Revisiting the Risk of Automation

Melanie Arntz[1,2], Terry Gregory[1] and Ulrich Zierahn[1,*]

[1]ZEW Mannheim

[2]University of Heidelberg

Economics Letters, Volume 159, October 2017

In light of rapid advances in the fields of Artificial Intelligence (AI) and robotics, many scientists discuss the potentials of new technologies to substitute for human labour. Fuelling the economic debate, various empirical assessments suggest that up to half of all jobs in western industrialized countries are at risk of automation in the next 10 to 20 years. This paper demonstrates that these scenarios are overestimating the share of automatable jobs by neglecting the substantial heterogeneity of tasks within occupations as well as the adaptability of jobs in the digital transformation. To demonstrate this, we use detailed task data and show that, when taking into accounting the spectrum of tasks within occupations, the automation risk of US jobs drops, ceteris paribus, from 38 % to 9 %.

JEL Classification: J23, J24, O33

Keywords: Artificial Intelligence, Robotics, Automation, Jobs

* corresponding author, zierahn@zew.de, ZEW Centre for European Economic Research, L7,1, 68161 Mannheim, Germany

Introduction

The prevailing view in academic research is that technological advances in the 19th and 20th century did not result in net job losses (Autor 2015). Nevertheless, a new wave of technological anxiety floods the public and academic debate, fearing that technology will displace jobs on a large scale in the near future. The reasons put forward are rapid advances in the fields of robotics and artificial intelligence (AI) (c.f. Brynjolfsson and McAfee 2014). These arguments receive increasing attention in the economic debate (e.g. Acemoglu/Restrepo 2017; Nordhaus 2015; Pratt 2015) and, e.g., have been discussed in the special session on *Economics and Artificial Intelligence* at the 2017 ASSA meetings.

Fueling this debate, various risk assessments predict that up to half of the workforce is at risk of automation in the next two decades, thus provoking political demands such as a robot tax or basic income. As a key drawback, these assessments neglect the general finding that tasks vary substantially within occupations (Autor and Handel 2013) and adjust to computerization (Spitz-Oener 2006). As a prominent example, Frey and Osborne (2017) develop automation scenarios by applying assessments from robotics and AI experts on the automatibility of occupations to the corresponding employment shares in the US economy. As a result, book-keeping, accounting and auditing clerks are assigned a 98% probability of being automated in the near future, irrespective of the task variation across workplaces within this profession. However, according to our task data, many workers in such highly exposed occupations also perform tasks that machines struggle with, such as problem solving or influencing. We demonstrate that neglecting this variation leads to an overestimation of the overall risk of automation in the economy.

Our analysis exploits detailed information on the task content of jobs. After linking expert-based assessments regarding the occupational risk of automation to individual job tasks and characteristics, we then re-estimate the automation potential in two scenarios: In a first scenario, we replicate the results of former studies with our data assuming homogenous tasks within occupations. In a second scenario, we allow for differences in the tasks across workplaces within occupations. Overall, we find that the automation risk of US jobs drops from 38 % to 9 % when allowing for workplace heterogeneity. Occupation-level assessments of automation potentials thus are severely upward-biased.

1 The occupation-level vs. job-level approach

Our approach relates experts' assessments of occupation-specific automation risks as collected by Frey and Osborne (2017) to individual job characteristics. The underlying idea is to assume that the experts' assessment regarding the potential for automating a particular occupation contains valid information, but captures only a representative occupation that masks important differences at the level of workplaces within its profession. Hence, we merge the information on occupation-specific automation risks to workers in that particular occupation for whom the PIACC database also includes individual-level information on socio-economic and job-related characteristics. We then estimate the following fractional response model: $a_o = \sum_{k=1}^{K} \beta_k x_{ki} + \epsilon_i$, where a_o represents the automation risk of individual i's occupation o as provided by Frey and Osborne (2017) and where x_{ki} represents the K job characteristics of worker i including 25 tasks as well as variables to capture that the actual content of a specific task may vary across individuals and firms depending on firm-size, gender, education or competences.[1] The coefficients β_k measures the influence of the K job characteristics on the occupation-specific automation potential, which we then use to predict automation potentials for each individual job i, $\hat{a}_i = \sum_{k=1}^{K} \widehat{\beta_k} x_{ki}$. Note that by projecting occupation-level risks of automation on job-level characteristics, our approach allows for within-occupational heterogeneity in the predicted automation potential as jobs may differ in K job characteristics.

Since the original automation potential a_o is measured at the occupational level it does not contain the job-level-variation in the exposure to automation which implies a measurement error in the dependent variable. An additional measurement error in this variable results from the discrepancy between the Standard Occupational Classification (SOC) for which the occupation-specific risk assessments are available and the PIACC data that contains the ISCO classification. We rely on the Bureau of Labor Statistics' correspondence table to transfer the SOC-level automation potentials to the ISCO classification. As a consequence, j multiple values of the automation potential a_{oj} are assigned to each individual i in the PIAAC data, whenever there is no one-to-one assignment between both classifications. We use the Expectation-Maximization (EM) algorithm proposed by Ibrahim (1990) for the estimation of

[1] In accordance with the literature, we assume that the technological potential for automating task k is constant across workplaces.

the fractional response model described above in order to cope with these measurement errors. Essentially, the method estimates the weights for each individual's multiple values a_{oj} that maximize the likelihood of being the "correct" automation risk, given workers' characteristics. This yields a weighted occupational risk of automation that varies across individuals given their characteristics. Hence, the imputation method addresses both the assignment problem due to the mismatched occupational classifications and, at the same time, yields individual level automation risks. See Ibrahim et al. (2005) for applications.

Based on the coefficient estimates, we predict the automation risks at the job level $\hat{a}_i$ (job-level approach). For comparison, we additionally estimate occupation-level automation potentials by keeping tasks constant within occupations to replicate former studies. For this, we calculate the occupation-level median of our variables $\tilde{x}_{ko}$, predict the occupation-specific automation potential $\widehat{a_o} = \sum_{k=1}^{K} \widehat{\beta_k} \tilde{x}_{ko}$ and apply these risks to all workers in occupation o.

2 Re-assessing the risk of automation

Descriptive statistics for the explanatory variables as well as average marginal effects resulting from the fractional response model described in Section 2 are shown in the Appendix. Overall, we find that the automation potential is lower in jobs that require programming, presenting, training or influencing others. In contrast, the risk of automation is higher in jobs with a high share of tasks that are related to exchanging information, selling or using fingers and hands. This resembles the evidence from the task-based literature which argues that routine tasks are subject to automation, whereas interactive or cognitive tasks are less likely to be substituted by machines and computers (see Acemoglu and Autor 2011).

Figure 1: Employment shares by automation risk in the US: occupation-level vs. job-level approach

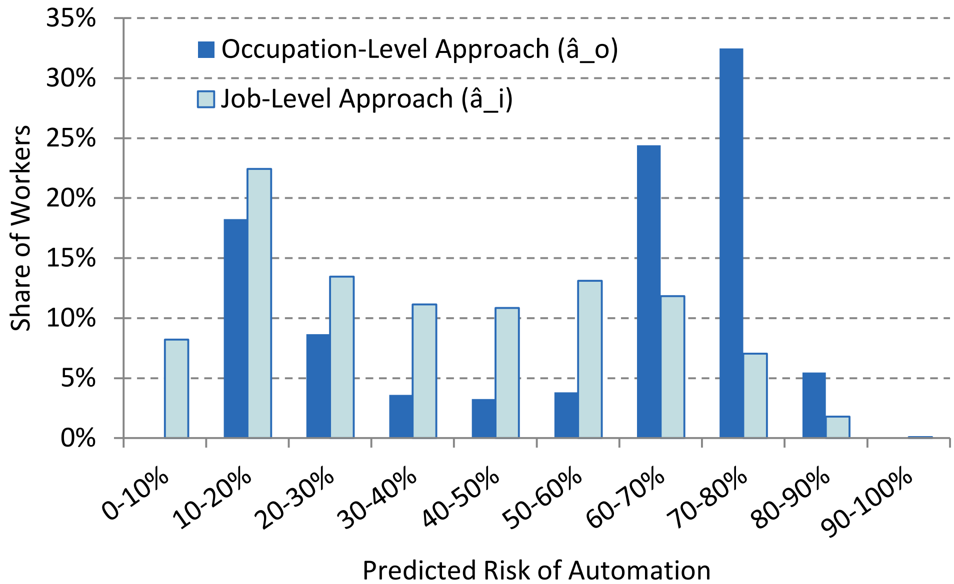
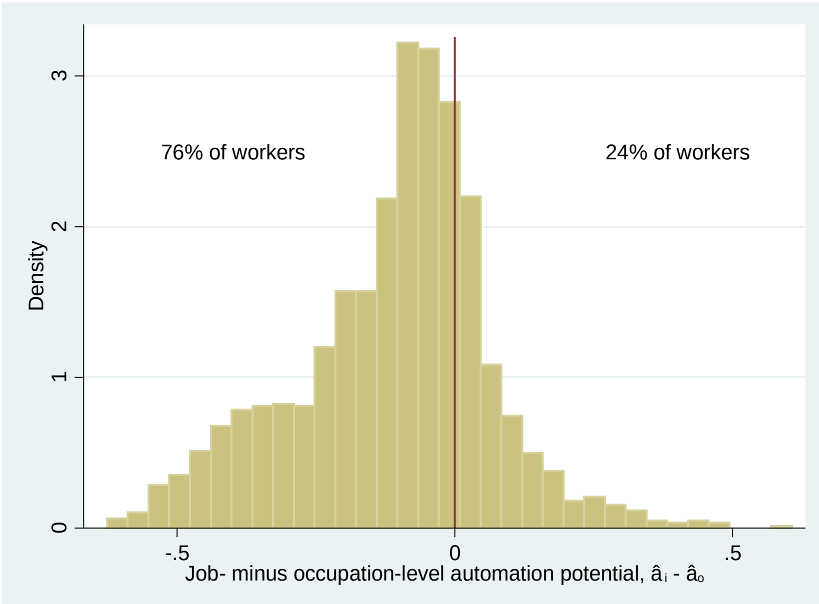


Figure 1 compares employment shares by decile of the predicted automation risk distribution in the US between the occupation-level vs. job-level approach. The results for the occupation-based approach resemble the bi-polar structure from previous risk assessments, i.e. the majority of jobs are assigned either a very high or a very low risk of automation. Accordingly, 38 % of the workers perform jobs with a risk of automation above 70 %. In contrast, when using the job-level approach, i.e. predicting the risk of automation at the level of individual jobs $\hat{a}_i$, we find only a moderate polarization in automation probabilities with most jobs being exposed to a medium risk of automation. Hence, only 9 % of all workers in the US face a risk of automation that exceeds 70 %.

Figure 2: Difference between job- and occupation-level automation potential: Distribution across workers in the US



Evidently, our job-level approach suggests a much lower share of jobs at risk of automation compared to previous occupation-level assessments, as the job- is below the occupation-level estimate for 76% of all workers (Figure 2). This is counterintuitive: as both predictions are computed using the same model, parameter estimates and data, one would expect the two estimates to be similar. The *only* difference between both approaches is that we use occupation-level median task values in the occupation-level approach, whereas we use the individual-level task information in the job-level approach. The difference between the job- and occupation-level results therefore is solely driven by the fact that the majority of jobs involve non-automatable tasks more often compared to the occupational median job, as workers of the same occupation specialize in different non-automatable tasks. As an example, consider Numerical and Material Recording Clerks (ISCO08=43), for which our occupation-level estimates suggest a high (74.4%) risk of automation. According to the data, many clerks of this profession specialise in niches that involve non-automatable tasks such as presenting, planning or problem solving. Taking the large and heterogeneous range of their tasks into account suggests that only 18.2% of them actually face a high risk of automation. Put differently, the average worker does a job that is much less automatable than the median job in this occupation. Technically speaking, the contribution of each task to an individuals' automation potential, $\hat{\mu}_{ki} = \hat{\beta}_k x_{ki}$, is negatively skewed within occupations for 72% of all tasks and there is no systematic correlation between the tasks' contributions to automation potentials ($corr(\hat{\mu}_{ki}, \hat{\mu}_{k'i})$). As a consequence, the occupational median job underestimates the relevance of many specialized jobs with less automatable tasks, thus leading to an upward-bias of automation risks.

3 Conclusion

Our study reveals a serious and systematic upward bias in occupation-level estimates of automation potentials compared to a job-level approach, as workers specialize in non-automatable niches within their profession. One potential reason for our results could be that workers increasingly focus on a diverse set of tasks that complement these technologies, as would be consistent with the evidence by Spitz-Oener (2006). This implies that the exposure to automation should be measured at the level of jobs rather than occupations. When doing so, still one in ten jobs is highly exposed. While some of these workers may only need to adjust their tasks, others might actually lose their jobs. Yet, whether this leads to net job losses depends on the relative sizes of job-creation and job-destruction effects (Acemoglu and Restrepo 2017, Gregory et al. 2016).

Acknowledgements

The authors gratefully acknowledge helpful comments by Holger Bonin, Stephan Dlugosz, Bernd Fitzenberger, Christina Gathmann, Anna Salomons, Susanne Steffes, Arne Warnke and two anonymous referees as well as financial support by the OECD. The usual disclaimer applies. An earlier version of this paper is available as OECD Social, Employment and Migration Working Papers No. 189.

Appendix

Table 1: Determinants of automation potentials in the US

	Variable		Model Results AME	Std. Err.	p	Descriptive Statistics mean or share	sd
worker characteristics	gender (0: male, 1: female)		-0,011	0,004	0,006	0,472	0,499
	age group	1: 16-19				3,1%	
		2: 20-24	0,006	0,012	0,613	9,0%	
		3: 25-29	0,051	0,012	0,000	11,7%	
		4: 30-34	0,008	0,012	0,475	11,7%	
		5: 35-39	0,016	0,012	0,166	10,9%	
		6: 40-44	0,031	0,012	0,008	10,7%	
		7: 45-49	0,029	0,012	0,017	11,6%	
		8: 50-54	0,021	0,012	0,081	12,4%	
		9: 55-59	0,024	0,013	0,059	9,4%	
		10: 60-65	0,060	0,013	0,000	9,5%	
	education	1: low (ISCED 0, 1, 2)				8,0%	
		2: medium (ISCED 3, 4, 5B)	-0,087	0,012	0,000	57,6%	
		3: high (ISCED 5A, 6)	-0,165	0,013	0,000	34,4%	
skills	literacy (mean score)		0,000	0,000	0,000	277,447	44,096
	numeracy (mean score)		0,000	0,000	0,030	263,608	50,170
	problem solving (mean score)		0,000	0,000	0,003	279,890	39,473
job characteristics	sector (0: private; 1: public, non-profit)		-0,053	0,004	0,000	0,255	0,436
	firm size	1: 1-10				20,5%	
		2: 11-1000	0,032	0,005	0,000	68,0%	
		3: >1000	-0,002	0,007	0,794	11,5%	
	responsibility for staff (0: yes; 1: no)		0,043	0,004	0,000	0,670	0,470
	educational job requirements (0: ISCED0-4; 1: ISCED 5-6)		-0,151	0,005	0,000	0,377	0,485
	required job experience (0: <1 year; 1: >1 year)		-0,014	0,004	0,001	0,513	0,500
	payment scheme (0: piece/hourly wage, no wage; 1: monthly/yearly wage)		-0,024	0,004	0,000	0,414	0,492
	yearly income (percentile rank)		-0,085	0,004	0,000	0,330	0,470
	not challenged enough (0: yes; 1: no)		-0,049	0,006	0,000	0,073	0,260
	more training necessary (0: yes; 1: no)		0,019	0,004	0,000	0,788	0,409
	level of computer use (0: simple; 1: moderate or complex)		0,003	0,004	0,475	0,629	0,483
	cooperating with others (1: non of the time; 2: <1/4 of time; 3: <1/2 of time; 4: <3/4 of time; 5 all of the time)		-0,003	0,002	0,094	3,713	1,355
tasks (share of working time (%), see OECD Working Paper for details)	exchanging information		0,255	0,069	0,000	0,054	0,047
	training others		-0,912	0,084	0,000	0,024	0,028
	presenting		-1,544	0,122	0,000	0,008	0,017
	selling		0,879	0,074	0,000	0,018	0,032
	consulting		0,152	0,081	0,059	0,038	0,031
	planning own activities		-0,564	0,083	0,000	0,039	0,036
	panning activities of others		-0,649	0,088	0,000	0,020	0,027
	organizing own schedule		-0,305	0,078	0,000	0,051	0,041
	influencing		-1,429	0,081	0,000	0,030	0,031
	negotiating		0,075	0,084	0,373	0,024	0,028
	solving simple problems		-0,281	0,074	0,000	0,050	0,038
	solving complex problems		-0,441	0,090	0,000	0,023	0,024
	working physically for long		-0,224	0,059	0,000	0,048	0,064
	using fingers or hands		0,343	0,058	0,000	0,068	0,063
	reading instructions		-0,472	0,061	0,000	0,043	0,040
	reading professional publications		-1,286	0,116	0,000	0,013	0,018
	reading books		-1,476	0,107	0,000	0,008	0,019
	reading manuals		0,082	0,085	0,330	0,022	0,024
	writing articles		-1,192	0,342	0,000	0,001	0,006
	filling forms		-0,255	0,064	0,000	0,033	0,034
	calculating shares or percentages		-0,238	0,074	0,001	0,029	0,031
	complex math or statistics		-0,431	0,156	0,006	0,003	0,011
	internet use for work-related info		-0,328	0,084	0,000	0,035	0,025
	using programming language		-1,427	0,179	0,000	0,003	0,011
	using communication software		-0,517	0,143	0,000	0,005	0,013
	constant		1,332	0,084	0,000		

References

Acemoglu, D., and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. Handbook of labor economics, 4, 1043–1171.

Acemoglu, D., and Restrpo, P. (2017): Robots and Jobs: Evidence from U.S. Labor Markets, MIMEO.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3), 3–30.

Autor, D. H. and M. J. Handel (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. *Journal of Labor Economics*, 31(2), S59–S96.

Brynjolfsson, E., and McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. WW Norton and Company.

Frey, C. and M.A. Osborne (2017). The Future of Employment: How Susceptible are Jobs to Computerization?. *Technological Forecasting and Social Change*, 114: 254–280.

Gregory, T., Salomons, A. and Zierahn, U. (2016). *Racing With or Against the Machine? Evidence from Europe.* ZEW Discussion Paper No. 16-053.

Ibrahim, J. G. (1990). Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association*, Vol. 85, No. 411, 765-769.

Ibrahim, J.G., Chen, M.-H., Lipsitz, S. R. and Herring, A. H. (2005): Missing-Data Methods for Generalized Linear Models: A Comparative Review, *Journal of the American Statistical Association*, 100(469): 332-346.

Nordhaus, W. D. (2015). Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth, NBER Working Paper 21547.

Pratt, G.A. (2015): Is a Cambrian Explosion Coming for Robotics? Journal of Economic Perspectives 29(3): 51-60.

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, Vol. 24, 235-270.