London School of Economics and Political Science
tlc.groups@lse.ac.uk

Training for the Future: A Sector-Based Approach to the Analysis of the Relationship Between Automation and Training

LSE GROUPS 2019

Group 7

Sarah Ang

Maurice Hirt

Cherry Nguyen

Haoyu Zhang

Xiaoyun Zhang

Abstract

It is widely acknowledged that automation is becoming the future of work. This points towards the possible return of Solow's paradox of high levels of technology investment and low productivity gains, bringing into focus the importance of job training. Although the risk of automation varies across sectors, making the case for differentiated responses in the form of training, there is currently little focus on comparative approaches to job training. Using a sector-based approach, this paper examines the relationship between the risk of automation and job training by building on the ONS measure of the risk of automation from 2011 to 2017 and extending it to the sectorial level, using the data from the UKCESS over the same period. This paper finds a curvilinear relationship between the risk of automation and the proportion of employees trained as well as a positive interaction of education and automation. Additionally, a significant relationship between training and other sector-level factors including the size of the firms, the proportion of charities and the lack of information on employees' educational background is observed. These findings provide insight into gaps in the provision of training and serve to inform employers' current and future training decisions.

Key words: automation, job training, sector, education, technology

I. Introduction

Solow's productivity paradox of high levels of technology investment but low productivity gains (Brynjolfsson, 1993) in the 1970s was due to gaps in job training, as mismanagement of new technology hampered productivity gains. With automation, this problem may re-emerge as although all sectors are at risk, training levels vary significantly (GOV.UK, 2013), signaling that certain sectors are not adequately prepared.

Using a sector-based approach and drawing from the UK Commission for Employment and Skills Survey (UKCESS) and Office for National Statistics (ONS) data, this paper seeks to answer the question: *what is the relationship between the risk of automation and job training in the UK?* This paper hypothesises that sectors with higher risk of automation such as manufacturing and agriculture tend to provide lower levels of employee training. As automation is introduced into the workplace, organisations in sectors with heavily manual processes will replace their employees with new technology, diminishing employers' incentives to provide job training. In addition, this paper presents new method for calculating the risk of automation, discusses the interaction between the risk of automation and proportion of the mediumly educated workers in a sector, and examines the relationship between the proportion of charities and firm size on training provision. In this case, automation is defined as the creation and application of technology to monitor and control the production and delivery of products and services (International Society for Automation, 2019). For the purpose of the study, employee training includes all types of on-the-job and off-the-job training employees receive.

The first section of the paper situates the research within the current context of the determinants of training levels. The second part presents the case selection, data, and the methodology, including the new construction of sector-level risk of automation. To conclude, findings from the regression analysis are discussed and scope for future research is offered.

II. Theoretical Framework

The following contextualises this paper in terms of (1) how it follows the previous discussion of various factors affecting the provision of training and (2) its comparative, sector-based approach to automation and training.

1. Determinants of training

i) Firm size

Larger establishments tend to provide more training for their employees due to apprenticeship programmes (Westhead, 1998). Economies of scale and the ability to pool risks also provide incentives for employers to invest in training (Black et al.,1999).

ii) Education

Lynch and Black (1998) find a positive relationship between the number of years of schooling for employees and employers'human capital investments. Better educated workers have stronger cognitive skills, reducing costs and increasing returns to training (Booth, 1991). However, Ariga and Brunello (2006) note that the marginal costs of on-the-job training increase with higher levels of education, leading to higher opportunity costs of training; over-education thus decreases training (Sicherman, 1991). Moreover, Bartel and Sicherman (1995) suggest that technological changes could mitigate the impact of education on the provision of training.

iii) Public/Private/Charity

Public and charity sectors are more likely to offer training for their employees than for-profit sectors. Non-profit sectors lack clear goals and measurable performance outcomes and are therefore more responsive to institutional pressures (eg. competition), seeking training as an indication of legitimacy and good reputation (Yang, 2006). Private profit-seeking firms prioritise profits, which limits and discourages expenditure on training (Booth, 1991).

iv) Skill shortages

Skill shortages are associated with the provision of training by employers in order to bridge the skill gaps (Gashi, Pugh and Adnett, 2010). Similarly, Westhead (1998) notes that the nature of the job market influences the amount of training employees receive, with businesses located in labor markets with current skill shortages training existing staff members.

v) Technology

Findings on the link between technological change and employer-provided training have so far proven mixed. Some authors find a positive relationship between the rate of technological change and formal training due to the firms' efforts to retain their competitiveness through skill development (Bartel and Sicherman, 1998; Gashi, Pugh and Adnett, 2010). There is a mutually reinforcing relationship between training and technology adoption, with new technology leading to higher training provision and vice versa. On the other hand, Brown and Campbell (n.d.) describe an overall negative correlation between technological change and training in the semiconductor industry. As the industry is subject to high risk of automation and competitive pressure to improve products in the market, technological changes will eliminate low-skilled jobs and trainings, while increasing more high-skill jobs and trainings, widening the skill gap between the two groups. However, Sinde Cantorna and Diéguez Castrillón (2005) find that introduction of new advanced technologies has no impact on employers' decision on employee training provision. Instead, extraneous factors such as the availability of public funding, company size, numerical flexibility, human capital stock and specificity of required qualifications are more significant.

2. Sector-based approach

Nevertheless, approaches to the impact of technology on training have focused either on general technological progress or on specific sectors (Brown and Campbell, n.d; Sinde Cantorna and Diéguez Castrillón, 2005). However, given that the risk of automation varies across sectors (Office for National Statistics, 2017), there is a lack of comparative informaion. This study seeks to fill the gap by focusing on the relationship between automation and training across sectors in the UK.

III. Data and Case Selection

The datasets focus on the UK as it provides the most comprehensive data and the results will allow reference to other developed economies with a similar market structure.

This paper draws from two data sets. The first is the summary of responses to the UKCESS which provides over 100 questions in 2011, 2013, 2015 and 2017. Proxies for relevant factors which affect the level of training provided and the dependent variable, the proportion of employees trained, are drawn from here. The percentage of staff trained was chosen instead of investments in training as the required costs of training to cope with automation and other obstacles might vary across sectors. The second dataset is from the ONS report on the Probability of Automation in England: 2011 and 2017, and is used to determine the main independent variable, the risk of automation for each sector.

IV. Methodology

1. Measuring factors which affect training provision

Based on the literature review, the most frequently occurring determinants of training provision are selected for analysis and are then coded into the UKCESS measures (Table 1).

Variable affecting training	UKCESS measure
Size of firm	% of micro, small, medium and large companies per sector
Classification of sector	% of for-profit, publicly-funded and voluntary (other) firms
Medium education	% of employers who reported that 20%-80% of employees have at least level 4 education
Unknown education	% of employers who reported that they have no knowledge about the percentage of employees who have at least level 4 education

Table 1

2. Risk determination on a sectorial basis

This paper defines the risk of automation as the percentage of jobs within a sector that are at risk of being automated. Given its anticipatory nature, the risk of automation, instead of the current level of automation itself, has been selected as an independent variable to reflects the challenges different sectors are facing more accurately. However, as a sector-based dataset that corresponds to the categories of the UKCESS could not be obtained, a new measure needs to be constructed.

The sector level values for the risk of automation were calculated by first determining the occupational composition of the 14 sectors from 2011-2017, as based on the UKCESS data. Subsequently, an occupation-based measure of automation by the ONS (The Probability of Automation in England: 2011 to 2017) is used (see Appendix A for ONS methodology). Subsequently, calculating the average risk of automation for each band of occupations from ONS data over the time period. The new measure, the risk of automation in each sector, is hence calculated using the weighted sum of the probability of automation of occupations in the sector, using the formula

$$risk\ of\ automation\ of\ sector\ \lambda = \sum_{i=managers}^{elementary\ staff} p_i \mu_i$$

$$p_i = \ risk\ of\ automation\ per\ occupation\ i\ of\ sector\ \lambda$$
$$\mu_i = share\ of\ occupation\ i\ of\ sector\ \lambda$$

Although some potential limitations and sources of error for the measure can occur (see Appendix A), a convergent validity test (Adcock and Collier, 2001) for the year 2017 with data generated by McKinsey (McKinsey Global Institute, 2017) for 11 equivalently defined US sectors (A *Future that Works: Automation, employment and Productivity, 2017*) yields a Pearson correlation of 0.93 (see Appendix A).

3. Processing the data

To obtain an appropriate dataset with enough observations, the UKCESS results from 4 different years are considered. Since the sector categories change slightly in 2017, a weighted calculation is undertaken to merge ‘agriculture’ and ‘mining and utilities’ into ‘primary sector and utilities’ to match with previous years. Similarly, ‘transport’ and ‘communication’ are collapsed to account for differences in definition between the two sectors over the period.

Having collected all relevant variables in a new dataset, an OLS regression is performed. The data is pooled and not treated as panel data even though the observations originate from 4 different years (2011 to 2017). The reason for the collectivisation of the data is to increase sample size. The graphical analysis of the main variables across time reveals that no significant time effects can be found (see Appendix B). Thus, the data is pooled.

Firstly, a graphical analysis of the plots is undertaken. Several partial F-tests are then performed to eliminate insignificant variables such as the skill gaps, vacancies and the impact of skill gaps on business performance (among others). Moreover, based on theoretical considerations (Bartel and Sicherman 1995), an interaction between the percentage of firms which report that 20%-80% of their workforce has at least obtained education level 4 (medium educated workforce) and the risk of automation is introduced, which proves to be significant in a F-test. Additionally, based on a plot between the risk of automation and the dependent variable (see Figure 2 Section V.2.a), a quadratic term is introduced which yields a p-value of 0.06.

Further, a test for influential points (outliers and leverage points) is undertaken which yields two observations (see Appendix B for further discussion). As a result, a regression is run without these points which yields an even more significant model, particularly for the quadratic term which has a p-value of 0.001 (Appendix B). However, there is no reason to believe that the two influential points are erroneous and should therefore not be discarded. Nonetheless, it can be argued that the p-value of the quadratic term is close to 0.05 with the influential points, and 0.001 without the terms, warrants its inclusion in the model.

V. RESULTS

1. Descriptive analysis

Generally, employee training across sectors has not varied significantly across time, apart from in public administration (Figure 1). One possible explanation is the change in fiscal policy (Office for National Statistics, 2016) which might affect government expenditure and hence resources allocated to training. The only significant variation in employee training that can be observed is between different sectors. These differences can be attributed to the effects of automation and other variables which will be discussed below.

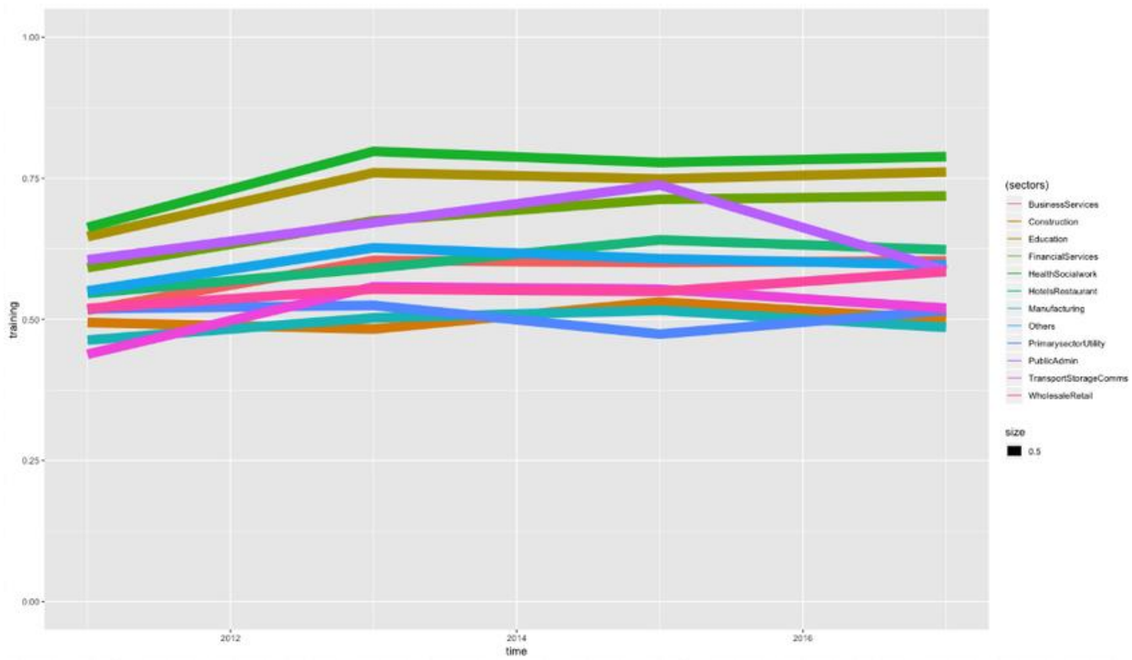

Figure 1

2. Regression model

Overall, there is a negative correlation between automation and training while controlling for other variables, but a positive correlation between the quadratic automation term and training. The finding also demonstrates the effects of other variables and the interaction between education with the risk of automation (Table 2). The significance levels of all variables are lower than 0.05, confirming one variable can determine the dependent variable, ceteris paribus.

Residuals				
Min -0.103313	1Q -0.023350	Median 0.009079	3Q 0.023566	Max 0.074770
Coefficients	Estimate	Standard Error	T-Value	P-Value
Intercept	4.79178	1.73666	2.759	0.008698***
Automation	-13.96448	6.31136	-2.213	0.032697**
Automation2	10.60681	5.54030	1.914	0.062731*
Medium. Education	-3.71641	1.56094	-2.381	0.022123**
Micro. Size	-0.23906	0.05587	-4.279	0.00114***
Charity. Other	0.27457	0.06070	4.523	5.33e-05***
Unknown. Education	-0.51634	0.15810	-3.266	0.002242***
Automation:Medium.Education	8.67389	3.54464	2.447	0.018891**
Residual Standard Error: 0.04275 on 40 degrees of freedom				
Multiple R-Squared: 0.8295			Adjusted R-Squared: 0.79996	
F-Statistic: 27.8 on 7 and 40 DF, p-value: 1.839e-13				
For p-value: 0 (***) 0.01 (**) 0.05 (*) 0.1 (.) 0				

Table 2

a. Automation

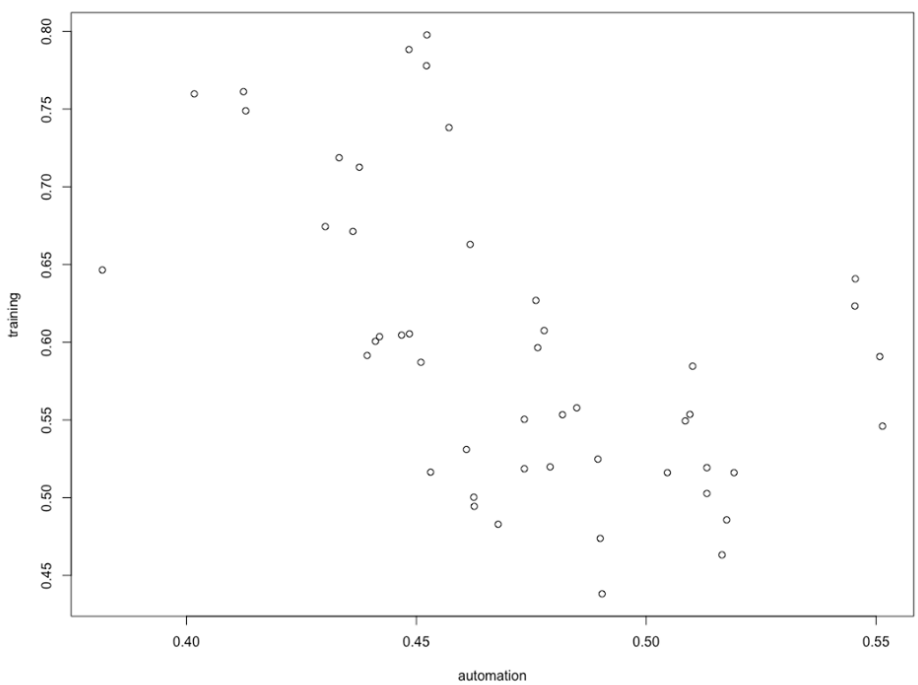

Figure 2

The quadratic term reveals that the risk of automation has a curvilinear relationship with training (Figure 2). The percentage of employees trained decreases with increasing risk of automation until a minimum at a risk of automation of 0.5 is reached. Thereafter, the share of trained employees rises as the risk of automation increases further. This observation is in accordance with the coefficients for automation. A positive coefficient for the quadratic implies a convex function with a global minimum.

This relationship could be explained by introducing two opposing effects of the risk of automation. One possible explanation is that with the introduction of automation, employee training declines as employers are uncertain about the effectiveness of the existing training in dealing with the risk of automation and the transformation of the production processes. Another possible explanation comes from Piore (1968), who argues that costs of adjustment in training increases with technological change. Employers will not proceed with training investment, unless the potential gains from those trainings exceed the adjustment costs. Even if organisations grant a certain allowance to respond to the introduction of a new technological change, adjustment costs can rise substantially when the firms are facing higher threats of automation. As a result, employers might prefer to lay off the current employees rather than providing them with the adequate training to meet the required standards and instead hire new experienced workers.

However, there are also potential gains from automation such as higher productivity, increased competitiveness, and cost reduction. Gashi et al. (2010) argue that undertaking training can increase employees' skill intensity, retaining competitiveness and generating maximum profit from utilising competitive advantages of the new technology. Moreover, the authors suggest a two-way effect, stating that improved skills can decrease the costs of adopting new technology. These findings help explain the positive relationship between the level of risk of automation and training. As organisations are facing a higher threat of technological change, the need to update employees' skill sets becomes more urgent to increase organisational competitiveness. The potential decline in the costs of technological adoption also incentivises employers to offer more training. Establishments might also be more informed of the risks and suitable training strategies to cope with automation after that threshold, thus decreasing the risk of uncertainty. As a result, at a certain level of automation, employers perceive the benefits of employee training to outweigh the risks of uncertainty and cost pressures, resulting in a marginal increase in the proportion of employees receiving training. The quadratic result corresponds with Bartel and Sicherman's (1998) findings that high rates of technological change may facilitate or reduce training due to the uncertainty of costs and benefits in human capital investment.

b. Interaction between education and automation

This section analyses the relationship between the percentage of firms with the mediumly educated workforce and training, and its interaction with automation on employer-provided training. The proportion of firms within the sector with medium workforce education yields a negative coefficient of - 3.72 and the interaction between the mediumly educated workforce and the risk of automation has a positive coefficient of 8.67. Overall, the marginal effect of a higher percentage of mediumly educated workforce depends on the risk of automation and can be summarised as:

$$\frac{\Delta training}{\Delta medium_{education}} = \beta_{medium_{education}} + \beta_{medium_{education} \times\, automation} automation$$
$$= -3.72 + 8.67 automation$$

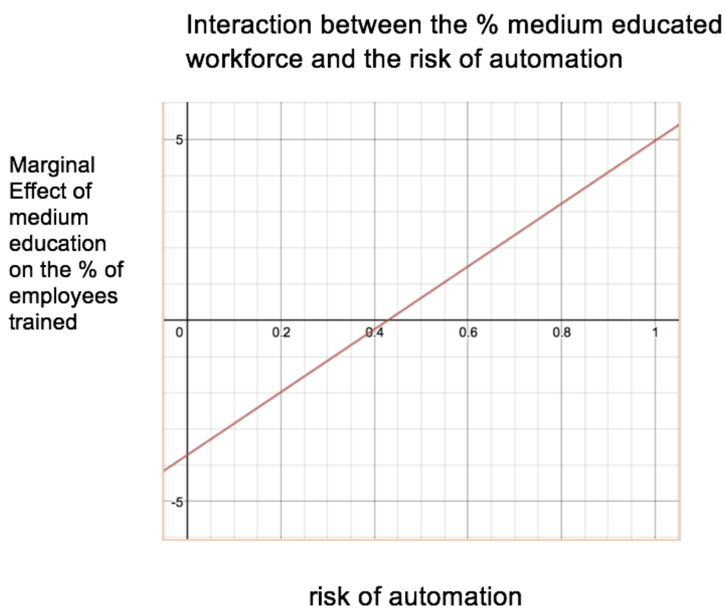

Figure 3

It is reasoned that at a low level of risk of automation (below 0.43) a higher share of mediumly educated workforce has a negative relationship with training. It could be argued that if the workforce is educated and there is a relatively low risk and magnitude of automation, then employees are already equipped with the necessary skills to deal with new technology. This can be considered an extension to the argument that employees who are overeducated for their positions receive less training (Bartel and Sicherman, 1998). However, if the risk of automation is above 0.43, the relationship between the risk of automation and training becomes positive. If the workforce already possesses some education, it becomes effective for the employers to build onto these skills rather than laying off employees and hiring new workers. This increasingly positive marginal effect of medium level of education with a higher risk of automation is visualized in Figure 3 above.

c. Other independent variables

The correlation between the percentage of charitable establishments in a sector and the proportion of employees receiving training yields a coefficient of 0.28, explained by the need for the charities to gain legitimacy and reputation through training provision. As profit-seeking firms have more quantifiable measurements of performance, sectors with higher proportion of charitable organizations may resort to training as an indicator for performance measurement (Yang, 2006).

Regarding size, the percentage of micro-size firms (of 2-4 employees) is negatively correlated with the outcome variable. On average, a 1% increase in the share of micro-companies leads to a 0.24% decrease in training. One possible explanation is that larger firms could exploit the economies of scale by spreading the training costs over a larger number of employees, whereas small firms cannot do so, thus discouraging those firms to invest in training (Lynch and Black, 1998; Westhead, 1998).

By presenting a correlation coefficient of –0.52 and a p-value smaller than 0.05, the result shows that when an employee's level of education is unknown to the employers, they are less likely to provide training. This is in line with the view of education as a positional good that indicates the trainability of the employees (van de Werfhorst, 2011). Moreover, when the level of educational qualifications is unknown to employers, employers are uncertain about the returns of training to productivity (Becker, 1962).

VI. Conclusion

Previous research has found contradictory evidence on whether a positive (Bartel and Sicherman, 1998) or negative (Brown and Campbell n.d.) relationship exists between technology and training. Using a sector-based approach for relationship between the risk of automation and training, the analysis finds that the relationship is far from straightforward. The initial hypothesis for a negative relationship can only be confirmed partially.

Ceteris paribus, this paper provides empirical evidence for a curvilinear relationship between the risk of automation and the percentage of employees trained. Sectors with higher risks of automation tend to provide less training. However, from a threshold of a risk of 0.5, this relationship reverses and more training is provided. This relationship could be explained by employers weighing the costs of training to respond to a new technological change against potential gains from training. Moreover, the interaction between the risk of automation and education adds another dimension to the relationship. The cross-sector comparison also finds am increasingly positive marginal effect of medium level of education with a higher risk.

For the control variables such as the share of micro-companies and charities, the relationship with training confirms previous research on an enterprise-level.

Altogether, the findings from this paper present possible implications on the policymaking regarding human capital investment as sectors with a risk of automation smaller than 0.5 experience decreasing levels of training. Unfortunately, time constraints prevent this study from accessing organisation-level data, which could create a more robust model. Moreover, a fixed regression for more observations could identify potential developments of the variables across time and yield more accurate results. Lastly, this paper only assesses the incidence of training but not the intensity of the training. With improvements implemented, future studies could shed additional light on the provision of job training in the era of automation.

VII. Acknowledgements

The authors appreciate assistance from the LSE GROUPS team, particularly Dr. Ellis Saxey for providing supports throughout the programme, our supervisor Julia Leschke for her comments and invaluable expertise, and Yan Wang for her support regarding the statistical analysis.

Bibliography

Adcock, R. and Collier, D. (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *The American Political Science Review*. [American Political Science Association, Cambridge University Press], 95(3), pp. 529–546. Available at: http://www.jstor.org/stable/3118231.

Ariga, K. and Brunello, G. (2006) 'Are Education and Training Always Complements? Evidence from Thailand', *Industrial and Labor Relations Review*. Sage Publications, Inc., 59(4), pp. 613–629. Available at: http://www.jstor.org/stable/25067560.

Bartel, A. P. and Sicherman, N. (1998) 'Technological Change and the Skill Acquisition of Young Workers', *Journal of Labor Economics*. [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago], 16(4), pp. 718–755. doi: 10.1086/209904.

Becker, G. S. (1962) 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*. doi: 10.1086/258724.

Black, D. A., Noel, B. J. and Wang, Z. (1999) 'On-the-Job Training, Establishment Size, and Firm Size: Evidence for Economies of Scale in the Production of Human Capital', *Southern Economic Journal*. Southern Economic Association, 66(1), pp. 82–100. doi: 10.2307/1060836.

Booth, A. L. (1991) 'JOB- RELATED FORMAL TRAINING: WHO RECEIVES IT AND WHAT IS IT WORTH?', *Oxford Bulletin of Economics and Statistics*. doi: 10.1111/j.1468-0084.1991.mp53003004.x.

Brown, C. and Campbelll, B. (no date) *Technological Change, Training, and Job Tasks in a High-Tech Industry*. Available at: http://www.irle.berkeley.edu/worktech/training.pdf (Accessed: 13 June 2019).

Brynjolfsson, E., 1993. The productivity paradox of information technology. *Communications of the ACM*, *36*(12), pp.66-77.

Gashi, A. N., Pugh, G. and Adnett, N. (2010) 'Technological change and employer- provided training: evidence from UK workplaces', *International Journal of Manpower*. Emerald, 31(4), pp. 426–448. doi: 10.1108/01437721011057010.

GOV.UK (2013) *Employer Investment in Training*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/306433/ukcess13-employer-investment-in-training.pdf?fbclid=IwAR0xw-ukMDLb7oCU4-GXtRKmX6xAwJmsxD91WsCXtPPJ0Lr-dgcl4901MmQ (Accessed: 13 June 2019).

International Society for Automation (2019) *What is Automation?*, *International Society for Automation*. Available at: https://www.isa.org/about-isa/what-is-automation/ (Accessed: 13 June 2019).

Lynch, L. M. and Black, S. E. (1998) ‘Beyond the Incidence of Employer-Provided Training’, *Industrial and Labor Relations Review*. Sage Publications, Inc., 52(1), pp. 64–81. doi: 10.2307/2525243.

McKinsey Global Institute (MGI) (2017) *A FUTURE THAT WORKS: AUTOMATION, EMPLOYMENT, AND PRODUCTIVITY*. Available at: www.mckinsey.com/mgi. (Accessed: 13 June 2019).

Office for National Statistics (2016) *UK Perspectives 2016: Spending on public services in the UK - Office for National Statistics*, *Office for National Statistics*. Available at: https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicsectorfinance/articles/ukperspectives2016spendingonpublicservicesintheuk/2016-05-25 (Accessed: 13 June 2019).

Office for National Statistics (2017) *The probability of automation in England - Office for National Statistics*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/theprobabilityofautomationinengland/2011and2017 (Accessed: 13 June 2019).

Piore, M. J. (1968) ‘On-The-Job Training and Adjustment to Technological Change’, *The Journal of Human Resources*. [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], 3(4), pp. 435–449. doi: 10.2307/144796.

Sicherman, N. (1991) ‘“Overeducation” in the Labor Market’, *Journal of Labor Economics*. The University of Chicago Press, 9(2), pp. 101–122. doi: 10.1086/298261.

Sinde Cantorna, A. I. and Diéguez Castrillón, I. (2005) ‘The effect of the implementation of advanced manufacturing technologies on training in the manufacturing sector’, *Journal* *of European Industrial Training*. Emerald, 29(4), pp. 268–280. doi: 10.1108/03090590510597124.

Van de Werfhorst, H. G. (2011) ‘Skills, positional good or social closure? The role of education across structural–institutional labour market settings’, *Journal of Education and Work*. Routledge, 24(5), pp. 521–548. doi: 10.1080/13639080.2011.586994.

Westhead, P. (1998) *Factors associated with the provision of job-related formal training by employers*, *International Journal of Entrepreneurial Behaviour & Research*. doi: 10.1108/13552559810235592.

Yang, S. (2006) *Organizational Sectors and the Institutionalization of Job-Training Programs: Evidence from a Longitudinal National Organizations Survey*, *Sociological Perspectives - SOCIOL PERSPECT*. doi: 10.1525/sop.2006.49.3.325.

Appendix A

Methodology for ONS measure of automation

This provides the information on the risk of automation for each specific occupation, by processing data of UK PIAAC (a sample of 8,892 individuals and the probability of automation of each individual) and Frey and Osborne probabilities of automation, using a modified OECD method. By regressing automation probability (OF) on characteristics of jobs as measured by tasks, a parameter that determines the influence of each job characteristic on the probability of automation is calculated, and substituting these values into PIAAC and APS (another dataset similar to PIAAC but with a larger sample size but without task variables), automation probability of each individual is calculated, with the occupation group the individual belongs to turns out to be the most significant variable.

Convergent validity test (Adcock and Collier, 2001)

The generated data for the risk of automation is plotted against existing data from McKinsey (McKinsey Global Institute, 2017) which yields a Pearson correlation of 0.93.

Risk of automation per sector McKinsey-generated data

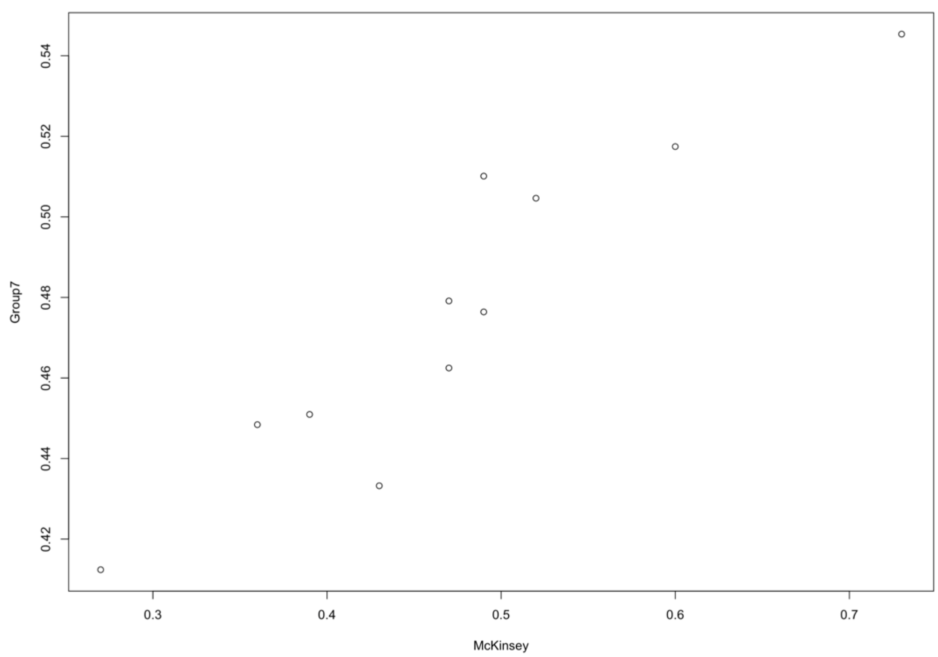

Figure 1

Due to differences in some definitions of sectors, only the relationship of 11 sectors in 2017 could be examined. These sectors include:

Primary Sectors and Utilities

Manufacturing

Construction

Wholesale and Retail

Hotels and restaurants

Transport, Storage and Communications

Financial services

Public admin.

Education

Health and social work

Arts and other services

Appendix B

Variables across time

Risk of automation across time between different sectors

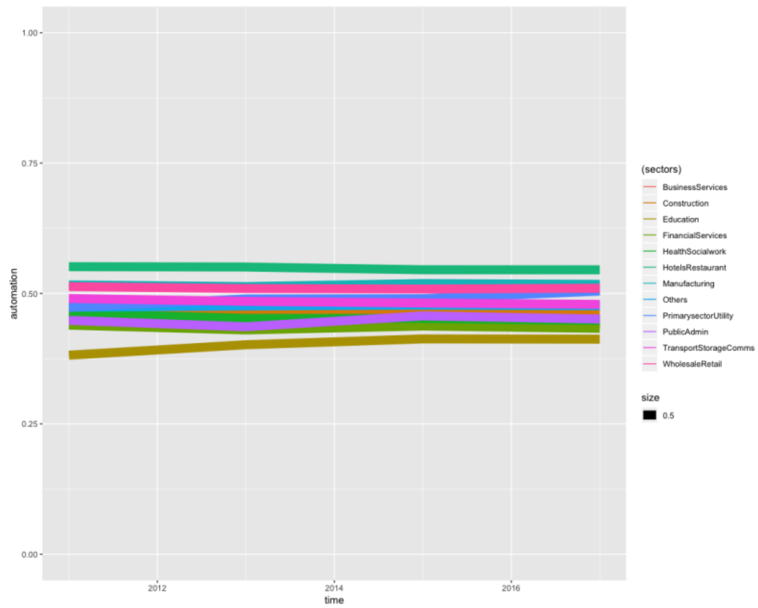

Figure 2

Percentage of employees trained across time between different sectors

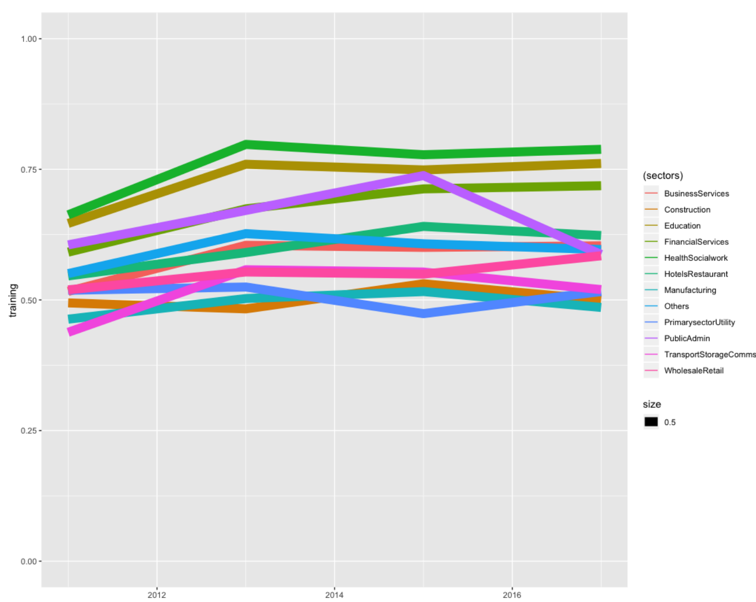

Figure 3

Percentage of micro-companies on each sector across time

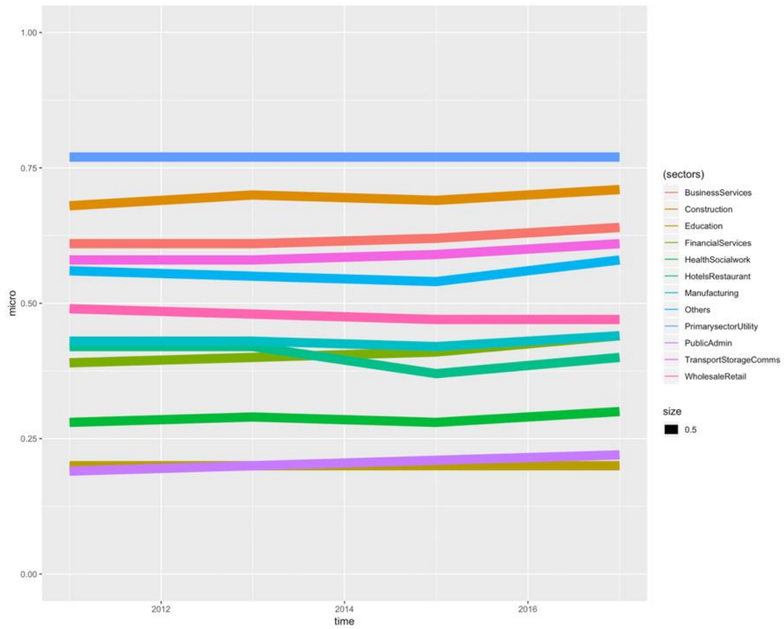

Figure 4

The negligible effect over time allows for a pooling of the data.

Cook's Distance

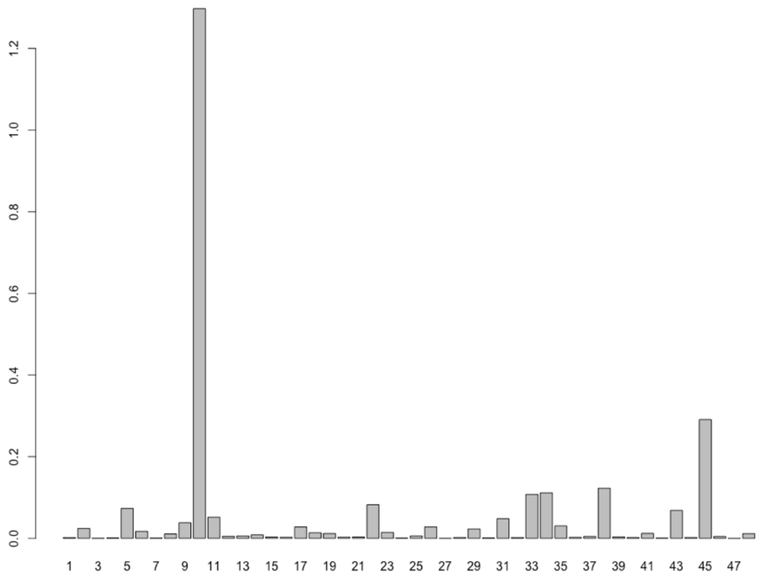

Figure 5

Observation 10 and 45 with Cook's Distance above 0.3 which implies influence on regression.

Regression table excluding outliers

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.084282 -0.023915  0.001543  0.022025  0.086402

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           7.22744    1.75210   4.125 0.000195 ***
automation          -24.07591    6.54335  -3.679 0.000722 ***
automation2          21.19063    6.01445   3.523 0.001128 **
medium.1             -3.40446    1.42564  -2.388 0.022015 *
micro                -0.19138    0.05304  -3.608 0.000885 ***
charity.other         0.30490    0.05620   5.425 3.49e-06 ***
unknow               -0.48824    0.17023  -2.868 0.006703 **
automation:medium.1   7.68184    3.24379   2.368 0.023066 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03888 on 38 degrees of freedom
Multiple R-squared:  0.8652,    Adjusted R-squared:  0.8404
F-statistic: 34.85 on 7 and 38 DF,  p-value: 1.212e-14
```

Table 1

Possible explanation for the outlier in automation-training model

In the automation-training model where the effects of other independent variables are not cleared, we have an outlier, the observation of education in 2011, which increases the p value of the regression model hugely. One possible explanation is the education in 2011 has an abnormally high percentage (12%) of employers who don’t know the level of education of their employees, compared to 5% in 2013, 2% in 2015 and 1% in 2017. According to our model, there is a negative correlation between percentage of employers who don’t know the level of education of their employees and level of training provided, which can explain the deviation of the outlier.