



# 15 Years of Research on US Employment and the Minimum Wage

Paul Wolfson<sup>1</sup> — Dale Belman<sup>2</sup>

*Abstract.* Statistical analysis of the minimum wage and employment has been very active for the last quarter century, including more than 37 studies of US data since the December 2000 AER exchange involving Card, Krueger, Neumark and Wascher. In this meta-analysis of the 37 that report results suitable for this technique, the most important finding is a considerable shift toward the origin in the ‘consensus range’: from the interval  $[-0.3, -0.1]$  to  $[-0.13, -0.07]$ . The minimum wage has negative employment effects, but these have become notably smaller and are largely localized to teenagers.

## 1. Introduction

The employment effects of the minimum wage have long been a controversial topic among economists. Conventional theory clearly predicts that a higher minimum wage goes hand in hand with lower employment, although not the strength of the relationship. Following Brown, Gilroy and Kohen (1982) close review of the econometric literature on its employment effects, the interval  $[-0.3, -0.1]$  has become the widely cited consensus range of the employment elasticity of the minimum wage.<sup>2</sup>

Reinvigorated by the 1991 ILR-Cornell conference on the minimum wage and a series of increases in Federal, State and Local minimum wages, research on the effects of the minimum wage has blossomed over the last quarter century. More than 800 scholarly articles in English have been published on some aspect of the minimum wage in the years since the conference, more than 600 since the exchange between Neumark and Wascher (2000) and Card and Krueger (2000) at the end of 2000. A substantial plurality considers the effect of the minimum wage on some aspect of employment.

Both the quantity and variation in estimates pose a challenge not only for researchers but also for policy makers. Summarizing the findings is not straightforward. There is no immediate answer to the often-asked question, ‘What is the effect of the minimum wage on employment?’ One approach to answering this question has been to choose estimates which the reviewer finds particularly convincing because of methodology, data, authorship,

---

We thank Tim Bartik for the initial suggestion to conduct this analysis. It is dedicated to the memory of Alan Krueger whose work contributed so much to the origin and continuation of the literature that we analyze here.

<sup>1</sup>Tuck School of Business, Dartmouth College, Hanover, NH, USA.

<sup>2</sup>School of Human Resources & Labor Relations, Michigan State University, East Lansing, MI, 48824 USA.

E-mail: dalebelman@outlook.com

and, perhaps (or, perhaps, especially), the results. Although picking and choosing the ‘best’ studies has a long history in reviews of many literatures, it is particularly unsatisfactory here if only because of the tone of many of the discussions of estimates of the employment effect.<sup>3</sup> Too often the choice of ‘best’ appears to coincide with the reviewer’s views on the minimum wage. Meta-analysis can be a more disciplined, transparent, and reproducible method of obtaining a common estimate of the effect of the minimum wage across studies.<sup>4</sup>

Building on the meta-regression work of Doucouliagos and Stanley (2009, DS hereafter), and Stanley and Doucouliagos (2012, SD hereafter), this paper attempts an objective approach, applying meta-regression to estimate the employment elasticity of the minimum wage for the 739 estimates for the US from the 37 analyses in the first 15 years of the 21st century that reported commensurable estimates. With meta-regression, a formal statistical technique, it is possible to systematically address conditions such as heteroskedasticity, heterogeneity, non-independence, and selection bias that would otherwise preclude deriving reliable meta-estimates of the employment elasticity.

Overall, we find that the range of the employment elasticity has shifted toward zero since Brown *et al.* (1982), from  $[-0.3, -0.1]$  to  $[-0.13, -0.07]$ , but, in contrast to prior meta-analyses Doucouliagos and Stanley (2009), Belman and Wolfson (2014), our estimates are negative and statistically significant, albeit of small magnitude. Teenagers, and eating and drinking establishments together account for more than half of the estimates in our sample. Estimating separate models for teens and for eating and drinking places has little effect on our estimated range, moving it from  $[-0.13, -0.10]$  to  $[-0.11, -0.07]$ .<sup>5</sup> The minimum wage then has negative employment effects, but estimates of them have become smaller and are largely localized to teenagers, who comprise a declining share of the labor force. The declining magnitude of the negative effect lessens the constraints on minimum wage policies.

## 2. Meta-analysis and meta-regression

Meta-analysis involves the application of statistical techniques to a collection of research results, both to combine them and to explore patterns, especially problematic ones: not only publication bias but also systematic differences.<sup>6,7</sup> As with many techniques of inferential statistics, the starting point is (i) a random sample in which (ii) the observations are independently and identically distributed. Failure of one or more of these assumptions is common in applications of meta-analysis: in particular, the assumptions that the sample is random and that the observations are distributed identically. Failures of these assumptions are often addressed under the labels of heteroskedasticity, heterogeneity, and selection bias (in particular publication bias).

In the simplest situation (i.e. a random sample with *iid* observations) meta-analysis can involve just an unweighted mean. However, the identical distribution assumption immediately fails to hold unless the estimates each derive from samples of the same size; otherwise the estimates will be heteroskedastic. Consequently, the simplest commonly used specification of meta-analysis starts from the assumption that the only important difference among the studies under consideration is the specific random sample used; in all other ways, the studies may be considered homogeneous. The meta-analysis estimate is then a weighted average of individual estimates where the weights reflect the relative sizes of the samples.

In both the simplest case and the simplest commonly used case, meta-analysis is primarily a technique for combining estimates to generate a single one with a smaller standard error than each of the individual studies.<sup>8</sup>

A step up in complexity allows for arbitrary heterogeneity beyond the heteroskedasticity addressed above: that is, a more extensive failure of the assumption of identical distribution. It is based on two assumptions. First, differences exist in the mean value of each  $y_i$ , and second, these differing mean values are themselves drawn from a common distribution.<sup>9</sup> The heterogeneity can be due to differences in the study design, details of the policy examined by different studies, and so forth. In any case, the heterogeneity is not modeled parametrically.<sup>10</sup>

Meta-regression, the technique which the present analysis primarily relies on, models the differences parametrically:  $y_i \sim N(x_i\beta, \sigma_i^2 + \tau^2)$ .<sup>11</sup> Typically, the  $X$  variables are indicator variables that describe characteristics of individual studies and of individual estimates within studies that report more than one. Meta-regression not only allows for a straightforward way to address non-identical distribution of observations but also for the breakdown of the assumptions of independence and random sampling. For instance, questions may be raised concerning the independence of estimates from the same study or from studies by the same researcher, or of estimates that rely on the same data source (for instance the CPS). A simple solution available to meta-regression is the use of dummy variables for the study, researcher or data source, respectively.

### 2.1. Publication bias

Assuming that the meta-analyst has followed proper procedure in constructing the sample, publication bias may result when a set of estimates is not representative of a random sample of all possible estimates from the underlying population. Good research may not see the light of day because of decisions at some stage of research or publishing.<sup>12,13</sup> Absent appropriate corrective techniques, the effect estimated by meta-regression will necessarily be a biased estimate of the true effect size. Publication bias is a form of the well-understood issue of sample selection; the typical solution for sample selection requires access to a sample that includes observations (i.e. studies) which has not been filtered by selection.<sup>14</sup> This solution is not available in this instance because it is not possible to assemble information on the sample of unpublished studies.

Building on work on meta-regression in medical research, Hedges and Vevea (2005) draw on several authors to discuss a family of corrections for publication bias. In this approach, observations are weighted based on the inverse probability of publication, and this probability depends positively on the size of the effect and negatively on either the p-value or the standard error of the estimated effect.

Publication bias in economics literature may not behave similarly to that in medical literature. For instance, there has been reason to suspect that studies of the employment consequences would be less likely to be published if they did not report a statistically significant, negative effect (Card and Krueger, 1995a,b; Doucouliagos and Stanley, 2009). In the years since their initial study reporting no statistically significant minimum wage effect (Card and Krueger, 1994) and the subsequent uproar, several studies have appeared reporting similar results. It appears that publication bias in this literature is not the same as it once may have been; at the least, it is no longer as severe for studies that do not report adverse employment consequences.<sup>15</sup> A more flexible approach than that suggested by DS is needed.

## 2.2. *Meta-analysis as literature review*

So far, we have treated meta-regression as part of formal or inferential statistics. An alternative is to treat it as part of descriptive statistics, to interpret its results as descriptive of a research literature rather than as deriving *the best* measure of effect size. In this approach, it can complement or substitute for a verbal survey of the literature. The research on employment and the minimum wage that Brown *et al.* (1982) reviewed was both considerably more homogeneous and notably smaller than is currently the case. It was then not unreasonable to assess the pooled effect from examination and discussion of tables describing the studies and their estimates. Meta-regression, and meta-analysis more generally, makes it possible to concisely describe the much larger and more diverse current body of research and to do this in a disciplined and transparent process. Correcting for publication bias can demonstrate what the results in the literature would look like if it were not the outcome of censoring by either editors or researchers. Controlling for heterogeneity, for choices made in the course of doing the research, can guide future research by indicating how those choices have shaped the results.

## 2.3. *Graphical meta-analysis*

A funnel plot is the primary graphical tool of meta-analysis. It is a scatterplot of two variables central to meta-analysis: the effect central to a specific meta-analysis (in the present case the minimum wage elasticity of employment), and the estimated standard error or its reciprocal (known as precision).<sup>16</sup> The name comes from the shape of the plot in a well-behaved sample of results, like an upright or inverted funnel, depending, respectively, on whether the  $y$ -axis is the standard error or precision. In a well-behaved sample, precise estimates (with small standard errors) are typically near the mean or median values of the effect size and the range of estimates around the central tendency increases as precision falls.

As with other statistical graphs, funnel plots are useful for becoming familiar with the data and for identifying errors, especially important since the data for meta-analyses are often collected and recorded manually as part of the study; that is, transcription errors are more common in meta-analyses than in the original studies that are the source of their data. However, the most common function of funnel plots, once errors have been recognized and corrected, is to display any hints of publication bias. Sampling error implies that the results from a random sample of analyses will be symmetrically distributed about the mean. A large deviation from symmetry in a funnel plot is a strong hint that publication bias has affected the sample.

# 3. Applying meta-analysis to the minimum wage literature

## 3.1. *Independence*

The minimum wage literature presents a more complicated problem to the meta-analyst than a typical medical study. Consider first the problem of defining the sample. Often, a study will report not just one or even a few estimates of what is ostensibly the same parameter but many that are based on the use of different subsamples or estimation techniques. This has long been the norm in the literature on the employment effect of the

minimum wage literature; Table A1, which lists the studies used in this analysis, shows that the number of estimates from each analysis ranges between 1 and 86. An unsettled issue in meta-analysis is how to respond to this variation.

Retaining more than one estimate per study in the sample raises several questions. First is the likelihood that estimates from the same study are not independent if only because of legitimate decisions that the analyst(s) have made in the course of the research. Another is that unless an appropriate adjustment is made, the study reporting the most estimates will effectively have more influence on the result of the meta-analysis than others.<sup>17</sup> Including only one estimate per study begs other questions, beginning with which estimate to include. If ‘the best’, how is that to be identified? Perhaps the single estimate included in the meta-analysis should instead be the average (or some weighted average) for each study. However, there are presumably traits that differentiate estimates even within the same study, so choosing ‘the best’ estimate sacrifices information (or variation) up front by not including it while using the average neglects variation by blurring differences.<sup>18</sup>

Not only may estimates from the same study not be mutually independent, so too may those from different studies. Card and Krueger’s (1994) creation of data to analyze is unusual; many minimum wage studies share the same data, most often the CPS but the QCEW and CES are also often used. We might consider systematic differences in the estimates of the minimum wage elasticity of employment for teenagers versus eating and drinking establishments as the result of heterogeneity. Are systematic differences across data sources due to heterogeneity between the data sources or to lack of independence within them?

What both examples make clear is that at least in this literature, the *iid* assumption of independence and homogeneity, does not comprise two distinct conditions each of which can fail separately. The difference between them is blurry. Meta-regression can address these in the same way without having to make too sharp a distinction. It is common to think that the solution for heterogeneity is to include the appropriate regressors in the estimation, and that the solution for dependence comes from appropriate specification of the error variance–covariance matrix. A simpler if less efficient alternative to specifying the variance–covariance matrix is to use appropriate fixed effects (this also sidesteps the often-problematic assumption that the dependence is orthogonal to the heterogeneity). That is, dummy variables can be used equally well for either condition.

### 3.2. Modeling (uncategorized) failures of the *iid* assumption

Sampling distribution theory does not begin to capture the heterogeneity across estimates due to different decisions. The profusion of methods, datasets, sectors of the labor market, and periods under study, each of which may account for differences in findings, make this an especially challenging issue for meta-studies of the minimum wage. The *iid* assumption may also fail in this literature for a variety of reasons in addition to the dependency and/or heterogeneity that result from the sample’s containing multiple estimates per study and per dataset.<sup>19</sup> For each potential cause of failure, indicator variables that identify estimates with a common feature can control for the failure of the *iid* assumption. All of these elements contribute to failures in the *iid* assumption, both independence and homogeneity, and ideally, should be modeled to understand the source of different estimates of the effect of interest.<sup>20</sup>

One (latent) factor that has received some attention is variation in analysis quality (Neumark, 2014, 2015; Neumark and Wascher, 1998,2007; SD, section 6.2.2). To the extent that

all estimates in a single analysis similarly reflect that study's quality, indicator variables for each study can account for it.

### 3.3. *Publication bias*

Inspired by Heckman (1979), SD develop an approach for addressing publication bias which uses the estimated standard error or its square as a proxy for the inverse Mills ratio.<sup>21</sup> We take this approach, as our starting point for addressing publication bias in the minimum wage literature. It is not sufficiently flexible as the probability of publication can vary only monotonically with the standard error. We have instead used a third order polynomial in the standard error, which allows the likelihood of publication to vary non-monotonically, and, following suggestions of SD, we included interactions of each term with the dummy variables used to describe the estimates in our specification search (discussed below).<sup>22</sup>

## 4. Data

We began our search by entering the phrase 'minimum wage' into Google Scholar and several electronic databases of published articles or working papers: ISI-Web of Science, Econlit and the NBER. We limited ourselves to analyses of US data that were published after 2000, either as articles in journals or as working papers.

The beginning of the *New Minimum Wage Research* can be dated to a 1991 conference at Cornell University, and the exchange between Neumark and Wascher (2000) and Card and Krueger (2000) in the December 2000 issue of the *American Economic Review* marks the end of its first period. Our study includes analyses of US data that have appeared after December 2000. We identified 60 analyses, working papers, and published articles that both satisfied these criteria and included at least one estimate of the effect of the minimum wage on employment. In 37 of these 60, either the analysis explicitly reported one or more elasticities and their standard errors or it was possible to calculate them from the estimates of a semi-log equation in which the dependent variable was a binary indicator of employment status, in combination with other data included in the analysis. The number of estimates per analysis varied greatly.<sup>23</sup>

The information that we recorded to control for heterogeneity included:

- (1) The analysis containing the estimate, and the researchers involved
- (2) Data frequency (e.g. quarterly)
- (3) Geographic reach of the analysis (e.g. national versus a treatment group in a single city)
- (4) Whether the estimate is for hours or for either employment state or number of jobs
- (5) Whether the estimate has been published in a peer reviewed journal (i.e. not a working paper)
- (6) Whether the analysis is a quasi-experiment, with distinct treatment and control groups
- (7) Whether reason exists for worrying about the accuracy of standard errors (à la Bertrand *et al.*, 2004, or Donald and Lang, 2007)
- (8) Whether the estimate refers to members of demographic groups, industrial sectors or a combination of the two

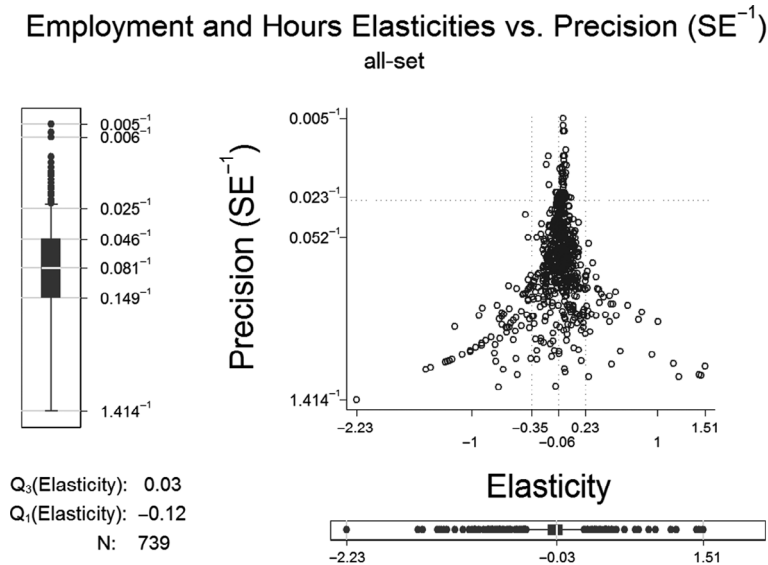
- (9) Whether the data for the estimate are ultimately derived from establishment surveys, household surveys, or both
- (10) The target of the estimate (e.g. teenagers, Eating and Drinking Establishments)
- (11) The primary dataset used (CPS, QCEW, QWI, or other)

**5. Funnel plots, descriptive statistics, and simple estimates**

Figure 1 displays the data in a funnel plot and two marginal box plots. Elasticity is displayed on the horizontal axis, with the tick marks indicating the minimum and maximum values in the data,  $-2.23$  and  $1.51$ : the unweighted mean,  $-0.06$ : one standard deviation either side of the mean,  $-0.35$  and  $0.23$ , and elasticities of  $\pm 1$ . The vertical dashed lines indicate the mean and the mean  $\pm$  one standard deviation. Precision,  $SE^{-1}$ , is shown on the vertical axis, which is drawn to a logarithmic scale. The tick marks, labeled not with the precision values which have no obvious meaning but with the SE value from which they are calculated, are of the same unitless scale as elasticity. The most precisely estimated elasticity in the data, at the top, has a standard error of  $0.005$  and the least precise has a standard error of  $1.414$ . The mean value of precision is  $0.052^{-1} = 19.1$  and one standard deviation above mean precision is  $0.023^{-1} = 43$  (one standard deviation below the mean is negative, so is not shown). The horizontal dashed line indicates the 90th percentile of precision.

The horizontal boxplot at the bottom right of Figure 1 displays the median and extreme values of elasticity; the other two members of the five-number summary, not easily shown on this graph, are listed to the left, along with the sample size. The median elasticity,  $-0.03$ , is slightly larger in absolute value than the mean, indicating perhaps a very slight skew to the left. The vertical boxplot on the upper left displays the distribution of

**Figure 1.** Funnel plot of all estimates in the sample: all-set

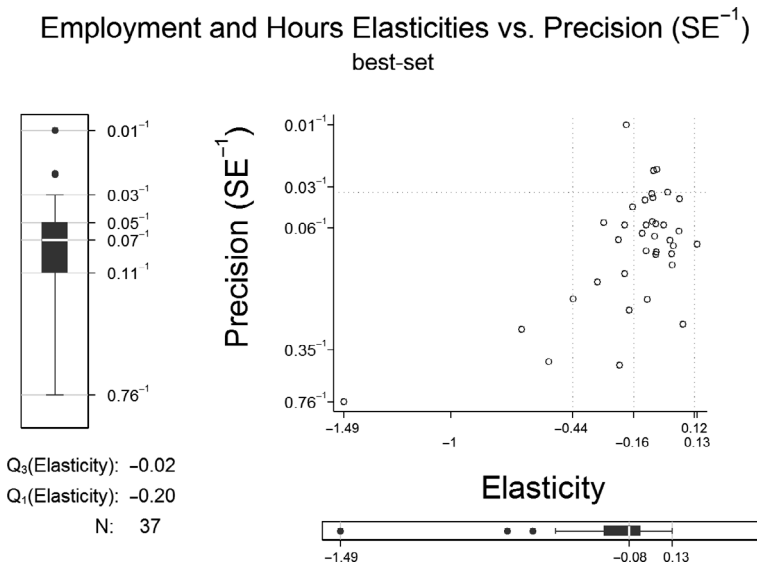


precision, with tick marks labeled by its reciprocal, SE. The five-number summary is  $(1.414^{-1}, 0.149^{-1}, 0.081^{-1}, 0.046^{-1}, 0.006^{-1})$ , and the 90<sup>th</sup> and 99<sup>th</sup> percentiles are  $0.025^{-1}$  and  $0.006^{-1}$ . The funnel plot is not symmetric about its mean, with substantially fewer estimates on the lower right than the lower left, suggesting some publication bias. However, the funnel plot is not truncated (certainly not severely so — perhaps moderately pruned), so publication bias is likely not severe.

Figures 2 and 3 are funnel plots of the best-set and average-set, respectively. Neither much suggests a funnel, although each is symptomatic of publication bias because (left) skewed or (right) truncated. Both characteristics suggest that sampling error and differences in sample size are not the primary causes of differences in estimates within each sample. At  $-0.16$ , the mean elasticity of the best-set is considerably to the left of the all-set mean,  $-0.06$ , although the distance between the medians,  $-0.08$  and  $-0.03$ , is half as great. The mean and median of the average-set,  $-0.09$  and  $-0.05$  respectively, are each considerably closer to the corresponding value of the all-set.

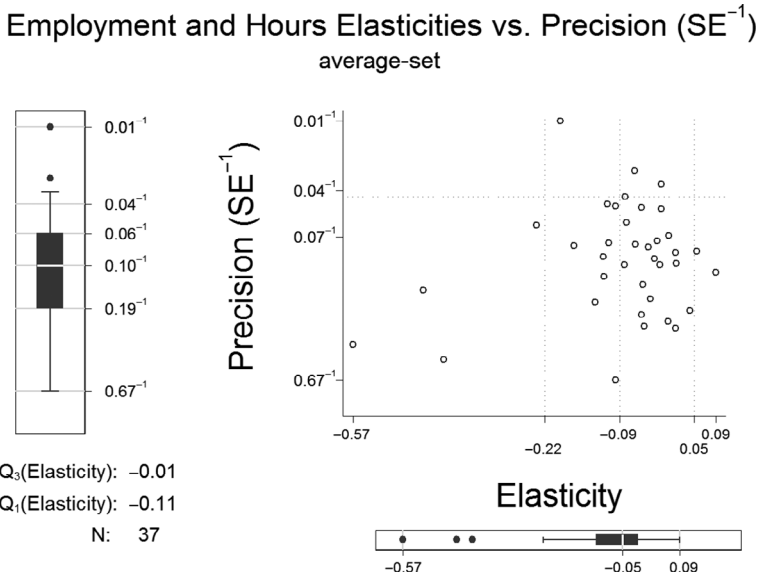
Table 1 contains additional information about each sample. The first two rows, number of observations and mean elasticity, repeat information from the funnel plots. The precision-weighted mean elasticities (row 3), are much smaller than the unweighted mean for the all-set and best-set samples; but more than one-third larger for the average-set sample. For all the means, both weighted and unweighted, the sample variation is small enough that one would feel conventionally confident in concluding that the reason each differs from zero is not due to sampling error; that is, each is easily statistically significant by any conventional measure. However, all three samples are skewed (row 4), the all-set moderately skewed and the other two highly skewed.<sup>24</sup> Furthermore, in each, the elasticity is negatively correlated with the size of its standard error (row 5), especially in the best-set, suggesting that estimates substantially greater than the mean are less likely to be published than others, that is, suggesting publication bias.<sup>25</sup>

**Figure 2.** Funnel plot of ‘Best’ estimate from each study: best-set





**Figure 3.** Funnel Plot of the mean of estimates from each study: avg-set



Rows 6 through 8 of Table 1 report the weighted means for the homogeneous and heterogeneous treatment effects meta-analysis models applied to the best-set and average-set samples.<sup>26</sup> At issue is whether it is reasonable to believe that these cross study elasticity estimates are the result of random draws from the same population, and are *iid* or whether the elasticities are drawn from different distributions and are not *iid* (are heterogeneous). Recall that the *Homogeneous Treatment Effects* model incorporates the assumption of heteroskedastic errors and a common mean in its estimate; the *Heterogeneous Treatment Effects* model dispenses with the assumption of a common mean (i.e., allows for heterogeneity). Cochran’s Q statistic is used to determine whether it is possible to reject the key assumption of the *Homogeneous Treatment Effects* model, that the different estimates are heteroskedastically distributed about a common mean. For both the best-set and the average-set samples, the p-values strongly indicate rejection. There is strong evidence of heterogeneity and a need to control for additional factors.

Rows 9 through 12 of Table 1 provide estimates of the precision-weighted elasticity with corrections for publication bias. There is an estimate for each of our samples, and for all-set sample there is a second estimate when dummy variables to control for the presence of multiple estimates from a single article are included.<sup>27,28</sup> Control for publication bias is provided by inclusion of a third degree polynomial in the standard error. Turning first to publication bias, the 95 per cent CIs indicate do not span 0, indicating there is significant publication bias for each estimate; it is negative for the all-set and average-set samples, but is positive for the best-set estimates. Although the estimates of the elasticities are smaller for the all-set and average-set estimates, controlling for publication bias in the best-set sample results in the estimated elasticity being more negative than the best-set precision-weighted mean. Except for the best-set estimate, the regression estimates of the mean elasticity are closer to zero than the simpler precision-weighted means. In addition, the estimates range from the middle of the consensus range that Brown *et al.* (1982) defined (the

**Table 1.** Estimates of simple models of the employment elasticity

	<b>all-set (Whole Sample)</b>	<b>best-set</b>	<b>avg-set</b>
1. Nobs	739	37	37
2. Mean Elasticity	-0.063 (0.011)	-0.162 (0.046)	-0.086 (0.022)
3. Precision-Wtd. Mean Elasticity	-0.024 (0.002)	-0.115 (0.015)	-0.12 (0.014)
4. Skewness (Elasticity)	-0.799 (0.090)	-3.202 (0.388)	-2.02 (0.388)
5. Correlation (Elasticity, SE)	-0.222 (0.036)	-0.810 (0.099)	-0.567 (0.135)
6. Homogeneous Treatment Effects 95% CI		-0.115 (-0.128, -0101)	-0.024 (-0.027, -0.022)
7. Cochran's Q p-Val		158 (3.30e-17)	1,109 1.60E -209
8. Heterogeneous Treatment Effects 95% CI		-0.088 (-0.125, -0.052)	-0.057 (-0.073, -0.040)

**Weighted regression estimates (Using third degree polynomial in SE to correct for publication bias)**

9. Elasticity 95% CI	-0.014 (-0.033, 0.005)	-0.048 (-0.080, -0.016)	-0.196 (-0.286, -0.107)	0.001 (-0.007, 0.009)
10. Mean Publication Bias 95% CI	-0.010 (-0.011, -0.010)	-0.015 (-0.016, -0.014)	0.050 (0.047, 0.054)	-0.010 (-0.013, -0.007)
11. Includes Deviation-Coded Analysis Fixed Effects	No	Yes	No	No
12. Q p(Q)	2,488 8.0E-188		159 4.1E-17	95.6 2.68E-07

Standard Errors in parentheses (Clustered by title for the 'Whole Sample' regression estimates, and heteroskedasticity-robust for the other regression estimates).

best-set), to considerably closer to zero (the all-set) to virtually zero (the average-set). Finally, in all three cases, Cochran's Q statistic strongly indicates that heterogeneity must be addressed in each sample.

## 6. Exploring heterogeneity

In gathering the data for this study, we coded for 14 different dimensions of heterogeneity, where a dimension refers to a particular type of difference in the estimated elasticity, e.g., the data frequency, whether the elasticity is for a measure based on hours of employment or an employment measure, etc. Within each dimension, there are two or more categories, coded by indicator variables. SD recommend initially including all the indicator variables in the regression and deleting variables with the smallest magnitude t-statistic, one-by-one, until the t-statistic of all remaining variables exceed some minimal size. One difficulty with this procedure is that the degree of multicollinearity in this set of variables

is too much for all to be simultaneously included as regressors. This suggests examining the sets of variables within each dimension beforehand to determine which capture heterogeneity in either the elasticity or publication bias.

We have relegated a detailed presentation of this examination to Appendix S1, listing only the highlights here. The points below are based primarily on regressions that include both dummies for the article from which the estimate is drawn and a simple correction for publication bias, run on subsamples defined by the specific heterogeneity indicator variable.<sup>29</sup>

- (1) Data with an annual frequency or longer generates meta-estimates that are larger (farther from zero in a negative direction) than those based on higher frequency data.
- (2) Meta-estimates for the subsample defined by the use of a measure of employment are statistically significant (those for the subsample using hours of employment are not).
- (3) Estimates that rely only on establishment data generate small, statistically insignificant meta-estimates of the elasticity. Estimates that rely on demographic data or a mix of both establishment and demographic data generate meta-estimates that are larger and statistically significant.
- (4) The meta-estimates for the following subsamples are larger than those for their complement, and are statistically significant (those for the complements are not)
  - (a) not ‘Eating & Drinking Establishments’
  - (b) not ‘Retail (not E&D)’
  - (c) teenagers.

Meta-estimates for the subsample that considers only women are about one-third as large as for the complementary set, and not statistically significant.

- (5) Meta-estimates for the subsample that considers only men are positive and statistically significant while those for the complement are negative and significant.
- (6) Meta-estimates for the subsample without a high school diploma are much larger (negative) than for the complementary subsample and are statistically significant.

Estimates for teenagers comprise over 40 per cent of the sample. The dummy variables for several of these other dimensions are highly correlated with the teenage dummy, so these results likely contain less information than at first appears.<sup>30</sup>

## 7. Prelude to meta-estimates: variable selection

To control for heterogeneity, SD and DS recommend beginning with all the variables coded to capture heterogeneity and then, one-by-one, deleting variables that do not satisfy conventional standards of statistical significance<sup>31</sup> However, the variables coded to capture the heterogeneity in the minimum wage are very multi-collinear.<sup>32</sup> It is not possible to include all simultaneously in a regression, and the choice of which to exclude initially is necessarily arbitrary and may bias the outcome. Trying a sufficiently large (random) sample of these variables to feel confident that this has not occurred would be a mammoth undertaking.

The LASSO (Tibshirani, 1996) is a more defensible and disciplined method for selecting variables to include in the regression equation. Belloni *et al.* (2014) describe both a *naive* approach and an approach that is more reasonable for policy analysis (the double LASSO) because it avoids attributing causal relationships to the policy variable which more plausibly belong to latent variables. Attempts with both approaches were ultimately unsuccessful but in a way that is informative about the state of the minimum wage research: we return to this later.<sup>33</sup> Our variable selection method was to begin with the variables suggested by the explorations in heterogeneity discussed in the last section, and then, following SD, remove the variable with the largest p-value until all had p-values no larger than 0.05.

## 8. Meta-estimates

Table 2 contains meta-estimates using explanatory variables based purely on the exploration of heterogeneity. The two columns on the left are from meta-regressions that incorporated study fixed effects, the two on the right from meta-regressions that excluded them. All estimates incorporate third degree polynomial correction for publication bias. Within each pair of columns, the left one includes all the dimensions of heterogeneity suggested by the exploration, the right one includes only those that remained after a one-by-one elimination of the variable that had the t-statistic which was smallest in absolute value, until all the remaining terms were associated with a p-value no larger than 0.05.<sup>34</sup>

Before turning to a detailed examination of the numbers in Table 2, consider first some large-scale patterns. The specification with the full set of heterogeneity terms (columns 1 and 3) leads to meta-estimates of the overall elasticity that are generally about 50–70 per cent larger in magnitude than the corresponding ones from a reduced set of heterogeneity terms (columns 2 and 4), and with considerably larger 95 per cent CIs. The estimates of publication bias are small but the 95 per cent CIs for those from regressions that include study fixed effects do not include zero, whereas those without them span the origin. Point estimates from specifications that included study fixed effects were, by and large, shifted to the left, away from zero, relative to the corresponding estimates from specifications that lacked them. This is the case not only for the estimates of publication bias and overall elasticity at the top of the table but also for coefficients of six of the nine heterogeneity terms. The exceptions are SourceData, Females, and NoHS (i.e. the elasticity for adult men who did not complete high school). The rest of the discussion will be about the estimates from the pared-down specification with study fixed effects, column 2.

Overall, the precision-weighted mean elasticity is about  $-0.08$ . For teens, eating and drinking establishments, and retail (not E&D), the point estimate of the elasticity is about 20 per cent to 50 per cent greater in magnitude than  $-0.08$ , but all four point estimates are well within the 95 per cent CIs of the other three, and very close to the lower end of the range for teenagers as defined by Brown *et al.* (1982), even when not actually within the range. There is some evidence of minor publication bias, and this likely contributes to the difference between these figures and theirs; the correction for the effect of publication bias shifts the point estimate to the left by a statistically significant but very small 0.013. Using low-frequency data results in a point estimate,  $-0.072$ , that is slightly closer to zero, not farther away, as one would expect if much of the effect of the minimum wage is in the longer run. Whether employment is measured by hours on the one hand or jobs or the employment ratio on the other appears to have almost no effect on the employment

**Table 2.** Elasticity meta-estimates based on explorations of heterogeneity

	With study fixed effects		Without study fixed effects	
	(1)	..(2)	(3)	(4)
Elasticity	-0.128	-0.082	-0.112	-0.068
95% CI	(-0.301, 0.046)	(-0.140, -0.023)	(-0.308, 0.085)	(-0.125, -0.012)
Mean publication bias	-0.011	-0.0130	0.0022	-0.0001
95% CI	(-0.012, -0.010)	(-0.0138, -0.0120)	(0.0019, 0.0024)	(-0.0004, 0.0002)
<b>Heterogeneity terms</b>				
Elasticity for <sup>a</sup>				
Teens	-0.171	-0.125	-0.123	-0.080
(N = 300)	(-0.373, 0.031)	(-0.217, -0.033)	(-0.299, 0.087)	(-0.139, -0.020)
Retail (not E&D)	-0.143	-0.099	-0.090	
(N = 70)	(-0.331, 0.044)	(-0.180, -0.018)	(-0.287, 0.107)	
E&D	-0.155	-0.100	-0.106	-0.057
(N = 188)	(-0.347, 0.038)	(-0.192, -0.007)	(-0.299, 0.087)	(-0.114, -0.0002)
Low freq. data	-0.118	-0.072	-0.113	-0.088
(N = 26)	(-0.288, 0.051)	(-0.126, -0.019)	(-0.31, 0.084)	(-0.150, -0.025)
SourceData	-0.102		-0.112	-0.064
(N = 343)	(-0.280, 0.076)		(-0.312, 0.088)	(-0.131, .002)
Males	-0.106		-0.053	
(N = 25)	(-0.279, 0.066)		(-0.256, 0.15)	
Females	-0.144	-0.089	-0.178	-0.103
(N = 34)	(-0.320, 0.032)	(-0.155, -0.024)	(-0.39, 0.034)	(-0.207, .001)
NoHS	-0.185		-0.188	
(N = 29)	(-0.502, 0.132)		(-0.568, 0.191)	
Employment	-0.130	-0.080	-0.117	-0.093
(N = 586)	(-0.302, 0.041)	(-0.138, -0.023)	(-0.322, 0.089)	(-0.158, -0.028)

Source: 1 if the original data is based on establishments or firms.

Notes: Lofreq = 1 if Frequency is Annual or Less

Confidence Intervals (CIs), in parentheses, based on standard errors that are calculated after clustering on each article.

Bold indicates statistically significant at  $\alpha = 0.05$

<sup>a</sup> The point estimates below are estimates of the elasticity associated with each heterogeneity term, not the deviation from the elasticity point estimates at the top of the column

elasticity; the point estimate of  $-0.080$  for this variable scarcely differs from the overall value of  $-0.082$ .

Finally, how important to the estimates is pooling them all and trying to control for well-defined heterogeneity rather than subsetting the sample? That is, suppose we distinguish the two major subsets that together comprise two-thirds of the sample: teens (308 estimates, 42 per cent of the sample) and Eating and Drinking Establishments (188 and 25 per cent, respectively). We proceed by first estimating a *Homogeneous Treatment Effects* model meta-estimate for each of our samples, conducting a Q test for heterogeneity and if homogeneity is rejected, we then estimate a *Heterogeneous Treatment Effects* model. The next step is to estimate meta-regressions with a control for publication bias for the E&D and Teens samples. The results of this procedure are in Table 3.

For both Teenagers and E&D establishments, the estimate of the elasticity from the *Homogeneous Treatment Effects* model is close to zero –  $-0.035$  for teens,  $-0.012$  for E&D

**Table 3.** Elasticity meta-estimates for teen and E&D subsamples

	Teens	E&D Establishments
Homogeneous Treatment Effects		
Elasticity	-0.035	-0.012
95% CI	(-0.039, -0.031)	(-0.016, -0.009)
Cochrane's Q (df)	1,174 (299)	251 (187)
p-val	0.000	0.001
Heterogeneous treatment effects		
Elasticity	-0.057	-0.018
95% CI	(-0.069, -0.046)	(-0.023, -0.012)
Meta-regression <sup>a</sup>		
Elasticity	-0.110	-0.065
95% CI	(-0.199, -0.021)	(-0.113, -0.017)
Mean publication bias	-0.016	-0.015
95% CI	(-0.018, -0.014)	(-0.017, -0.013)

<sup>a</sup>Estimates based on precision weights and a regression equation that includes study fixed effects and a third degree polynomial in *se* to capture publication bias.

Confidence Intervals (CIs) based on standard errors that are calculated after clustering on each article.

establishments – but the 95 per cent confidence interval is strictly to the left of the origin. For both, Cochrane's Q test rejects homogeneity. For both, the estimate from the *Heterogeneous Treatment Effects* model is slightly more negative, -0.057 and -0.018, respectively, and once again the 95 per cent confidence intervals are strictly to the left of the origin. In neither of these models are controls in place for either publication bias or lack of independence across estimates from the same analysis. For those, we turn to the next set of estimates, from meta-regression. The mean values of publication bias are only slightly larger than the one in column 2, Table 2, but both lie outside that estimate's 95 per cent confidence interval in Table 2. Finally, the point estimates of the employment elasticities when considering each subsample on its own (i.e. Table 3) is smaller than the corresponding figure in Table 2 (for teens, -0.11 here versus -0.125 in Table 2: for E&D Establishments, -0.065 here versus -0.10 in Table 2). However, all these values in both tables fall well within the 95 per cent confidence interval of the corresponding value in the other table. Stepping back a bit to gain some perspective, the different approaches for calculating the elasticities generate noticeable but small differences in the point estimates, and either way, the elasticities for teenagers and for E&D establishments are small but statistically significant at conventional levels.

Previously, we mentioned our ultimately unsuccessful attempts to use the LASSO. It is worth describing them briefly so that our interpretation is comprehensible. In the naïve LASSO, the variable selection procedure is run only once, on the dependent variable (i.e. the estimated employment elasticity). When we did this, after requiring the inclusion of study fixed effects to control for the lack of independence across estimates within the same sample, the naïve LASSO indicated that no other variables belonged in the equation. When we further required the inclusion of the variables in the second column of Table 2 and ran the double LASSO on these variables, the result was the same: no additional variables.

What can be made of this result; that in the presence of article fixed effect dummy variables, additional dimensions of heterogeneity (gender, the frequency of the data, and whether the study considers teen or E&D Establishments) do not improve the predictive power of the metaregression equations? Although the single LASSO may omit confounding

variables correlated with the treatment, the double LASSO guards against this.<sup>35</sup> The list of regressors in the equation defined by the double LASSO is the response variable as a function of the treatment variable and all the variables that appear in either set of LASSO results. The results from the single and double LASSOs indicate that the article dummies absorb all the inter-article sources of heterogeneity, making further control for heterogeneity unnecessary. This strongly suggests that the articles which constitute the current minimum wage research are heterogeneous in dimensions that are not easily captured by explicit measures of heterogeneity of the sort we have defined.

## 9. Conclusion

We have identified 60 analyses of US data that have been completed since the exchange between Card and Krueger, and Neumark and Wascher in the December 2000 issue of the *AER*. From the 37 of these that either report elasticities of employment with respect to the minimum wage and their standard errors or provide sufficient information to calculate them, we have gathered 739 estimates. Although publication bias has a detectable effect on the magnitude of the point meta-estimates, inducing a left shift, away from zero, the estimated standard errors of the meta-estimates are hardly, if at all, affected. Furthermore, the magnitude of the shift in the point estimates is small once these standard errors are taken into consideration.

Building on the best judgment of Brown *et al.* (1982), the consensus range for the elasticity of employment with respect to the minimum wage has been from  $-0.3$  to  $-0.1$ . It has proven durable and has been widely cited in the literature.<sup>36</sup> Although the method of obtaining this range reflected best practice in its day, it is the product of non-systematic judgments about article quality, methods, and data. It also reflects a generation of research and labor market conditions which are now nearly forty years in the past.

As we have detailed in this paper, a main advantage of meta-regression is its averaging of estimates across a potentially large body of research in a fashion which is transparent and reproducible, and which explicitly allows for the issues of heteroskedasticity, publication bias, and heterogeneity across these studies. The improvement in precision achieved by bringing large numbers of estimates together better distinguishes effects which are not significant from those which are very small in magnitude.

This latter strength speaks directly to our findings. Our meta-estimate of the overall elasticity is  $-0.082$ : for Teens the figure is  $-0.125$ ; for retail (not E&D),  $-0.099$ ; for E&D Establishments,  $-0.10$ ; and for women,  $-0.089$ . Incorporating the separately estimated values for Teens and for Eating and Drinking Establishments (Table 3), the range is extended from  $-0.13$  to  $-0.07$ . What was the smallest effect in the Brown *et. al.* range,  $-0.10$ , is now the midpoint of a much narrower range. Estimation using the single and double LASSO show that, with inclusion of article fixed effects, further control for inter-article heterogeneity does not improve the fit of the meta-regression.

Why might the minimum wage have lesser effects on employment? Given the controls for heterogeneity, it is unlikely that it is due to the evolution of methods or a change in the focus of studies. One (speculative) possibility is that the labor market has become much less important to the lives of teenagers and teenagers have become much less important to the functioning of the labor market over the last 15 years.<sup>37</sup> Furthermore, as reported in Table 3, the current teen elasticity is close to the bottom of the conventional range. On the

other hand, over the last 25 years the importance of Eating and Drinking Establishments to the labor market has grown, with their share of employment increasing by at least 25 per cent (both for production and non-supervisory workers, and for all employees).<sup>38</sup> Such shifts in the structure of labor markets toward employment that is less sensitive to changes in the minimum wage have likely played a central role in the shift of the range in minimum wage employment effects.

Contrary to our earlier analysis (chapter 4 of Belman and Wolfson, 2014) which reported no significant minimum wage effects for a smaller sample of studies of US data, here we find a negative relationship between the minimum wage and employment but one that is closer to zero than Brown *et al.* reported more than 35 years ago. This is also in contrast with an earlier meta-analysis of Doucouliagos and Stanley (2009) that reported meta-estimates of  $-0.01$  with a range from  $-0.003$  to  $0.065$ .

## Notes

<sup>1</sup>Brown *et al.*'s (1982) conclusions were more nuanced than reflected in later literature. The  $-0.1$  to  $-0.3$  summarized 'typical' results for teenagers from time series models. The typical range for cross sectional research was 0 to  $-0.75$ ; smaller effects typified studies of young adults. Nevertheless, authors as prominent as David Neumark and William Wascher refer to the  $-0.1$  to  $-0.3$  range.

<sup>2</sup>Leonard (2000) points to derogatory personal comments referring to Card and Krueger's (1994) results by prominent economists who had not themselves conducted minimum wage research. In their 2000 exchange, Card and Krueger (2000) and Neumark and Wascher (2000) cast much more decorous aspersions on each other's work.

<sup>3</sup>Meta-regression has been criticized as a mechanistic approach to understand the minimum wage literature. A narrative review of the literature has the advantage of providing insights into the strengths and weaknesses of each article, but it does not provide a measure of the effect or empirical estimates of how various decisions, such as studying teens rather than the broader labor force or use of difference in difference methods affects empirical measures. See *Minimum Wages*, (2008, Neumark and Wascher) and *What Does the Minimum Wage Do?* (2014, Belman & Wolfson) for the recent detailed narrative reviews of the literature

<sup>4</sup>Separate estimates for teenagers slightly but non-insignificantly moves the top of the range down to  $-0.11$  while separate estimates for E&D establishments shifts the lower end to  $-0.07$ .

<sup>5</sup>This section draws heavily on Harbord and Higgins (2008), Hedges and Vevea (2005) and SD.

<sup>6</sup>Many techniques are well established and widely used for meta-analyses in the medical and physical sciences, and the Stanford School of Medicine has a research center, METRICS, devoted to their development and dissemination <http://metrics.stanford.edu/> on January 17, 2016.

<sup>7</sup>A formal statement of this model is that the estimates,  $\{y_i\}$ , are assumed to be independent and distributed according to  $y_i \sim N(\theta, \sigma_i^2)$ , where the variances differ only because of differences in the size of the sample used for each estimate. Obviously, a distribution other than the normal can be used.

<sup>8</sup>In the conventional meta-analytic framework, the model that allows for heteroskedasticity is known as the *Fixed Effects Model* or *Common Effects Model*, whereas the second model, which allows for more extensive failure of the *iid* assumption is known as the *Random Effects Model*. Because these terms, *Fixed Effects* and *Random Effects*, have such well-defined and widely understood meanings in panel econometrics, using them here would almost certainly lead to confusion. Instead we will refer to *Homogeneous Treatment Effects* for the Common or Fixed Effects Model (which adjusts only for differences in the sample size of each study), and *Heterogeneous Treatment Effects* for the Random Effects Model.

<sup>9</sup>Formally, the  $\{y_i\}$  are independently distributed according to  $y_i \sim N(\theta_i, \sigma_i^2)$  and the  $\{\theta_i\}$  are in turn independently distributed according to  $\theta_i \sim N(\theta, \tau^2)$ .



<sup>10</sup>SD is an extended presentation of meta-regression.

<sup>11</sup>This suspicion is so widely mentioned that specific citations to it are beside the point. Instead, google ‘publication bias null hypothesis’ or ‘publication bias statistical significance’, or read the Wikipedia article ‘Publication Bias’ (December 26, 2018).

<sup>12</sup>It may be on the part of researchers who decide not to engage in certain types of research or submit results for publication, or on the part of editors who decide not to publish an analysis, not due to its quality, but because its results contradict theoretical priors or, are not statistically significant or thought to be sufficiently interesting. The set of estimates available for meta-analysis are then not a random sample of all possible estimates.

<sup>13</sup>In a conventional selection problem, some information is available on those who are not selected, and more is available on those who are. In this situation, the standard solution is Heckman’s (1979), which purges the bias from the point estimate.

<sup>14</sup>Neumark and Wascher (1998) argue that what is interpreted as publication bias in time series may be due to structural change.

<sup>15</sup>Occasionally, the  $y$ -axis displays the degrees of freedom, the sample size, or its square root.

<sup>16</sup>Considering the passions surrounding the employment effect of the minimum wage, the phrase ‘stuffing the ballot box’ comes to mind.

<sup>17</sup>Section 2.4.4 of SD is a concise discussion of questions about multiple estimates per study.

<sup>18</sup>The section below labeled Data includes a detailed list of differences across studies in our sample.

<sup>19</sup>Tests for heterogeneity exist, the most widely recognized being Cochran’s  $Q$  statistic which is distributed  $\chi^2$  under the null hypothesis of homogeneity.

<sup>20</sup>The rationalization is that researchers engage in p-hacking of some sort to generate an acceptable result and this is most likely to be successful with small samples (which are associated with larger standard errors). For an extensive but informal discussion of p-hacking, see the Wikipedia entry on Data Dredging: [https://en.wikipedia.org/wiki/Data\\_dredging](https://en.wikipedia.org/wiki/Data_dredging) (26 December 2018).

<sup>21</sup>Hypothesizing that p-hacking would result in p-values just less than 0.05, we also examined the effect of including a dummy variable that indicated p-values between 0.04 and 0.05. The p-value of the coefficient of that dummy was consistently and remarkably large, so we report no results along these lines.

<sup>22</sup>Appendix S1 lists these studies and the number of estimates garnered from each. The mean number of estimates in a study is 20 (s.d. = 20.6) and the five-number summary is {1, 6, 14, 25, 86}. The total number of estimates in our study is 739. We excluded estimates that were replications of prior work and included in the analysis to provide a basis of comparison.

<sup>23</sup>Bulmer (1979, p. 63) suggests that when the absolute value of the skewness statistic is less than 0.5, the distribution is reasonably symmetric; when it is between 0.5 and 1, the distribution is moderately skewed, and when it is greater than one, the distribution is highly skewed.

<sup>24</sup>In the all-set and best-set samples, the correlations with respect to the square and cube of the standard errors are very similar to those shown (relevant because of our later use of a 3rd degree polynomial in the *SE to address publication bias*). For the average-set, the correlations decline by 15–20 per cent with each increase in the exponent.

<sup>25</sup>We do not estimate these models for the all-set sample because of concern with dependence across estimates within the same sample.

<sup>26</sup>Throughout our analysis, all dummy variables are deviation coded. When there is only one set of dummies, e.g., sex, the constant term of the regression is then the simple mean of the dummy for each sex. When more than one set of dummies is coded, e.g., both sex and teen/not-teen, the constant term is the simple mean of the four dummies for female teens, male teens, female adults, and male adults. For more information about deviation-coded dummies, see <https://stats.idre.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/#DEVIATION> (retrieved 20 May 2019).

<sup>27</sup>The three analyses that each contain only one estimate are grouped together to make up the excluded dummy.

<sup>28</sup>Statistical significance here refers to a test size of 0.05.

<sup>29</sup>For example, only five of the nine principal components for this set of variables have eigenvalues greater than or equal to 1, and regressing the teenager dummy on the other dummies generates  $R^2$  and adjusted- $R^2$  values of 0.50. Including study-specific fixed effects in this regression raises the  $R^2$  to 0.87 and the adjusted- $R^2$  to 0.86.

<sup>30</sup>This procedure is justified with reference to Hendry's General-to-Specific framework; the method was developed to capture the salient characteristics of an underlying data-generation process (DGP) in a reasonably sparse specification. It typically involves a variety of specification and residual tests at each step to ensure model validity (Hoover and Perez 1999). It is not obvious that thinking in terms of DGPs is useful in meta-analysis where the purpose is to describe or characterize a body of research.

<sup>31</sup>In a principal component analysis of the 16 dummy variables included to capture heterogeneity (Lofreq, National, Multi-State, State, City, employment, Published, Quasi-experiment, reliable SEs, Source Data, E&D, Retail, Teens, NoHS degree, Females only, Males only), only the first seven have eigenvalues greater than or equal to 1, and together they explain 74% of the variance. When 34 dummies for the studies are also included, only the first 33 components have eigenvalues no smaller than 1 (jointly accounting for 95% of the variance), the 5 smallest eigenvalues are less than 0.01 and the two smallest are less than 5E-8.

<sup>32</sup>We use a version of the LASSO technique that account for clustering of the observations (in our case, by the analysis from which each was derived): Belloni *et al.* (2015). We thank Christian Hansen for making available his Stata program for this procedure.

<sup>33</sup>This search procedure used regressions with deviation-coded dummy variables, so dimensions of heterogeneity were excluded when they ceased having any statistically significant information relative to the mean elasticity, not when the elasticity relevant to the corresponding dimension of heterogeneity was not statistically significant. Thus, in column 2, elasticity estimates based on source data (and for males) do not appear because they do not differ by a statistically significant amount from the overall mean (the elasticity at the top of the column) rather than because they do not differ by a statistically significant amount from zero.

<sup>34</sup>The double LASSO is the methodology preferred by Belloni *et al.* (2014)

<sup>35</sup>More recently, Brown (1999) concluded that the range has both shrunk and moved closer to zero.

<sup>36</sup>According to CPS data, from 2000 to 2015 the teen employment rate fell from about 0.45 (roughly its mean since the second half of the 20<sup>th</sup> century) to about 0.28. Corresponding figures for the labor force participation rate are 0.52 and 0.34. During this period, the teen share of employment fell from 5.3% to 3.2% and the teen share of the labor force fell from 5.8% to 3.6%.

<sup>37</sup>According to CES data, from 1990 to 2015, the share of Food Services and Drinking Places (NAICS 722) in employment has risen from 8.0% to 9.9% for Production and Non-Supervisory Employees, and from 7.2% to 9.2% for All Employees.

## References

- Belloni A., Chernozhukov V. and Hansen C. (2014) 'High-Dimensional Methods and Inference on Structural and Treatment Effects', *Journal of Economic Perspectives* 28(2): 29–50.
- Belloni A., Chernozhukov V., Hansen C. and Kozbur D. (2015) 'Inference in High Dimensional Panel Models with an Application to Gun Control', *WP* Oct: 6).
- Belman D. and Wolfson P. (2014) *What Does the Minimum Wage Do?*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Bertrand M., Duflo E. and Mullainathan S. (2004) 'How Much Should We Trust Differences-in-Differences Estimates?', *Quarterly Journal of Economics* 119(1): 249–275.
- Brown C. (1999) 'Minimum Wages, Employment, and the Distribution of Income' in Ashenfelter Orley and Card David (eds.) *Handbook of Labor Economics*, Vol. 3B. New York, NY: Elsevier.
- Brown C., Gilroy C. and Kohen A. (1982) 'The Effect of the Minimum Wage on Employment and Unemployment', *Journal of Economic Literature* 20(2): 487–528.

- Bulmer M. G. (1979) *Principles of Statistics* NYC. NY: Dover Publications.
- Card D. and Krueger A. (1994) 'Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania', *The American Economic Review* 84(4): 772–793.
- Card D. and Krueger A. (1995a) 'Time-Series Minimum-Wage Studies: A Meta-analysis', *The American Economic Review* 85(2): 238–243.
- Card D. and Krueger A. (1995b) *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton, NJ: Princeton University Press.
- Card D. and Krueger A. (2000) 'Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply', *American Economic Review* 90(5): 1397–1420.
- Donald S. and Lang K. (2007) 'Inference with Difference-in-Differences and Other Panel Data', *Review of Economics and Statistics* 89(2): 221–233.
- Doucouliafos H. and Stanley T. (2009) 'Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis', *British Journal of Industrial Relations* 47(2): 406–428.
- Harbord R. and Higgins J. (2008) 'Meta-regression in Stata', *The Stata Journal* 8(4): 493–519.
- Heckman J. (1979) 'Sample Selection Bias as a Specification Error', *Econometrica* 47: 153–161.
- Hedges L. and Vevea J. (2005) 'Selection Method Approaches' in Rothstein H., Sutton A. and Borenstein M. (eds.) *Publication Bias in Meta-analysis – Prevention, Assessment and Adjustments*. Chichester, England: John Wiley and Sons Ltd.
- Hoover K. D. and Perez S. J. (1999) 'Data mining reconsidered: encompassing and the general-to-specific approach to specification search', *Econometrics Journal* 2: 167–191.
- Leonard T. C. (2000) 'The very idea of applying economics: The modern minimum-wage controversy and its antecedents', *History of Political Economy* 32(Suppl\_1): 117–144.
- Neumark D. (2014) "Employment effects of minimum wages". *IZA*. World of Labor (May).
- Neumark D. (2015) 'The Effects of Minimum Wages on Employment', *FRBSF Economic Letter* 2015–2037.
- Neumark D. and Wascher W. (1998) 'Is the Time-Series Evidence on Minimum Wage Effects Contaminated By Publication Bias?', *Economic Inquiry* 36: 3.
- Neumark D. and Wascher W. (2000) 'Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania: Comment', *American Economic Review* 90(5): 1362–1396.
- Neumark D. and Wascher W. (2007) 'Minimum Wages and Employment', *Foundations and Trends in Microeconomics* 3(1–2): 1–182.
- Neumark D. and Wascher W. (2008) *Minimum Wages*. Boston: MIT Press.
- Stanley T. and Doucouliagos H. (2012) *Meta-Regression Analysis in Economics and Business*. New York: Routledge.
- Tibshirani R. (1996) 'Regression shrinkage and selection via the lasso', *Journal of the Royal. Statistical Society B* 58(1): 267–288.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

**Appendix S1** Exploring heterogeneity.

**Table S1** Studies used.

**Table S2** Exploring heterogeneity in the whole sample (part 1).

**Table S3** Exploring heterogeneity in the whole sample (part 2).

**Table S4** Exploring heterogeneity in the best-set sample (part 1).

**Table S5** Exploring heterogeneity in the best-set sample (part 2).