—to expand exports, and therefore income and employment, in the other countries.

International investment also plays a role in the transmission of business fluctuations. This was particularly true before the chaotic changes of the 1930's led to a drying up of the flow of international capital.[18] Increasing income and rising business activity in any country create a demand for loanable funds for investment; at the same time, confidence and the willingness to lend are likely to improve. Some of the new capital funds may come from abroad. The rise in international lending increases employment and incomes in the borrowing countries, as men are put to work on investment projects, and leads to an increased demand for capital-goods exports from the lending and other countries. Thus, a rise in foreign loans by the United States to, say, various South American countries would lead to an increase in domestic investment, incomes, and imports of consumers' goods in the latter countries, and it would also generate an increased demand for capital goods from the United States and other industrial economies.

HOW A DEPRESSION MAY BE "IMPORTED"

Suppose that a business downswing is in progress in part of the world. Other countries cannot escape feeling some effect. They soon experience a declining demand for their exports; employment

[18] For a study of the behavior of international investment in the 1920's and 1930's, see United Nations, *International Capital Movements During the Inter-War Period,* 1949.

and incomes fall in their export industries; and this deflationary influence spreads into the domestic industries. As incomes decline, so do imports; and, consequently, still other countries are drawn into the deflationary spiral.

Any country experiencing a decline in exports relative to imports is faced with a balance-of-payments problem. It can meet the situation created by its increasingly unfavorable balance of payments in several ways. If it has large reserves of gold and foreign exchange, it can utilize them to pay for excess imports and wait for the world demand for its exports to revive, in the meantime stimulating domestic activity to absorb the unemployment in the export industries. This is the ideal solution.[19] The other methods available all involve taking steps to reduce imports in line with the decline in exports, with the result that the international depression already under way is accentuated.

The other methods of adjustment available are of three sorts. The first method is to undergo a sufficient decline in domestic incomes and employment so that imports are eventually reduced enough to restore equilibrium in the balance of payments. The deflationary process, by reducing prices and costs, will also provide some stimulus to exports. This method exposes the economy to the full effects of the depression abroad and also helps to transmit it to still other countries. This is the classical method of adjustment where exchange rates remain fixed, as under the gold standard, and the government imposes no restrictions on purchases from and payments to other countries.

The second method is that of exchange depreciation. If exports fall more than imports, the external value of the country's currency can be permitted to decline. As a result, imports become more expensive in terms of the domestic currency, and exports become cheaper to foreigners. In this case, no deflation within the country may be necessary. Depreciation of the currency takes the place of internal deflation. While this method helps to protect a country against external deflationary pressures, it nonetheless adds to the depression elsewhere. This is so because of the decline in imports

[19] It is to the great credit of the United States that it has deliberately followed this ideal solution in the last decade, during which it experienced a substantial balance-of-payments deficit. In this case, the deficit was due not to a decline in American exports because of depression abroad but primarily because of the large volume of American aid and loans to other countries.

which occurs. To the extent that the currency depreciation permits one country to take away part of the export markets of another, the situation is further aggravated.

The final method of adjustment was extensively used during the 1930's and the first years after World War II. Through higher tariff duties, exchange control, import quotas, and similar types of intervention, a country can directly and immediately restrict its imports to offset a decline in exports and thus avoid the need to undergo internal deflation. It can stimulate activity at home to offset the decline in the export industries and at the same time take action to restrict its imports. It is also possible to stimulate exports through export subsidies, bilateral trading agreements, and similar devices. These are beggar-my-neighbor policies. Exports are maintained by taking markets away from other countries, and the restriction of imports also accentuates the depression in other parts of the world.

Thus it is possible, through currency depreciation and various sorts of trade controls, to insulate a country to some extent from depressions that begin elsewhere. But the method of insulation tends to make the depression more severe in other countries. In addition, the last method of adjustment described, which involves direct restrictions on the free flow of international trade, tends to reduce permanently the volume of world trade, to interfere with the international division of labor, and to impede the movement of capital to the parts of the world where it can be most productive.

RECONCILING INTERNAL AND EXTERNAL STABILITY

The preceding analysis brings into focus the dilemma faced by all countries that engage in international trade: how to maintain stable, full employment at home and, at the same time, prevent serious disequilibrium from developing in their international balance of payments.[20] There would be no problem if *all* countries were able to keep their economies stable. But if deflation begins in one important country, other countries will find themselves in a position in which their exports are falling relative to their imports. Granted that a country's reserves of gold and foreign exchange cannot stand

[20] See the useful article by Edward Marcus, "Countercyclical Weapons for the Open Economy," *Journal of Political Economy*, vol. 62, December, 1954, pp. 479–493.

a serious drain for very long, two alternatives are possible. Employment may be maintained by various means, in which case the country will have to alter the external value of its currency or directly restrict imports. In this case, internal stability is maintained at the expense of external stability. The other alternative is to undergo the necessary amount of internal deflation until current payments are again brought in line with current receipts in the balance of payments. In this case, internal stability is sacrificed in order to restore equilibrium in the balance of payments.[21]

The same dilemma faces a country that may want to expand more rapidly than the rest of the world. The accelerated expansion means a rise in imports relative to exports and hence disequilibrium in its balance of payments. Either imports must be curbed in the ways previously described or else the rate of expansion must be retarded.

THE IDEAL SOLUTION

Ideally, internal stability can be maintained in the face of deflation abroad by the liberal use of gold and foreign-exchange reserves and by the movement of short-term capital in the right direction. If a country finds a gap opening in its balance of payments because of a decline in exports, it can permit gold to be exported and the government can sell freely its holdings of foreign currencies to importers and others who must make payments abroad. If this can be done, there is no need to bring about a reduction in imports, and the government is free to follow an expansionary policy at home to offset the decline in employment in the export industries. At the same time, by maintaining its imports, the country does not add further to the deflationary forces already at work in other countries.

Under the proper conditions, short-term capital movements

[21] A similar but less difficult dilemma faces a country that wants to avoid an inflationary boom that has begun in other countries. If no steps are taken, the inflation abroad will cause a rise in the value of its exports, which will inflate incomes and prices at home; and the resulting inflow of gold and foreign-exchange reserves will make it more difficult for the monetary authorities to control the volume of credit. In addition, higher prices for imported foods and raw materials will raise costs and prices in domestic industries and exert an upward pressure on wage rates. In this case, appreciation of the external value of the currency may serve to reduce exports (by making them more expensive to foreigners) and to increase imports, which become cheaper in terms of domestic currency. Here again, external stability would be sacrificed for internal stability. If direct trade controls were already in existence, they could be relaxed; and tariffs could be reduced in order to encourage imports.

could work in the same direction. If, when a temporary gap opens in the balance of payments, short-term capital can be attracted by a moderate rise in interest rates or by a very slight decline in the exchange rate, again internal deflation and a reduction in imports can be avoided. Before the 1930's (and again in recent years), short-term capital moved freely among the major money markets of the world and was quite sensitive to small changes in interest and foreign-exchange rates.

Although this is the ideal way in which a country can adapt itself to a foreign-born deflation, the difficulties involved are obvious. Most countries cannot stand a prolonged drain on their limited reserves of gold and dollars, although the position of the European countries in this respect is now much better than it was at the end of the 1940's. Only the United States could afford a substantial loss of gold over a considerable period; and at the beginning of the 1960's, some observers were beginning to express concern over the persistent deficit in the American balance of payments.

A primary objective of the International Monetary Fund is to provide a central pool of gold and foreign currencies which member nations can draw on when faced with temporary balance-of-payments difficulties. The additional reserves thus made available have helped various countries to meet temporary balance-of-payments deficits, although the Fund's resources were completely inadequate to meet the world-wide dollar shortage that existed in the first years after World War II. Even now, although the Fund's resources have been increased and there has been a large movement of gold from the United States to the rest of the world, the Fund could be of only limited help in coping with the balance-of-payments problems that would arise in a moderately severe world-wide depression. Indeed, there is some question as to whether the monetary reserves of the free world are adequate for continued rapid expansion in the volume of world trade and for the temporary emergencies that must be expected from time to time.[22]

INTERNATIONAL MEASURES TO ACHIEVE STABILITY

Shortly after World War II, when international trading relationships were still characterized by a variety of serious maladjustments,

[22] See, for example, two articles by Robert Triffin in *Banca Nazionale del Lavoro Quarterly Review*, March and June, 1959; Bernstein, *op. cit.*, chap. 9.

a report of the United Nations listed three main requirements "for the establishment of the kind of over-all international economic equilibrium and stability which would enable countries successfully to maintain their economies at a stable and prosperous level."[23] The requirements are as follows:

1. All countries should cooperate in making the structural readjustments necessary to restore and maintain equilibrium in their balance of payments. This means taking steps to eliminate chronic deficits or surpluses in the balance of payments of each country so that international trading and financial relationships will again be in approximate equilibrium—with a minimum of government restrictions to impede the international movement of goods and capital.

2. The necessary conditions should be created for the maintenance of a stable flow of international investment.

3. Some procedure should be devised to prevent the international propagation of cyclical fluctuations and particularly to prevent the cumulative contraction of world trade resulting from a decline in aggregate demand in one country.

RE-ESTABLISHMENT OF INTERNATIONAL EQUILIBRIUM

Let us look at these requirements more closely.[24] Although much remains to be done, it is encouraging to be able to report that substantial progress has been made toward the re-establishment of a viable international economic system capable of supporting both economic stability and growth.

The first requirement listed above called for the elimination of

[23] *National and International Measures for Full Employment,* 1949, p. 49. These requirements and the recommendations for meeting them are presented on pp. 49–69 and 87–99 of this report.

[24] It is worth comparing these requirements with those put forward a few years earlier in the League of Nations volume, *Economic Stability in the Post-War World,* p. 280:

1. The adoption of liberal and dynamic commercial policies.
2. The creation of international machinery to facilitate the smooth flow of international payments.
3. The creation of an international institution to stimulate international investment and, so far as possible, to give it a contracyclical character.
4. International action for the solution of the problems associated with primary products.
5. International coordination of national programs for maintaining a high and stable level of employment.

the basic economic maladjustments from which the world economy had been suffering since the Great Depression, and which were greatly aggravated by World War II. In the early postwar period, there was a world-wide dollar shortage; and most countries were unable to generate sufficient exports, especially to the United States, to pay for the imports (and other current obligations) that would normally go with a high level of domestic employment. Imports were held down by a network of trade restrictions, and a freely convertible currency was a comparative rarity. As a result, the volume of trade was much less than it might have been; full advantage was not taken of the international division of labor; and progress toward higher standards of living throughout the world was retarded.

A considerably brighter picture could be drawn at the beginning of the 1960's. The countries that had suffered most from the war had largely rebuilt their economies. Their capacity to export had greatly increased. As the balance-of-payments position of various countries improved, trade restrictions were reduced and something close to full currency convertibility was achieved in a good many countries.[25] Total output and living standards rose rapidly in western Europe, in Japan, and in North America. A heavy flow of grants and loans poured into the underdeveloped parts of the world. An unprecedented degree of international economic cooperation manifested itself: through international financial agencies such as the International Monetary Fund and the International Bank for Reconstruction and Development, the European Economic Community (with its objective of a common free-trade market for its six member countries), the Organization of European Economic Co-operation, the General Agreement on Tariffs and Trade (GATT), and a variety of other international agencies and activities.

The substantial progress toward international equilibrium and the widespread removal of trade restrictions have contributed to the enlargement of world trade and to the rise of living standards, particularly in the less backward countries. With larger reserves of gold and foreign currencies, and with the additional help available from the various forms of international cooperation mentioned, most countries have been able to absorb temporary disturbances in

[25] Various sorts of trade restrictions still remained, although on a substantially diminished scale.

their balance of payments without either significantly reducing domestic employment or reimposing severe restrictions on the international flow of goods or capital.

Although the situation had greatly improved, various sorts of international disequilibrium were still evident. Most of the underdeveloped countries, for example, had domestic programs that were inconsistent with balance-of-payments equilibrium. Of concern to many Americans was the seemingly chronic deficit in the American balance of payments. American payments abroad (for imports, economic and military aid, and government and private foreign investment) exceeded foreign payments to the United States in every year but one during the 1950's. One result was a significant reduction in the American gold stock, as well as a large increase in the dollar holdings of other countries.[26] As the 1960's opened, fear was being expressed that continuation of this situation would make it difficult for the United States to continue to perform, on the same scale as before, the leading role that it had thus far played in the restoration of international economic stability and in the stimulation of economic growth in the free world.

A HIGH AND STABLE LEVEL OF INVESTMENT

The second prerequisite for international stability mentioned on page 655 is "the maintenance of a stable flow of international investment." A high and stable level of international investment is important for several reasons. Only the wealthier nations can supply the funds required for economic development in those parts of the world where standards of living are lowest. Secondly, a large and stable volume of lending facilitates the achievement of international equilibrium and the maintenance of a multilateral trading system. For example, American loans to underdeveloped areas in Asia, Africa, or Latin America would help to support the demand for American exports, would finance the import surplus of the borrowing countries, and would provide the latter with dollars which could be used to buy goods in Europe as well as the United States (which in turn would help European countries to finance essential imports from the United States).

[26] For a review of these developments, see Bernstein, *op. cit.;* also Joint Economic Committee, *Staff Report on Employment, Growth, and Price Levels* (86th Congress, 1st session, 1959), chap. 11.

As we noted early in this chapter, the volume of international investment has been notoriously unstable, thus tending to accentuate international booms and depressions. If arrangements could be made not only to increase but also to *stabilize* the volume of international investment, this development would be a major contribution to the achievement of international economic stability. This is a further reason why it is important to have the right sort of international investment program, which would insure that international lending would be reasonably stable from year to year.

It is highly unlikely that such stability can be achieved by relying exclusively on private lenders.[27] The United Nations report previously mentioned suggests that this be done through greatly expanding the authority of the International Bank for Reconstruction and Development. Each lending country would set an annual target for itself. To the extent that the target was not met by private and governmental loans from the country concerned, the government would make the difference available to the International Bank. The latter would use the funds to finance general development programs in countries that were short of capital and would not be restricted to specific capital projects as at present. A good deal can be said both for and against such a program. Its great virtue is its emphasis on the need not only for enlarging the flow of international investment but also for keeping it stable.[28]

There was a substantial revival of private international investment during the 1950's. Private foreign investment from the United States showed a significant increase in the latter half of the decade, and there was a rise also in the export of private capital from the European creditor countries. The larger part of American foreign investment in the postwar period has been so-called "direct investment" in American-owned facilities abroad, and there has been a tendency for such investment to be concentrated in particular areas

[27] However, there is general agreement that everything possible should be done to increase the volume of private international lending, even though supplementary government measures may be necessary to make the flow of international investment reasonably stable.

[28] Jacob Viner has gone further and proposed that the lending activities of a much enlarged International Bank be put on a deliberately contracyclical basis. The new lending institution "would be obliged to lend freely when depression was threatening and to cut off lending and to press hard for repayment when employment conditions were buoyant." "International Finance in the Postwar World," *Journal of Political Economy*, vol. 55, April, 1947, p. 106.

(e.g., Canada, western Europe, the Middle East, and Latin America) and in particular industries (notably petroleum but also manufacturing).

While the revival of private foreign investment is encouraging, the problem of instability still remains. International investment by private lenders is likely to continue to be highly volatile and is almost certain to fall off sharply in the event of an international depression. Hence the need for the largest possible amount of international cooperation, both through the International Bank and by other means. Clearly, also, the flow of private investment will be more stable the more effectively we can maintain overall economic stability, particularly in the more important industrial and creditor nations.

STABILIZING THE FLOW OF INTERNATIONAL TRADE

The third requirement listed on page 655 gets directly at the problem of preventing the international propagation of business fluctuations. The authors of the United Nations report suggest a procedure which is hardly likely to be adopted by the governments concerned, though the proposal is logical in the light of the analysis in the preceding sections of this chapter. Under the arrangement proposed, any country that reduced its imports because of a domestic deflation would deposit with the International Monetary Fund an amount of its own currency sufficient to make up the deficiency. Countries whose exports to the depressed country had declined could use their own currencies to purchase the currency deposited by the first country. In this way, they could obtain the foreign exchange necessary to maintain their imports, despite the fall in their exports to the country experiencing a depression. Thus, a depression in one important country need not lead to a cumulative contraction in international trade. This is an attempt to secure the ideal solution described on page 653.

It is difficult to believe that this proposal has much chance of acceptance, and we shall not stop to analyze it in detail. In any event, as the United Nations report emphasizes, it is at best only a supplementary measure for obtaining world-wide economic stability. The most important requirement (which is not included in the list on page 655) has not yet been mentioned. It is nothing less than international cooperation to secure a high and stable level of employ-

ment in all countries and particularly, because of its dominant position, in the United States. If wide cyclical fluctuations continue to occur in important industrial countries, they are bound to be reflected in the volume of international trade; and other countries will be forced to take protective measures to maintain internal stability.

THE NEED TO MAINTAIN DOMESTIC STABILITY

We come then to the final simple—and difficult to attain—solution to the problem of preventing the international spread of business fluctuations and of permitting countries to have both internal stability and equilibrium in their balance of payments with freely convertible currencies. The answer is to maintain full employment without inflation in every country. The most effective way to international stability within the framework of a relatively free world-trading system is the maximum amount of internal stability in each country. In the words of the United Nations report, "Measures designed to maintain international demand and thus prevent the international propagation of economic recessions must be regarded as supplementary to domestic policies for ensuring full employment and not as substitutes for them."[29] It is highly desirable that domestic stabilization programs be worked out in consultation with other countries and that they be fitted into a program for the removal of remaining structural maladjustments that obstruct the expansion of international trade and investment. Obviously, we do not want countries to maintain internal stability by means that make it more difficult for other countries to maintain a high and stable level of employment.

The role of the United States is crucial in this respect. As one writer has put it, "The greatest single contribution the United States can make to the postwar world in economic affairs is to maintain her domestic economy at levels of high income and employment."[30] The dominant position of the United States in the world economy is so great that fluctuations in this country are reflected to some extent in the balance-of-payments position of many other countries; and, despite the attempts at insulation, their domestic

[29] *National and International Measures for Full Employment*, p. 69.

[30] N. S. Buchanan, *International Investment and Domestic Welfare*, 1945, pp. 215–216.

economies are likely to be affected also. The converse is much less true. Because of the size and self-sufficiency of the American economy, an inflationary or deflationary stimulus from abroad would have only a minor effect on the level of income in the United States.[31]

While, as we saw in Chapters 15 and 16, the United States has continued to experience minor recessions since the war, the American economy has been relatively stable; and this has made the problem of maintaining stability in other countries much simpler than if the United States had experienced one or more serious depressions. Also, the mild American recessions since the early 1950's have had a smaller impact on the economies of western Europe than most economists expected.[32] A high and rising level of aggregate demand has made for rapid growth in Europe (and other advanced economies); American foreign aid and investment and American government expenditures abroad have helped to support the balance of payments of other countries; governments have, on the whole, pursued policies that have helped to maintain demand when recession threatened; and, in various ways, other countries have been able to cope with temporary balance-of-payments difficulties without either generating serious unemployment at home or reverting to discriminatory restrictions on the flow of international trade. Inflation, rather than depression, has been the form in which the problem of instability has chiefly presented itself.[33]

All of this is highly encouraging, but several cautions are in order. (1) The more the world returns to liberal trade policies, the more room there is for the international transmission of economic disturbances. As one economist has recently observed, "the economies of Europe are as vulnerable as ever to external stimuli, and suffer

[31] The American balance of payments was, however, noticeably affected by the 1958 recession in Canada and Europe. Despite the American recession of 1957–1958, American imports were maintained better than exports.

[32] For discussion of the unexpectedly mild impact of the American recessions of the 1950's on business conditions abroad, see J. H. Furth, "The United States Balance of Payments in the Recession," *Quarterly Journal of Economics*, vol. 73, May, 1959, pp. 197–206; W. Lederer, "Major Developments Affecting the United States Balance of International Payments," *Review of Economics and Statistics*, vol. 38, May, 1956, pp. 177–192; H. K. Zassenhaus, "Direct Effects of a United States Recession on Imports: Expectations and Events," *Review of Economics and Statistics*, vol. 37, August, 1955, pp. 231–255.

[33] On all this, see the useful paper by Maddison, *op. cit.*

less from them only because of a general improvement in domestic economic policy and in standards of international economic behavior."[34] (2) The forces making for rapid growth and expanding aggregate demand in both Europe and America may eventually weaken. (3) Hence, it is essential that governments not only continue but also improve further the battery of stabilization tools that they have been using. (4) International cooperation needs to be strengthened further, and the resources of the international financial agencies should be further enlarged. International cooperation played a major role in the substantial measure of international stability achieved in the 1950's. (5) The deficit in the American balance of payments will eventually have to be corrected. Already this chronic deficit makes it more difficult for the American authorities to pursue appropriate policies in a recession.[35] (6) As restrictions on the flow of capital have been removed, the danger of destabilizing movements of short-term capital have increased. The movement of "hot money" was a serious problem in the 1930's, and it was again creating difficulties for some countries in the 1950's.[36]

Thus, despite the marked improvements since the 1940's, it is still necessary to recognize ". . . that there is still a great deal of instability in the international economy, that conflicts of economic interest do frequently arise, and that public and even political opinion does not fully recognise the virtues of reciprocity or the danger of destroying it by restrictive measures, even though there is sufficient understanding and solidarity to prevent a major collapse. Government policy and international institutional arrangements will need constant improvement if a reasonable equilibrium is to be preserved."[37]

INTERNATIONAL STOCK-PILING AND THE COMMODITY RESERVE PROPOSAL

We have already noted that countries which concentrate on the production and export of primary products—that is, foodstuffs and

[34] *Ibid.*, p. 119.

[35] Thus very low interest rates (adopted as an antirecession device) would presumably lead to an outflow of short-term capital seeking higher interest rates elsewhere. This would result in a further loss of gold and might undermine confidence in the dollar. Thus, the Federal Reserve authorities might have to forego use of an easy-money policy at the time that it was most needed.

[36] This was particularly true in Europe in 1957. See OEEC, *Tenth Annual Economic Review*, 1959, pp. 60–61.

[37] Quoted from Maddison, *op. cit.*, p. 145.

raw materials—are peculiarly vulnerable to fluctuations that originate in other countries. The demand for their exports fluctuates widely over the business cycle; the supply of these products is relatively inelastic; and, as a result, prices as well as the volume of exports are extremely unstable. Consequently, even a moderate depression in the rest of the world is likely to result in a sharp decline in the value of exports from these countries and to create serious balance-of-payments difficulties. The situation is usually aggravated by the behavior of international investment. The countries that export chiefly primary products are, for the most part, the underdeveloped areas of the world; and it is toward these areas that international capital tends to flow in times of prosperity. In depressions, international lending is drastically curtailed, and this situation adds to the already serious difficulties of the raw-material countries. The latter are forced to curtail their imports sharply, so that the deflationary forces already at work in the world are reinforced.[38]

THE DESIRABILITY OF INTERNATIONAL BUFFER STOCKS

With a view to reducing this sort of instability, a number of proposals have been made for the creation of an international buffer-stock agency which would buy a selected list of primary commodities when their prices tended to fall and sell them when their prices tended to rise.[39] To be effective, such an agency would have to be international in character and would need to be liberally financed by the participating governments. The buffer-stock organization would be prepared to purchase unlimited quantities of a selected list of commodities if their prices fell to some specified minimum level and would sell its accumulated stocks freely at predetermined maximum prices. Thus, price fluctuations of these commodities would be held to a narrow range set in advance. The buffer-stock agency would accumulate stocks in times of depressed demand and reduce its stocks during boom periods.

[38] The recent difficulties of the countries relying on exports of primary products can be followed in the *World Economic Survey*. Cf. also Bernstein, *op. cit.*, chap. 6.

[39] Cf., for example, League of Nations, *Economic Stability in the Post-War World*, chap. 19; W. W. Riefler, "A Proposal for an International Buffer-Stock Agency," *Journal of Political Economy*, vol. 54, December, 1946, pp. 538–546; and the further references cited in footnote 42. For other references to some form of buffer-stock proposal, see M. K. Bennett and Associates, *International Commodity Stockpiling as an Economic Stabilizer*, 1949, Appendix Note E.

The advantages of an adequately financed and properly managed buffer-stock program are obvious. First of all, the deliberate accumulation of stocks when prices are low represents a form of investment which stimulates economic activity in the same way as any other kind of investment. But it does more than this. It helps to maintain the incomes of primary producers without requiring radical shifts in resources to other industries. It helps to prevent serious distortions in the world price structure. It also provides a partial substitute for international lending, which is ordinarily interrupted during depressions. The main part of the funds for the accumulation of buffer stocks is likely to come from the wealthier industrial nations that ordinarily export capital, and the disbursements made by the buffer-stock agency would go chiefly to the countries that normally are capital importers. Finally, the countries benefiting from the buffer-stock operations would be able to maintain their demand for goods from other countries, with the result that the stimulus would be felt in industrial countries also.[40]

THE COMMODITY RESERVE PLAN

Proposals for contracyclical buffer-stock operations have sometimes been linked with plans for monetary reform, and several writers have put forward proposals for a "Commodity Reserve Currency." In these plans, the buffer-stock agency would have the power to create money (bank credit) and would finance its purchases of commodities with newly created money. During periods of price decline, the agency's purchases of commodities not only would tend directly to stabilize prices but also would increase the economy's liquidity by adding to the supply of money. During boom periods, sales out of accumulated stocks would not only directly check the rise in prices but also cause a contraction in the supply of money.

The best-known and most thoroughly developed of these buffer-stock proposals which incorporate the "coinage principle" is that of Benjamin Graham.[41] Graham's plan for an international Commod-

[40] Cf. League of Nations, *Economic Stability in the Post-War World,* p. 269.

[41] See his *World Commodities and World Currency,* 1944. This volume applies to the international sphere the proposals for a national commodity-reserve currency put forward in his *Storage and Stability: A Modern Ever-Normal Granary,* 1937. Commodity Reserve plans involving the creation of new money have also been put forward by F. D. Graham, F. A. Hayek, and R. F. Harrod. See F. D. Granam, "Transition to a Commodity Reserve Currency," *American Economic Re-*

ity Reserve System has two essential features that are not necessary parts of an international buffer-stock plan. One is the "coinage principle." Through the International Monetary Fund, the stockpiling agency would pay for commodities with newly created deposit credits. The second feature is use of the "commodity unit" principle. The commodity unit would consist of specified amounts of a selected list of storable commodities, and only composite commodity units would be bought and sold. There would be no transactions in individual commodities.

The plan would work approximately as follows: A "commodity unit" would be set up consisting of specified amounts of standardized, nonperishable commodities which move in world trade and are traded on organized exchanges. An International Commodity Corporation would be empowered to buy whole commodity units in unlimited amounts if the price of the unit fell to a specified minimum and to sell commodity units if the price rose to a predetermined maximum. Graham has proposed rather narrow limits to permitted price fluctuations—between 95 and 105 percent of whatever is taken as the base-period price. Purchases of commodity units would be paid for with new deposit credits created by the International Monetary Fund. These deposit credits would be a new international currency which would be used in international transactions as freely as gold.

Suppose that, because of a decline in the prices of some primary commodities, the price of the composite commodity unit should fall to the specified minimum. The Commodity Corporation would begin to buy all the commodities on its list in the fixed proportions required by the commodity unit. The sellers, presumably commodity dealers, would obtain in exchange deposit credits at the International Monetary Fund. They would deposit these sums in commercial banks in their own countries, and the commercial banks would deposit them at the central bank. Thus deposits in the country would increase, and both the commercial banks and the central bank would have larger reserves.

Let us assume that the decline in prices that set this process in mo-

---

view, vol. 31, September, 1941, pp. 520–525; also his essay in A. P. Lerner and F. D. Graham, *Planning and Paying for Full Employment,* 1946; F. A. Hayek, "A Commodity Reserve Currency," *Economic Journal,* vol. 53, June–September, 1943, pp. 176–184; R. F. Harrod, *Towards a Dynamic Economics,* 1949, Appendix.

tion was part of a general business recession which had led to a decline in exports from this country and to a resulting disequilibrium in its balance of payments. Domestic purchasing power, however, would be maintained through the purchases being made by the buffer-stock agency; and imports could be kept up with the help of the additional currency reserves which would have come into the hands of the central bank. The additional currency reserves would represent a net addition to the world's money supply, increasing liquidity in some countries without involving a loss of monetary reserves by any other country. Gradually, if the country in question continued to have a current deficit in its balance of payments because of a lower volume of exports, it would have to transfer its deposits at the International Monetary Fund to other countries having a surplus in their balance of payments. But this would not necessitate the sort of monetary deflation that might have been required if the buffer-stock purchases had not provided the first country with additional currency reserves. And the alternatives of exchange depreciation or import restrictions, in order to avoid internal deflation, would also presumably not be necessary.

One of the most controversial aspects of the Graham proposal is the method of financing commodity purchases through the creation of new deposit credits. It has been strongly suggested that it would be preferable to have the Commodity Corporation finance its purchases out of capital funds and loans floated in the international financial centers.[42] The result would be essentially the same, except that the funds accruing to the countries from which commodities were purchased would come out of the monetary reserves of the countries that subscribed to the capital and loans of the Commodity Corporation. Graham's suggestion would involve a net monetary expansion in the world as a whole during depressed periods when buffer-stock purchases were being made. Although, as we have seen, monetary expansion cannot alone cure a depression, the additional liquidity thereby provided would reinforce the more important and direct effects of buffer-stock operations.

[42] Cf. M. K. Bennett and Associates, *op. cit.*, chaps. 3–4. This volume presents a very useful and exhaustive evaluation of the Graham proposal. See also the more recent analysis by E. M. Harmon, *Commodity Reserve Currency*, 1959, which also refers to the similar proposal made by Jan Goudriaan.