Term Paper Proposal

Sarah Gust *ifo Institute at the University of Munich, gust@ifo.de*
Ann-Christin Kreyer *Max Planck Institute for Innovation and Competition, ann-christin.kreyer@ip.mpg.de*

This is where we put the abstract...

Introduction

Digitization has fundamentally changed labor demand. Middle-skilled workers were replaced by technology while the demand for high skilled workers and low-skilled workers has grown (Acemoglu and Autor, 2011; Autor and Dorn, 2013). Autor and Dorn (2013) show that increased employment at the lower tail of the earnings distribution is mainly due to an increase in service occupations. At the upper tail, technological changed led to a college wage premium: wages of college graduates relative to high-school graduates increased (Acemoglu and Autor, 2011). De La Rica et al. (2020) find that a one-standard-deviation increase in abstract tasks is related to a 3.3-log-point wage premium. For each standard deviation of routine tasks there is a 2.6 to 2.9-log-point wage penalty. The vast literature on the change of job tasks takes skills of workers as pre-defined. However, workers and firms can also invest in new skills via training. Massiv Open Online Courses (MOOC) have facilitated global access to ICT and programming courses. It is often claimed thate mployees who i.e. worked in non-programming jobs before could try to improve their job opportunties by reskilling and acquiring programming skills using MOOC (Garrido et al., 2016; World Economic Forum, n.d.). We are more skeptical that open courses can extensively contibute to retraining as they already require a decent level of ICT skills.

In this paper, we investigate the specific characteristics of workers that participate in on-the-job training and open education. The goal is to find out if training and especially new open educational programs are indeed an opportunity for middle skilled workers that are primarily effected by decreasing job opportunities to take on more abstract tasks. Or if training reinforces inequalities because only high-skilled workers receive and invest in training.

Becker (1962) distinguished between two kinds of training: specific and general. Specific training increases the marginal product of a worker within one specific firm while general training increases her productivity in many other firms. In a perfect labor market, workers are paid their marginal product. In such case, firms would not invest into general training of their employees as they could leave the firm and look for a better paid job. Instead, workers would pay for their general training as an investment into higher future wages. Lynch (1991) and Lynch (1992) find that on-the-job training tends to be firm specific in the US and thus wage raises cannot be taken along to subsequent employers. Off-the-job training by proprietary institution have little effect on wages in the current employment but raise future expected wages in subsequent employment. Acemoglu and Pischke (1999) argue that firms still invest in general training due to their monopsony power. Wages increase by less than the marginal productivity and firms can profit. Konings and Vanormelingen (2015) find that an increase in the share of trained workers by 10 percentage points raises the productivity by 1.7 to 3.2 percent while wages only increase by 1.0 to 1.7 percent.

Previous literature on training focuses on wage and productivity effects but the research on the specific characteristics of workers that participate in training is scarce. Applying a machine learning, we can identify the factors that drive the probability of receiving training from a large set of 1,460 variables of the survey of the Programme for the International Assessment of Adult Competencies (PIAAC). More specifically, we apply a Lasso regularization proposed by Tibshirani (1996). In the baseline model, we estimate a Lasso logistic model for the binary outcome variables of whether or not a person participated in on-the-job training or open education. In the second prt of then paper, we estimate a Lasso linear model for the number of on-thejob and off-the job trainings.

For our baseline model, we find that people in higher skilled occupation, with a higher educational level, and who require computer knowledge generally receive more training. Thus, training is likely to increase productivity in high skilled jobs and fuel wage growth at the upper tail of the wage distribution. It does not seem to support workers in climbing up the skill ladder and aquire more abstract tasks. The results for on-the-job traning and open education are very similar. Individuals that do not receive on-the-job training are also less likely to participate in open education. With the Lasso regularization, we improve upon the simple logistic model. The accuracy increases from around 0.5 in the logistic model to 0.7 for on-the-job training and 0.8 for open education in the Lasso logistic models. Our Lasso logistic model performs much better than the Lasso linear model. The regularized linear regression results partly confirm our findings from the Lasso logistic model. Higher education as well as the need to always be up to date and using the computer daily participate in more on-the-job trainings. There is also a trend visible that lower educational levels increase the number of open courses. Other seletected variables are more puzzling. However, the out-of-sample performance of the linear regularized model is worse compared to the logistic regressions which is why we rely on those results.

Data and Desriptive Statistics

To explore and identify the factors that drive the probability employees receive on-job or off-job training, we use the results of the survey of the Programme for the International Assessment of Adult Competencies (PIAAC) (GESIS - Leibniz Institute for the Social Sciences (n.d.)). The survey was conducted by the Organisation for Economic Co-operation and Development (OECD) with the goal to assess which skills adults need to manage challenges and tasks at work as well as in their personal life. The study targeted explicitly the skills in literacy, numeracy and adaptive problem solving. However, for our research question the focus lies on the comprehensive background information the study also provides. This comprises the respondents' past and current education and job-related trainings, subjective assessments of their skills and job requirements (GESIS - Leibniz Institute for the Social Sciences (n.d.)). The study was conducted in cycles and the first cycle consisted of three rounds which began in 2011/12. In the first round 24 countries took part. In the second round nine additional countries participated and in the last round individuals from five different countries were questioned. In total 40 countries participated in the first cycle comprising about 5,000 randomly selected adults who were between 16 and 65 years old. The second cycle started in 2018 and results are to be expected in 2022 (GESIS - Leibniz Institute for the Social Sciences (n.d.)).

For the study at hand the results of the first wave are used in a reduced form. The original dataset comprises 1,460 columns with 230,691 observations of respondents. However, the research question of this paper is to analyze the probability of trainings for middle-skilled workers compared to trainings for high-skilled and low-skilled workers in the wake of the increasing polarization of skills following from digitization. To answer this, the original PIAAC dataset is reduced to 52 variables in total. For example, information on the various test results conducted in the study are excluded as well as variables with no or very few observations. Additionally, we focus on columns that contain information on the current job to estimate the training probability. To achieve comparability across countries, questions which were only answered by respondents living in the United States are also excluded from analysis. The final dataset is cross-sectional with one observation representing the answers of one respondent.

The 52 variables we kept in the final dataset comprise information on the individual's background information, her past and ongoing formal or informal education, information on training activities, information on ICT skill use at work and the respective extensive and intensive margin, her subjective job requirements, information on her curent job and information on monthly income. Since there are many categorical variables in the dataset, we created dummies for the different categories. We obtain thus a total sum of variables of 180 that are used in the final dataset.

As we do not restrict the dataset in terms of respondents but only in terms of questions answered, our final dataset comprises 230,691 observations of individuals. Of those 230,691 individuals, 122,830 are female and 107,859 are male (see Figure 1). The age of the respondents is evenly distributed between the ages 16 to 64 with a female mean age of 39.95 years and male mean age of 39.38 (see Figure 1). The number of respondents working in skilled white-collar occupations is the highest with 73,090 respondents. 53,756 respondents work in semi-skilled white-collar occupations and 38,238 are working in semi-skilled blue-collar occupations. Occupations classified as elementary amount to 19,363 respondents (see Figure 2).

The key variables of this study are the the skill classification of the individual's job and her trainings comprising on-the-job training and distance or open training courses. On-the-job trainings comprise all trainings that are organized by the employer for his employees. Such trainings take place during working time and are often conducted by internal experts of the respective topic (Acemoglu and Autor, 2011). Open or distance education courses, on the other hand, comprise methods of learning and teaching without or only little face-to-face interactions and separations in time and space, e.g. MOOCs (Jung, 2019). In the PIAAC survey, the respective questions for these two types of trainings are the following: *How many of open or distance education activities did you participate in?* and *How many of organized sessions for on-the-job training or training by supervisors or co-workers did you participate in?*. The classification of the respondents jobs in terms of skills is evenly distributed across age groups and gender. However, respondents working in semi-skilled white-collar occupations are slightly younger than those working in skilled occupations or semi-skilled blue-collar occupations (see Figure **??**). Looking at the distribution of on-job trainings and occupational classification alone shows that occupations classified as skilled or semi-skilled white-collar occupations received slightly more trainings in the last year than blue-collar occupations (see figure 3). On average individuals working in skilled white-collar occupations participated in 3.6 on-job trainings and while individuals in semi-skilled white-collar occupations participated in on average approximately 3 on-job trainings. Individuals working in elementary occupations received the least on-job trainings with on average 2.58 trainings in the last year (see Table 1). The distribution of the off-job trainings among the job classifications is similar although the averages are lower for all job classifications (see Figure 3). The average number of off-job trainings for individuals working in skilled white-collar occupations amounts to 2.8 trainings, while individuals in semi-skilled white- or blue-collar occupations received on average 2.2 trainings in the past year. Again, individuals working in elementary jobs receive the least number of off-job trainings on average (see Table 2). The skeweness in the distribution of both off- and on-job trainings towards skilled white-collar occupations is evident.

Simple OLS regressions using only a few selected variables including gender, years of education to get the current job, number of children, occupational skill level and ICT skill use at work, reveal that there are positive correlations between the high-skilled jobs and the number of trainings respondents participated in. This holds true for on-the-job-training and distance or open educational training. However, as we have a dataset that provides many different information on the individuals, we want to be able to include as many variables and thus information as possible in the regressions. In the next chapter we therefore explain the Lasso regression.

Figure 1: Distribution of age and gender

(a) Gender distribution

(b) Age distribution

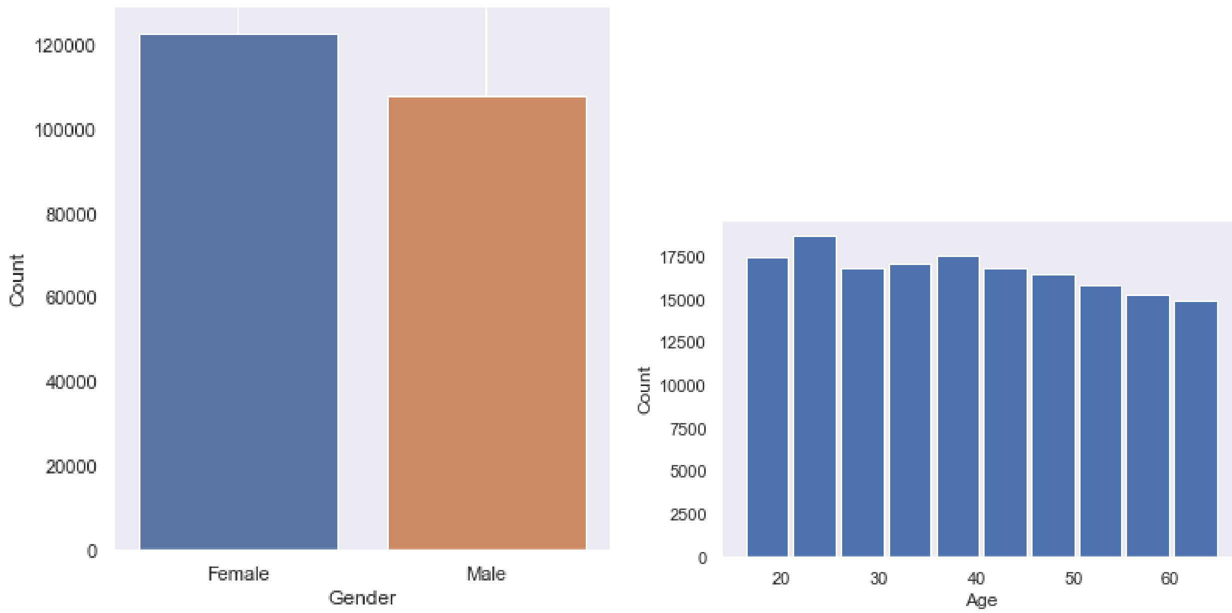


Figure 2: Distribution of occupational classifications

(a) Classification of jobs

(b) Age and Classification of jobs

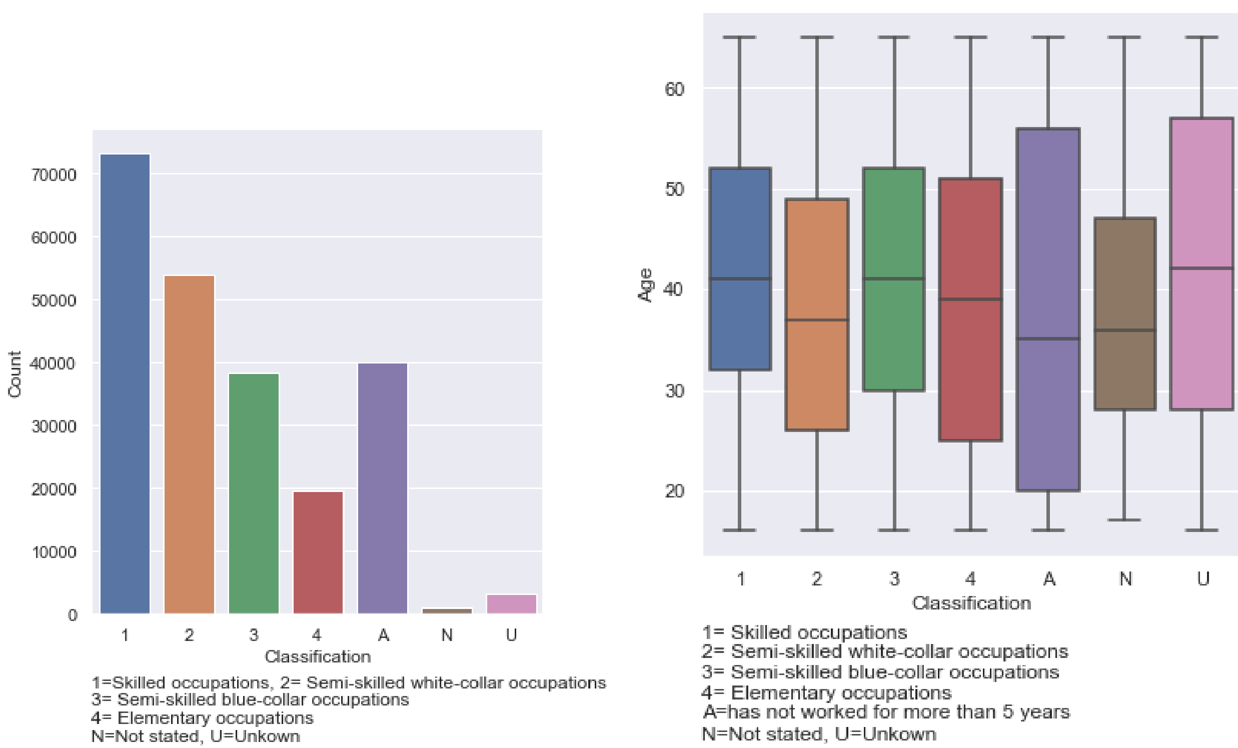


Figure 3: Open courses and classification of jobs

(a) Classification of jobs and trainings

(b) On-job trainings and classification of jobs

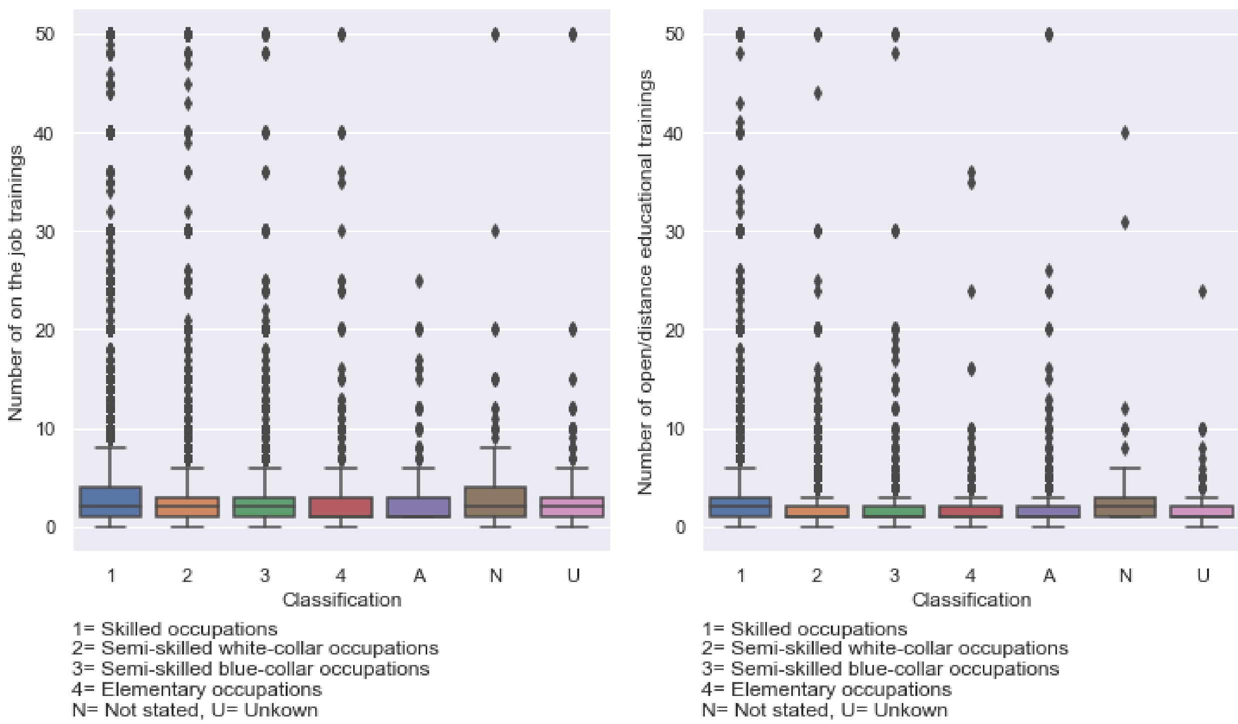


Table 1: Average number of on-job trainings per job classification

Job classification	Mean number of on-job trainings
1	3.66353
2	2.95611
3	2.586
4	2.45609
A	2.58834
N	3.61111

Job classification	Mean number of on-job trainings
U	3.28139

Table 2: Average number of open job training courses per job classification

Job classification	Mean number of open training courses
1	2.84033
2	2.21544
3	2.22948
4	1.7525
A	1.99916
N	3.16867
U	2.21875

Lasso Logistic Model

We now estimate the Lasso model for the binary outcome variables of whether or not a person received on-the-job training and whether or not a person participated in open education. We estimate the Lasso-regularized logistic model for the probability that a person received one specific training as follows:

$$\hat{\theta}_\lambda = argmin(-l_N(\theta)) + \lambda \sum_k |\theta^k| \tag{1}$$

where $l_N(\theta))$ is the log-likelihood function

$$l_N(\theta)) = \sum_i [y_i x_i \theta - log(1 + e^{x_i \theta})]$$

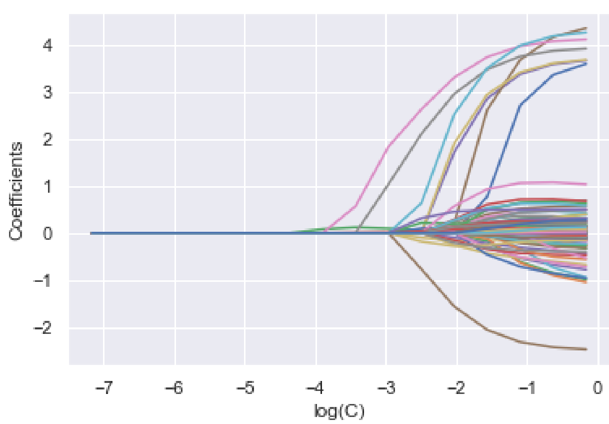
$\sum_k |\theta^k|$ is the lasso penalty that shrinks coefficients of little explanatory power to zero. $\lambda > 0$ is the penalty weight.

Training the Model

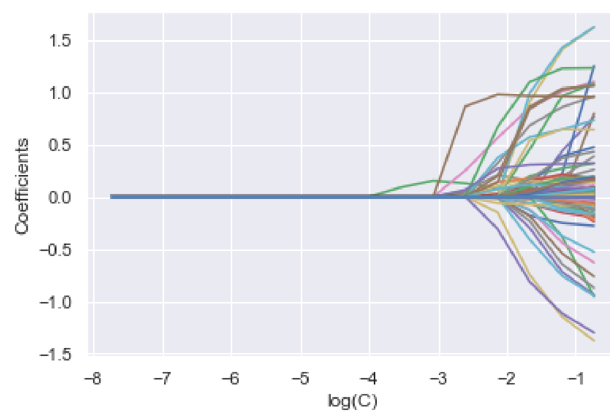
Figure 4 presents the Lasso regularization path for the logistic candidate models. The models are ordered from the most penalized to the least penalized model and the algorithm includes more non-zero coefficients in the model. We select λ via *5-fold cross validation* which leads to an optimal λ of 29.764 for on-the-job training and an optimal λ of 0.089 for open education.[1]

Figure 4: Logistic model: Lasso path

(a) On-the-job training



(b) Open education



Variable Selection

The Lasso logit model indentifies 152 columns[2] with non-zero predictive power for on-job training. We present the coefficients in Table A3. Let us first turn to the results for the skill level. *skill_4* are elementary occupations and represent the reference group here. *skill_3* is the dummy variable for semi-skilled blue-collar occupations, *skill_2* is the dummy for semi-skilled white-collar occupations, and *skill_1* is the dummy for skilled occupations. Working in a semi-skilled blue-collar occupation and working in a semi-skilled white-collar occupation increases the probability of receiving training by 00.45 % and 00.42 % respectively compared to working in an elemetary occupation. Working in a skilled occupation increases the probability of receiving training by 10.11

[1]Note that these results include randomization which may lead to different outcomes if run again.
[2]Including country and industry controls

% relative to working in an elementary occupation.

The dummy variable for whether a person was employed during studying for a qualification, *b_q10a_Yes*, has the highest positive explanatory power. If a person uses a computer on this specific job (*g_q04_Yes*), it increased the probability of receiving on-the-job training by 67.52 %. Having general computer experience (*computerexperience_Yes*) increases the probability of receiving on-the-job training by 38.93 %. People are 24.52 % more likely to participate in on-the-job training if they have the feeling that they need more training in order to cope well with their present duties (*f_q07b_Yes*). Moreover, employees are more likely to receive training if they work in larger companies, compared to smaller companies and if they have a higher educational level.

If a job does not involve keeping up to date with new services and products (*d_q13c_Never*), it lowers the probability of reveiving training by 37.60 %. Never participating in online discussions such as conferences (*g_q05h_Never*) reduces the probability of receiving on-the-job training by 39.00 %. Also, having a low education level, if a job needs less than one month of prior work experience, and working in a job without a contract have the most negative effects on the chances of participating in on-the-job training.

For open education, the Lasso logit model identifies 138 non-zero columns[3] The results look very similar to the on-the-job training. Working in a skilled occupation increases the probability of participating in open education by 11.03 %. The indicator for semi-skilled blue-collar or white-collar workers zero and thus excluded by the Lasso regularization. As in on-the-job training, the most important feature is being employed and using the computer on this specific job (66.60 %), or having general computer experience (33.45 %). Also working in a larger company and having a higher educational level increases the probability of participating in open education. Lower education, having no contract, and never using the computer for work tasks negatively affect the probability of participating in open education.

We can draw two main conclusions from these results. First, people in higher skilled occupation, with a higher educational level, and who require computer knowledge generally receive more training. Thus, traning is likely to increase productivity in high skilled jobs and fuel wage growth at the upper tail of the wage distribution. It does not seem to support workers in climbing up the skill ladder and aquire more abstract tasks. Secondly, on-the-job traning and open edu-

[3]Including country and industry controls

cation are complements. Individuals that do not receive on-the-job traning are also less likely to participate in open education.

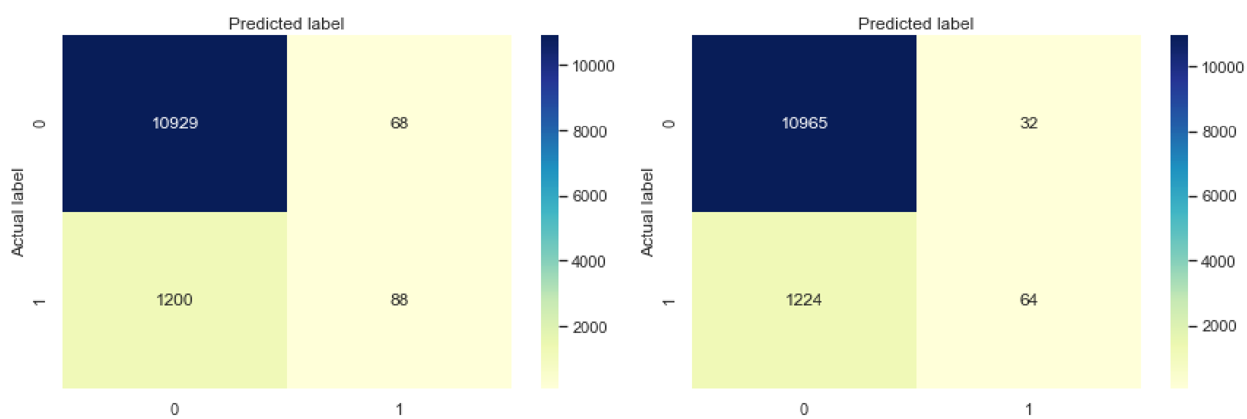
Accuracy of the Model

Finally, we evaluate the accuracy of the Lasso logit model. Figure 5a presents the confusion matrix for on the job training. 10929 are true negative prediction and there are 88 true positive predictions. The model yields 1200 false negative predictions and 68 false positive predictions. We present the confision matrix for open education in Figure 5b. For this model, we obtain 10965 true positive predictions and 64 true negative predictions. 1224 + 32 are incorrect predictions.

Figure 5: Logistic model: Confusion matrix

(a) On-the-job training

(b) Open education



In Table 3, we compare the accuracy of the Lasso logistic model with the unregularized logistic model. The test accuracy of the logistic model is 0.4858 and the test accuracy of the Lasso logistic model is 0.7292 for on-job training. The test accuracy for open education is 0.5021 for the logistic model and reaches 0.8188 with the Lasso logistic model. With the Lasso penalty we excluded unnecessary variables from our Logistic regression that caused over-fitting. The Lasso-regularized logistic model performs much better for both outcome variables.

Table 3: Accuracy of the Lasso logistic model

	Lasso logistic model	Logistic model
On-job training		
Training accuracy	0.7288	0.4962
Test accuracy	0.7292	0.4858
Open education		
Training accuracy	0.8185	0.4996
Test accuracy	0.8188	0.5021

Lasso Linear Model

We start with a simple linear regression model $y = X\beta + \varepsilon$, where $y \in \mathbb{R}^N$ is the predicted participation in on-job or off-job training, $X \in \mathbb{R}^{N\times k}$ are the vectors of covariates, and $\varepsilon \in \mathbb{R}^N$ is the residual with the standard assumptions of OLS. To select the set of covariates with the strongest predictive power from our set of 180 variables, we apply the Least absolute shrinkage and selection operator (Lasso) that was first proposed by Tibshirani (1996). We add the Lasso penalty equal to $\sum_k |\beta_k|$ to our linear model. The Lasso linear estimator $\hat{\beta}$ is then given by

$$\hat{\beta}_\lambda = argmin\{\sum_i (y_i - x_i^T\beta)^2 + \lambda \sum_k |\beta_k|\} \tag{2}$$

That is, we minimize the sum of squared residuals but the Lasso penalty puts a cost at every $\hat{\beta} \neq 0$ and thus, we penalize complexity and avoid over-fitting the model. $\lambda > 0$ is the penalty weight or the *tuning parameter*.

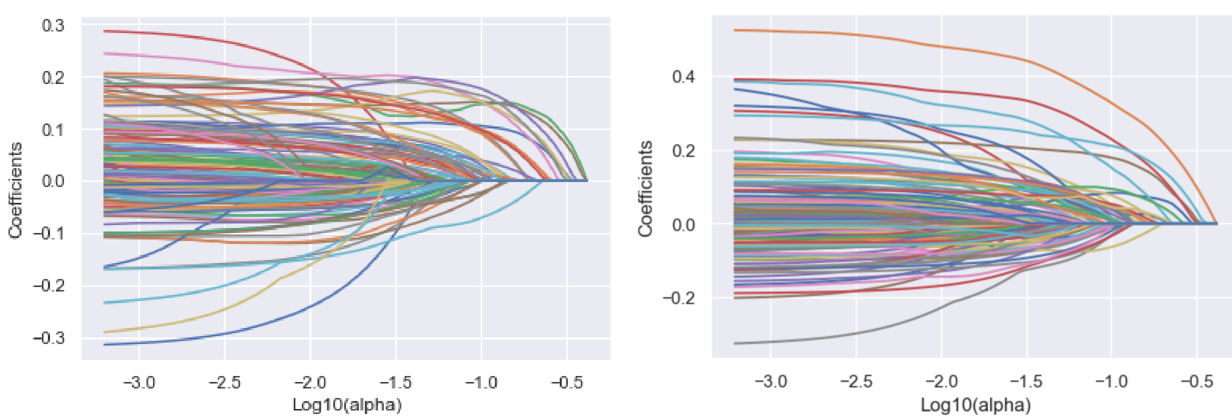
Training the Model

Figure 6 shows the lasso regularization path of candidate models $\hat{\beta}_1 \dots \hat{\beta}_t$ that we obtained by minimizing Equation 2 over a sequence of tuning parameters $\lambda_1 < \lambda_2 < ... < \lambda_T$ with on-the-job training and open education as dependent variable respectively. The vertical axis contains different levels of $\hat{\beta}$. The horizontal axis contains different levels for λ. Each vertical section along the horizontal axis, represents one candidate model. Moving from higher to lower λ, the algorithm includes more nonzero $\hat{\beta}_k$ and the model becomes more complex. To find the optimal value for λ we use *5-fold cross- validation*. We split the data into five random, evenly sized subsets and derive the lasso paths $\hat{\beta}_1^k \ldots \hat{\beta}_T^k$ on each of the folds but the fifth fold to train the models. Then we use the left out fifhts fold to obtain the out-of-sample error for each candidate model. The $\hat{\lambda}_t$ that minimizes the out-of-sample error is selected as the optimal $\hat{\lambda}_t$.

Figure 6: Linear model: Lasso path

(a) On-the-job training

(b) Open education



For the dependent variable on-the-job training, we obtain the optimal $\lambda = 0.0032$. For open education, we obtain the same optimal $\lambda = 0.0032$.

Variable Selection

When estimating the linear model using the optimal Lasso parameter for the number of on-job trainings a person has participated in in the past year, the Lasso regularization identifies 148 columns[4] with non-zero predictive power. The coefficients are presented in Table A1. We distinguish between four different skill levels: elementary occupations (*skill_4*), semi skilled blue-collar workers (*skill_3*), semi-skilled white-collar workers (*skill_2*), and skilled occupations (*skill_1*). The

[4] Including country and industry controls

reference group is *skill_4* which captures elementary occupations. All skill dummies coefficients are non-zero in the Lasso model. Individuals working in high-skilled jobs have fewer trainings than individuals in elementary jobs. Working in semi-skilled jobs reduces the number of on-the-job trainings as well but less strongly than for high-skilled jobs. Individuals in semi-skilled blue-collar occupations have the least reduction in the number of trainings compared to working in elementary occupations.

The number of on-the-job trainings is the strongest affected positively the higher the number of years the individual has worked. An increase of the working years (*c_q09*) by 4 years results in an increase of the trainings by approximately one training. Also, the higher the educational degree and the longer the individual had to study to get the current job, the higher the number of trainings she received in the last year (*yrsget* and *edcat8_Tertiary – research degree (ISCED 6)*). Jobs in which individuals have to keep up to date with new products or developments every day, *d_q13c_Every day*, imply that the individual participate in more on-the-job trainings. A higher willingness to learn or the wish to increase their skills and knowledge or if the respondent perceives her formal education as very useful for the current job increases the amount of trainings. Another positive effect can be observed if the individual works in a larger firm and if she supervises other employees.

Never using a programming language at work, *g_q05g_Never*, while using the internet for research purposes every day, *g_q05c_Every day*, have positive effects on the number of on-the-job trainings.

When increasing the age of the individual, the number of on-job trainings participated in in the last year decreases. The number of trainings is also negatively correlated with the straightforward usage of a PC at work in general, *g_q06_Straightforward* and *g_q04_Yes*, and with the usage of spreadsheed software daily for the job, *g_q05e_ every day*, specifically. Furthermore, if the job requires to keep up to date with new products or services less than once in a month (*d_q13c_Less than once a month*) the number of trainings are lower. A lower education and the subjective feeling that the formal education is not useful for the current job further decrease the number of trainings an individual received in the past year.

Moving on to the number of open courses the individuals participated in in the past year, the linear regression using lasso regularization comprises 150 coefficients[5] . The coefficients indicate that working in a high-skilled job, *skill_1*, or working in a semi-skilled blue-collar occupation increase the number of open courses compared to working in elementray occupations. Working in a semi-skilled white-collar occupation, *skill_2*, has a small negative effect on the number of open courses.

We observe the strongest increase in the number of open courses if the employee never uses spreadsheets at work, *g_q05e_Never*, and if the employee uses a computer in general for her work, *g_q04_Yes*. An increase in the number of years the individual had been employed by 3 years, increases the number of off-job trainings by one training (*c_q09*). Individuals who are working in firms with more than 1000 people are also participating in 0.28 more trainings than individuals who are working in smaller firms (*d_q06a_More than 1000 people*). In general, individuals with lower educational qualifications and who have no contract or are serving an apprenticeship participate in more open courses. Nonetheless, the willingness to learn is also an important factor for the number of open courses as well as the fear of loosing the current job.

As for on-job trainings, we observe a negative correlation between the number of open courses and age of the individual. An increase in age by 10 years reduces the number of trainings by 1.5 trainings. We observe another negative effect if the individual seldom or never uses the PC at work, e.g. never conducts work-related transactions over the internet, *g_q05d_Never* or if the individual never participates in real-time discussions on the internet as the negative coefficient of *g_q05h_Never* indicates. If the individual perceives her formal education as somewhat useful for the current job, *b_q10c_Somewhat useful*, the number of open courses decreases.

Overall, the regression results indicate that individuals with a higher education as well as the need to always be up to date and using the computer daily participate in more on-the-job trainings and open courses. But the results also imply that the number of trainings is reduced if the individual works in a high-skilled job. This is somewhat counterintuitive also when recalling the distribution of trainings among job classifications. Figure 3 clearly indicates that individuals working in high-skilled jobs receive more trainings. These discrepancies suggest that the model, although regularized, does not yield a good fit. The perfomance of hte models will be analyzed in the next step.

[5] Including country and industry controls

Out-of-Sample Evaluation Results

In this section, we show the performance of the Lasso regularized linear regression compared to an Ordinary-Least-Squares (OLS) regression. We compare the out-of-sample deviance, R^2, to evaluate the performance of the models. The out-of-sample deviance is the deviance between the model's prediction after being trained on the training data and the new data, namely the test data (see eq:OOS). The lower the out-of-sample deviance, the higher is the out-of-sample performance and thus the better the model (Taddy, 2019).

$$dev_{OOS}(\hat{\beta}) = \sum_i (n+i)^{n+m}(y_i - x_i'\hat{\beta})^2 \tag{3}$$

For the linear models, we can see in Figure 7 that the R^2 for the Lasso regularized model has a higher mean than the OLS model. Hence, the out-of-sample performance of model with the Lasso penalty is better as it allows for less noise in the data and reduces overfitting. The same holds for the models for off-job trainings as we show in Figure 7. The out-of-sample performance of the regularized model is higher than that of the linear model without regularization. This indicates that the regularization achieves a better out-of-sample fit than the OLS model. The regularized models achieve higher accuracies compared to the OLS models. Table 4 shows that for on-the-job trainings, the Lasso linear model achieves a test accuracy of 0.04921 which is slightly higher than the linear model that yields 0.04696 test accuracy. Turning to open courses it is also evident that the regularized model performs better on new data than the unregularized model. However, both the test accuracy of the OLS model and the Lasso regularized model are negative. This implies that the null model performs better than the model that includes regressors. This is line with the counterintuitive regression results in the previous chapter. Hence, in a next step we will apply a a logistic model as this might fit the data better than a linear model.

Figure 7: Linear model: Out-of-sample validation

(a) On-the-job training (b) Open education

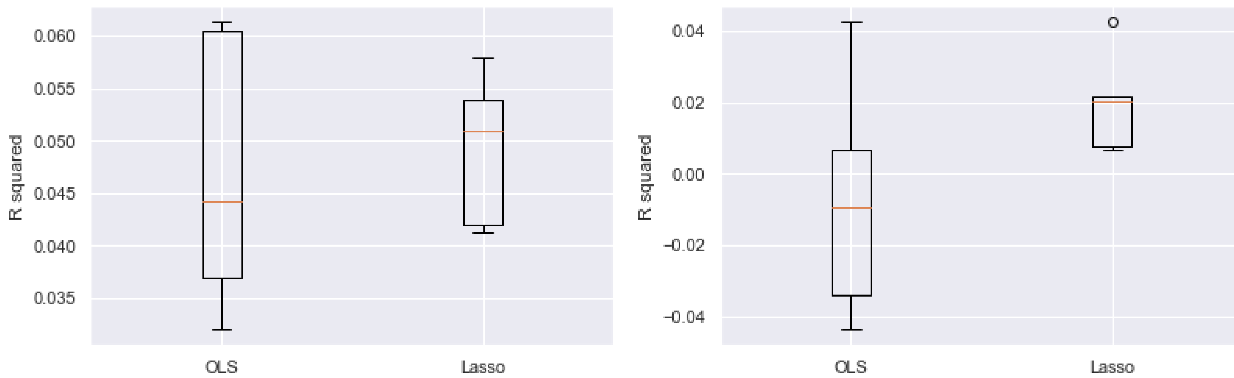


Table 4: Accuracy of the Lasso linear model

	Lasso linear model	OLS
On-job training		
Training accuracy	0.0755	0.0758
Test accuracy	0.0467	0.0441
Open education		
Training accuracy	0.0967	0.0989
Test accuracy	-0.0071	-0.0096

Conclusio and Future Research

This paper illustrates the use of Lasso regulariation to identify characteristics that influece the probability of receiving training. It is often claimed thate mployees who i.e. worked in non-programming jobs before could try to improve their job opportunties by reskilling and acquiring programming skills using MOOC (Garrido et al., 2016; World Economic Forum, n.d.). However, we find that people in higher skilled occupation, with a higher educational level, and who require computer knowledge generally receive more training. Thus trainings cannot be seen as a stepping stone to move from middle skilled occupations to higher skilled occupations. Both on-the-job traings and open education serve highly skilled workers. This supports our hypothesis that open courses are unable to provide extensive retraining as they already require a decent level of ICT skills.

Applying a Lasso logistic model to analyze training partcipation represent a great starting point for future research. Due to the lockdowns and economic shutdowns during the Covid-19 pandemic larger as well as smaller businesses are facing bancruptcy or had to reduce their workforce or limit the number of new hirings to reduce costs (Barrero et al., 2020). The pandemic also pushed digitization and the adoption of robots and thus replacing workforce (Zeng et al., 2020). Especiall the first lock-down has led to a massive increase in Google searches for MOOCs. The aim of our future research is to find out whether this demand shock led to any changes in the characteristics of people that partcipate in trainings. The OECD will launch a second cycle of the PIAAC suvey for 2022 to 2023. With this data, we can compare pre- and post-Covid-19 trends in training participation.

Appendix A

Table A1: Lasso linear regression for on-job training

	Coefficients	Feature
0	-0.292754	age_r
1	-0.0043783	j_q03b
2	0.179486	yrsget
3	0.266616	c_q09
4	0.11098	c_q10a
5	0.150865	readytolearn
6	0.052781	earnmthallppp
7	-0.159132	g_q05e_Every day
8	0.0204591	g_q05e_Less than once a month
9	-0.00896565	g_q05e_Less than once a week but at least once a month
10	-0.000838294	g_q05e_Never
11	0.0063086	d_q12b_A lower level would be sufficient
12	-0.00724959	d_q12b_This level is necessary
13	0.0179174	vet_True
14	0.0136302	g_q05a_Every day

	Coefficients	Feature
15	0.00939957	g_q05a_Less than once a week but at least once a month
16	0.0189398	g_q05a_Never
17	0.0761009	f_q07b_Yes
18	0.103459	g_q05c_Every day
19	-0.00483822	g_q05c_Less than once a month
20	-0.0369135	g_q05c_Less than once a week but at least once a month
21	-0.0173969	g_q05c_Never
22	0.0555923	g_q05d_Every day
23	-0.0223127	g_q05d_Less than once a month
24	0.0493511	g_q05d_Less than once a week but at least once a month
25	0.0759526	g_q05d_Never
26	-0.0183738	d_q12c_1 to 6 months
27	0.0460864	d_q12c_3 years or more
28	-0.0368091	d_q12c_7 to 11 months
29	-0.0568734	d_q12c_Less than 1 month
30	0.0301412	d_q12c_None
31	-0.0418013	edcat8_Post-secondary, non-tertiary (ISCED 4A-B-C)
32	-0.000107181	edcat8_Primary or less (ISCED 1 or less)
33	0.0606026	edcat8_Tertiary - bachelor/master/research degree (ISCED 5A/6)
34	-0.0430088	edcat8_Tertiary – bachelor degree (ISCED 5A)
35	0.0372705	edcat8_Tertiary – master degree (ISCED 5A)
36	-0.105282	edcat8_Tertiary – professional degree (ISCED 5B)
37	0.0992461	edcat8_Tertiary – research degree (ISCED 6)
38	-0.0647511	g_q04_Yes
39	0.0716229	g_q05f_Every day
40	-0.0312493	g_q05f_Less than once a month
41	-0.0510465	g_q05f_Less than once a week but at least once a month
42	0.0512698	g_q05f_Never
43	-0.0396136	leaver1624_Not in education, did not complete ISCED 3, aged 16 to 24
44	-0.00988211	computerexperience_Yes
45	0.0826479	b_q10a_Yes
46	0.125839	d_q13c_Every day
47	-0.162642	d_q13c_Less than once a month
48	-0.0942772	d_q13c_Less than once a week but at least once a month
49	-0.0514091	d_q13c_Never

	Coefficients	Feature
50	0.0399117	g_q05h_Every day
51	0.0795065	g_q05h_Less than once a month
52	0.0984154	g_q05h_Less than once a week but at least once a month
53	-0.05333	pared_Neither parent has attained upper secondary
54	0.0151634	b_q14a_Yes
55	-0.00350692	d_q03_The private sector (for example a company)
56	0.166267	d_q03_The public sector (for example the local government or a state school)
57	-0.0947999	b_q10c_Not useful at all
58	-0.0271457	b_q10c_Somewhat useful
59	0.149894	b_q10c_Very useful
60	-0.0985501	b_q01b_Engineering, manufacturing and construction
61	-0.0380066	b_q01b_General programmes
62	0.186284	b_q01b_Health and welfare
63	-0.0499852	b_q01b_Humanities, languages and arts
64	0.00095563	b_q01b_Science, mathematics and computing
65	-0.01057	b_q01b_Services
66	-0.0306996	b_q01b_Social sciences, business and law
67	-0.0795014	g_q06_Straightforward
68	0.00205237	j_q04a_Yes
69	0.083128	d_q06a_251 to 1000 people
70	0.0250342	d_q06a_51 to 250 people
71	0.0952915	d_q06a_More than 1000 people
72	0.0363452	g_q08_Yes
73	0.0472953	d_q06b_Increased
74	0.0024297	d_q06b_Stayed more or less the same
75	0.00693864	g_q05g_Every day
76	0.0336911	g_q05g_Less than once a month
77	-0.000745596	g_q05g_Less than once a week but at least once a month
78	0.145367	g_q05g_Never
79	0.0310869	d_q09_A temporary employment agency contract
80	0.0502096	d_q09_An apprenticeship or other training scheme
81	0.0118571	d_q09_An indefinite contract
82	0.0123539	d_q09_No contract
83	0.0340834	d_q09_Other
84	0.0232184	b_q14b_Other

	Coefficients	Feature
85	0.00987163	b_q14b_To be less likely to lose my job
86	0.0837637	b_q14b_To do my job better and/or improve career prospects
87	0.0879889	b_q14b_To increase my knowledge or skills on a subject that interests me
88	-0.00636807	b_q14b_To increase my possibilities of getting a job, or changing a job or profession
89	-0.011668	b_q14b_To obtain a certificate
90	0.0132574	b_q14b_To start my own business
91	-0.00414979	gender_r_Male
92	0.150531	b_q26a_t_Yes
93	0.038162	d_q04_t_Employee, supervising fewer than 5 people
94	0.177062	d_q04_t_Employee, supervising more than 5 people
95	0.126982	d_q14_Extremely satisfied
96	0.0739403	d_q14_Neither satisfied nor dissatisfied
97	0.109439	d_q14_Satisfied
98	-0.245979	skill_1
99	-0.200437	skill_2
100	-0.0343195	skill_3

Table A2: Lasso linear regression for off-job training

	Coefficients	Feature
0	-0.151999	age_r
1	-0.0138795	j_q03b
2	0.025427	yrsget
3	0.28891	c_q09
4	-0.0136086	c_q10a
5	0.224639	readytolearn
6	-0.0622156	earnmthallppp
7	0.0364067	g_q05e_Every day
8	0.216647	g_q05e_Less than once a month
9	0.0320373	g_q05e_Less than once a week but at least once a month
10	0.300763	g_q05e_Never
11	0.0360178	f_q07a_Yes
12	0.0791551	d_q12b_A lower level would be sufficient
13	0.0456419	d_q12b_This level is necessary
14	-0.0798626	vet_True

	Coefficients	Feature
15	-0.0307748	g_q05a_Every day
16	-0.0233737	g_q05a_Less than once a month
17	-0.117025	g_q05a_Less than once a week but at least once a month
18	-0.0186612	g_q05a_Never
19	0.0356576	g_q05c_Every day
20	0.00287966	g_q05c_Less than once a month
21	-0.0971144	g_q05c_Less than once a week but at least once a month
22	-0.00899534	g_q05c_Never
23	-0.127509	g_q05d_Every day
24	-0.0471858	g_q05d_Less than once a month
25	-0.093704	g_q05d_Less than once a week but at least once a month
26	-0.298242	g_q05d_Never
27	0.0262656	d_q12c_1 to 6 months
28	-0.0524124	d_q12c_3 years or more
29	-0.0107216	d_q12c_7 to 11 months
30	-0.0294456	d_q12c_Less than 1 month
31	0.013063	d_q12c_None
32	-0.000973135	edcat8_Post-secondary, non-tertiary (ISCED 4A-B-C)
33	-0.00962121	edcat8_Primary or less (ISCED 1 or less)
34	-0.099439	edcat8_Tertiary - bachelor/master/research degree (ISCED 5A/6)
35	0.0114505	edcat8_Tertiary – bachelor degree (ISCED 5A)
36	0.044422	edcat8_Tertiary – master degree (ISCED 5A)
37	-0.113532	edcat8_Tertiary – professional degree (ISCED 5B)
38	-0.0885221	edcat8_Tertiary – research degree (ISCED 6)
39	0.191632	edcat8_Upper secondary (ISCED 3A-B, C long)
40	0.298119	g_q04_Yes
41	0.111086	g_q05f_Less than once a month
42	-0.0641919	g_q05f_Less than once a week but at least once a month
43	-0.146666	g_q05f_Never
44	0.0449273	leaver1624_Not in education, did not complete ISCED 3, aged 16 to 24
45	0.139708	b_q10a_Yes
46	0.146047	d_q13c_Every day
47	-0.0845384	d_q13c_Less than once a month
48	0.0402694	d_q13c_Less than once a week but at least once a month
49	-0.0427484	d_q13c_Never

	Coefficients	Feature
50	0.0416819	g_q05h_Every day
51	-0.110962	g_q05h_Less than once a month
52	-0.0598274	g_q05h_Less than once a week but at least once a month
53	-0.18279	g_q05h_Never
54	0.00104829	pared_At least one parent has attained tertiary
55	-0.130121	pared_Neither parent has attained upper secondary
56	0.0399298	b_q14a_Yes
57	-0.0199741	d_q03_The private sector (for example a company)
58	0.0603001	d_q03_The public sector (for example the local government or a state school)
59	-0.0634292	b_q10c_Not useful at all
60	-0.183465	b_q10c_Somewhat useful
61	0.0237475	b_q10c_Very useful
62	0.0670203	b_q01b_Engineering, manufacturing and construction
63	-0.0135331	b_q01b_General programmes
64	0.210205	b_q01b_Health and welfare
65	-0.0246535	b_q01b_Humanities, languages and arts
66	0.160249	b_q01b_Science, mathematics and computing
67	0.0947835	b_q01b_Services
68	0.0288766	b_q01b_Social sciences, business and law
69	-0.0328424	g_q06_Moderate
70	0.00554308	j_q04a_Yes
71	0.0296164	d_q06a_11 to 50 people
72	0.0762181	d_q06a_251 to 1000 people
73	0.0771511	d_q06a_51 to 250 people
74	0.283448	d_q06a_More than 1000 people
75	-0.118146	g_q08_Yes
76	0.071801	d_q06b_Increased
77	-0.0112068	d_q06b_Stayed more or less the same
78	-0.0360649	g_q05g_Every day
79	-0.0360528	g_q05g_Less than once a month
80	-0.0200237	g_q05g_Less than once a week but at least once a month
81	-0.0114513	d_q09_A temporary employment agency contract
82	0.111856	d_q09_An apprenticeship or other training scheme
83	0.0812359	d_q09_An indefinite contract
84	0.151203	d_q09_No contract

	Coefficients	Feature
85	0.0240714	d_q09_Other
86	-0.0507221	b_q14b_Other
87	0.144671	b_q14b_To be less likely to lose my job
88	0.0289398	b_q14b_To do my job better and/or improve career prospects
89	0.0662145	b_q14b_To increase my knowledge or skills on a subject that interests me
90	0.0407479	b_q14b_To increase my possibilities of getting a job, or changing a job or profession
91	-0.0584367	b_q14b_To obtain a certificate
92	-0.0610496	b_q14b_To start my own business
93	0.0124119	gender_r_Male
94	0.13646	b_q26a_t_Yes
95	-0.0370352	d_q04_t_Employee, supervising fewer than 5 people
96	0.0892785	d_q04_t_Employee, supervising more than 5 people
97	0.0186283	d_q14_Extremely dissatisfied
98	0.0331462	d_q14_Extremely satisfied
99	0.0569407	d_q14_Neither satisfied nor dissatisfied
100	0.122717	skill_1
101	-0.00655556	skill_2
102	0.0684937	skill_3

Table A3: Lasso logistic regression for on-job training

	Coefficients	Feature
0	-0.0173091	age_r
1	0.00648863	j_q03b
2	0.0323998	yrsget
3	0.014883	c_q09
4	0.00407624	c_q10a
5	0.0688914	readytolearn
6	-4.95059e-08	earnmthallppp
7	0.389253	computerexperience_Yes
8	0.152316	d_q09_A temporary employment agency contract
9	-0.0685864	d_q09_An apprenticeship or other training scheme
10	0.000118794	d_q09_An indefinite contract
11	-0.249488	d_q09_No contract
12	0.268225	d_q09_Other

	Coefficients	Feature
13	0.0919625	f_q07a_Yes
14	-0.00312127	b_q01b_Engineering, manufacturing and construction
15	-0.079488	b_q01b_General programmes
16	0.192744	b_q01b_Health and welfare
17	-0.0830876	b_q01b_Humanities, languages and arts
18	0.0623138	b_q01b_Science, mathematics and computing
19	-0.0454431	b_q01b_Services
20	-0.00709438	b_q01b_Social sciences, business and law
21	0.0955665	b_q01b_Teacher training and education science
22	0.0628176	d_q06b_Increased
23	-0.0100635	d_q06b_Stayed more or less the same
24	0.0822428	d_q04_t_Employee, supervising fewer than 5 people
25	0.112577	d_q04_t_Employee, supervising more than 5 people
26	-0.162508	g_q08_Yes
27	-0.119125	pared_At least one parent has attained tertiary
28	-0.0624561	pared_Neither parent has attained upper secondary
29	-0.0492441	gender_r_Male
30	-0.655047	leaver1624_Not in education, did not complete ISCED 3, aged 16 to 24
31	-0.00587444	d_q13c_Every day
32	-0.132742	d_q13c_Less than once a month
33	-0.159028	d_q13c_Less than once a week but at least once a month
34	-0.375997	d_q13c_Never
35	-0.160666	j_q04a_Yes
36	0.343026	edcat8_Post-secondary, non-tertiary (ISCED 4A-B-C)
37	-0.162479	edcat8_Primary or less (ISCED 1 or less)
38	0.128055	edcat8_Tertiary - bachelor/master/research degree (ISCED 5A/6)
39	0.350122	edcat8_Tertiary – bachelor degree (ISCED 5A)
40	0.31369	edcat8_Tertiary – master degree (ISCED 5A)
41	0.165938	edcat8_Tertiary – professional degree (ISCED 5B)
42	-0.152984	edcat8_Tertiary – research degree (ISCED 6)
43	0.0797353	edcat8_Upper secondary (ISCED 3A-B, C long)
44	0.160175	g_q07_Yes
45	0.00541096	vet_True
46	0.100965	g_q05d_Every day
47	-0.104847	g_q05d_Less than once a month

	Coefficients	Feature
48	0.0893642	g_q05d_Less than once a week but at least once a month
49	0.0191254	g_q05d_Never
50	-0.308108	d_q14_Extremely dissatisfied
51	-0.00958417	d_q14_Extremely satisfied
52	-0.104017	d_q14_Neither satisfied nor dissatisfied
53	-0.0803196	d_q14_Satisfied
54	0.0423771	g_q05a_Every day
55	-0.0825071	g_q05a_Less than once a month
56	0.108813	g_q05a_Less than once a week but at least once a month
57	-0.0223103	g_q05a_Never
58	0.153748	g_q05e_Every day
59	0.10127	g_q05e_Less than once a month
60	0.118944	g_q05e_Less than once a week but at least once a month
61	0.162112	g_q05e_Never
62	0.0906379	g_q05f_Every day
63	-0.0120091	g_q05f_Less than once a month
64	-0.105233	g_q05f_Less than once a week but at least once a month
65	-0.0498476	g_q05f_Never
66	-0.151607	g_q05g_Every day
67	0.0534463	g_q05g_Less than once a month
68	0.0319572	g_q05g_Less than once a week but at least once a month
69	-0.0987864	g_q05g_Never
70	0.00587424	g_q05h_Every day
71	-0.00152421	g_q05h_Less than once a month
72	-0.0529479	g_q05h_Less than once a week but at least once a month
73	-0.390016	g_q05h_Never
74	0.245238	f_q07b_Yes
75	-0.125719	b_q10c_Not useful at all
76	0.431681	b_q10c_Somewhat useful
77	-0.0221019	b_q10c_Very useful
78	-0.0368781	d_q12c_1 to 6 months
79	-0.0163393	d_q12c_3 years or more
80	-0.137065	d_q12c_7 to 11 months
81	-0.277454	d_q12c_Less than 1 month
82	0.0594719	d_q12c_None

	Coefficients	Feature
83	0.0659232	g_q05c_Every day
84	-0.126711	g_q05c_Less than once a month
85	-0.0997177	g_q05c_Less than once a week but at least once a month
86	-0.250353	g_q05c_Never
87	-0.123279	d_q12b_A lower level would be sufficient
88	-0.177648	d_q12b_This level is necessary
89	0.111568	d_q06a_11 to 50 people
90	0.280167	d_q06a_251 to 1000 people
91	0.22403	d_q06a_51 to 250 people
92	0.293531	d_q06a_More than 1000 people
93	-0.277048	d_q03_The private sector (for example a company)
94	-0.1113	d_q03_The public sector (for example the local government or a state school)
95	0.675257	g_q04_Yes
96	0.801725	b_q10a_Yes
97	0.0374759	g_q06_Moderate
98	-0.0442568	g_q06_Straightforward
99	0.101065	skill_1
100	0.00415191	skill_2
101	0.00447484	skill_3

Table A4: Lasso logistic regression for off-job training

	Coefficients	Feature
0	-0.0176953	age_r
1	0.0474509	yrsget
2	0.0131248	c_q09
3	0.00448319	c_q10a
4	0.0861567	readytolearn
5	-3.07033e-08	earnmthallppp
6	0.220386	computerexperience_Yes
7	-0.21247	d_q09_No contract
8	0.0751546	f_q07a_Yes
9	-0.0165669	b_q01b_General programmes
10	0.179973	b_q01b_Health and welfare
11	0.0139165	b_q01b_Science, mathematics and computing

	Coefficients	Feature
12	0.0940462	b_q01b_Teacher training and education science
13	0.0570971	d_q06b_Increased
14	-0.00782241	d_q06b_Stayed more or less the same
15	0.0140561	d_q04_t_Employee, supervising fewer than 5 people
16	0.0829086	d_q04_t_Employee, supervising more than 5 people
17	-0.0656784	g_q08_Yes
18	-0.042624	pared_At least one parent has attained tertiary
19	-0.024773	gender_r_Male
20	0.048559	d_q13c_Every day
21	-0.0795516	d_q13c_Less than once a month
22	-0.0931521	d_q13c_Less than once a week but at least once a month
23	-0.274087	d_q13c_Never
24	-0.0359449	j_q04a_Yes
25	0.162247	edcat8_Post-secondary, non-tertiary (ISCED 4A-B-C)
26	0.252905	edcat8_Tertiary – bachelor degree (ISCED 5A)
27	0.157216	edcat8_Tertiary – master degree (ISCED 5A)
28	0.0540491	edcat8_Tertiary – professional degree (ISCED 5B)
29	-0.0268396	edcat8_Tertiary – research degree (ISCED 6)
30	0.15142	g_q07_Yes
31	0.0613931	g_q05d_Every day
32	-0.0462954	g_q05d_Less than once a month
33	0.0218131	g_q05d_Less than once a week but at least once a month
34	0.0188032	d_q14_Extremely satisfied
35	0.0423613	g_q05a_Every day
36	0.0798542	g_q05e_Every day
37	0.0237213	g_q05e_Never
38	0.121532	g_q05f_Every day
39	0.0351061	g_q05g_Less than once a month
40	-0.00863865	g_q05g_Never
41	0.0140042	g_q05h_Every day
42	-0.352158	g_q05h_Never
43	0.259307	f_q07b_Yes
44	0.107441	b_q10c_Somewhat useful
45	-0.056507	d_q12c_7 to 11 months
46	-0.141571	d_q12c_Less than 1 month

	Coefficients	Feature
47	0.0394685	d_q12c_None
48	0.141345	g_q05c_Every day
49	-0.137027	g_q05c_Never
50	-0.0905359	d_q12b_A lower level would be sufficient
51	-0.138258	d_q12b_This level is necessary
52	0.00834744	d_q06a_11 to 50 people
53	0.145794	d_q06a_251 to 1000 people
54	0.106445	d_q06a_51 to 250 people
55	0.162631	d_q06a_More than 1000 people
56	-0.238538	d_q03_The private sector (for example a company)
57	0.56252	g_q04_Yes
58	0.78252	b_q10a_Yes
59	0.0366449	g_q06_Moderate
60	0.11028	skill_1

References

Acemoglu, D., Autor, D.H., 2011. Skills, tasks and technologies: Implications for employment and earnings, in: Handbook of Labor Economics. Elsevier, pp. 1043–1171.

Acemoglu, D., Pischke, J.-S., 1999. Beyond becker: Training in imperfect labour markets. The economic journal 109, 112–142.

Autor, D.H., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the us labor market. American Economic Review 103, 1553–97.

Barrero, J.M., Bloom, N., Davis, S.J., 2020. Covid-19 is also a reallocation shock. National Bureau of Economic Research.

Becker, G.S., 1962. Investment in human capital: A theoretical analysis. Journal of political economy 70, 9–49.

De La Rica, S., Gortazar, L., Lewandowski, P., 2020. Job tasks and wages in developed countries: Evidence from piaac. Labour Economics 65, 101845.

Garrido, M., Koepke, L., Andersen, S., Mena, A., Macapagal, M., Dalvit, L., 2016. An examination of mooc usage for professional workforce development outcomes in colombia, the philippines,

& south africa. Seattle: Technology & Social Change Group, University of Washington Information School.

GESIS - Leibniz Institute for the Social Sciences, n.d. Programme for the international assessment of adult competencies (piaac).

Jung, I. (Ed.), 2019. Open and distance education theory revisited: Implications for the digital era, 1st ed. 2019. ed, SpringerBriefs in open and distance education. Springer Singapore; Imprint: Springer, Singapore. https://doi.org/10.1007/978-981-13-7740-2

Konings, J., Vanormelingen, S., 2015. The impact of training on productivity and wages: Firm-level evidence. Review of Economics and Statistics 97, 485–497.

Lynch, L.M., 1992. Private-sector training and the earnings of young workers. The American Economic Review 82, 299–312.

Lynch, L.M., 1991. The role of off-the-job vs. On-the-job training for the mobility of women workers. The American Economic Review 81, 151–156.

Taddy, M., 2019. Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw-Hill Education, New York.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

World Economic Forum, n.d. The future of jobs report 2020.

Zeng, Z., Chen, P.-J., Lew, A.A., 2020. From high-touch to high-tech: COVID-19 drives robotics adoption. Tourism Geographies 22, 724–734.