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

What is the topic of your dissertation? What were the reasons for choosing this topic? What is your research hypothesis? How do you define the central terms of your hypothesis? Why and for whom is it important to answer these questions?

- Motivate (one paragraph).
- Summarize what we know (one paragraph).
- Third paragraph: Tell us what you are doing!
- Describe research design.
- Value added.
- Summarize key findings.
- DO NOT write a "roadmap".
- No subsections in the Introduction

Literature

(We contribute to the literature on digitalization, job tasks, training, and job mobility.)

Computer capital and workers that perform routine tasks are substitutes whereas computer capital and workers that perform non-routine cognitive tasks are complements (Autor et al., 2003). The declining price of computer capital has led to a U-shaped labor demand function (Acemoglu and Autor, 2011; Autor and Dorn, 2013). This indicates that middle-skilled workers were replaced by technology while the demand for high skilled workers and low-skilled workers has grown. 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) develop abstract, routine and manual task measures and 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.

This vast literature on the change of job tasks has mainly focused on wage effects and takes skills of workers as pre-defined. However, workers and firms could also invest in new skills via training.

Becker (1962) distinguished between two kinds of on-the-job 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. 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 work-ers 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.

Evidence on job mobility behaviour of workers is more mixed. Zweimüller et al. (2003) findings support Becker (1962) human capital theory. Workers who received firm specific training quit less often and show less job searching behaviour. Workers who received general training increased their job searching activities and quit more often. Dietz and Zwick (2020) use German employer-employee data and find that training increases the retention probability. These studies focus on on-the-job training.

Lynch (1991) and Lynch (1992) compares on-the-job to off-the-job training. She focuses on young workers that are particularly mobile. She finds 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. Lynch (1991) shows that the probability of leaving an employer varies with respect to race, gender, and educational level. Workers with disabilities, black workers and workers with a high school degree or less increased the probabil-

ity of leaving the first employer. Working in a job with collective agree-ment or having a college degree decreased their probability of leaving the employer. The effect of training, disability, and education disappears when Lynch (1991) re-estimates the equation only for men, while these effects are particularly strong for women.

Applying a machine learning approach permits us to take a broader approach on this topic. Instead of restricting our estimation to a specific group of workers or countries, we identify the factors that drive the probability of leaving-a-job and job-switching from a set of ## variables.

- mostly effect on wages, inequality

Data

What is the epistemological framework of the dissertation? For empirical studies it should be made clear: Why were the specific methods of data analysis chosen? How was the data acquired?

To explore these questions 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 while also comprising comprehensive background information on the respondents past and current education, subjective assessments of their skills and job requirements as well as information on migration (GESIS - Leibniz Institute for the Social Sciences (n.d.)). The first cycle consisted of three rounds and 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 work-

ers 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 130 variables in total, including indices. For example, information on the various test results conducted in the study are excluded. To achieve comparability across countries, questions which were only answered by respondents living in the United States are also excluded. The final dataset is cross-sectional with one observation representing the answers of one respondent.

The 130 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 skills and the respective extensive and intensive margin, her subjective job requirements, information on her curent job and information on monthly income. 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 2).

The classification of the respondents jobs in terms of skills is also 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 Fiugre 3).

The key variables of this study are the the skill classification of the individual's job and her trainings comprising on-the-job training, seminars or workshops, distance or open training courses as well as private lessons. The simple OLS regressions 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, seminars or workshops, distance or open educational training as well as for private lessons. However, for semi-skilled jobs, the picture is slightly different. Here, the number of seminars or workshops and private lessons are positively correlated with the semi-skilled occupations.

- Name, source, unit, time, structure, number of observations, relevant population.
- Definition of (main) sample.

Figure 1: Distribution of Gender

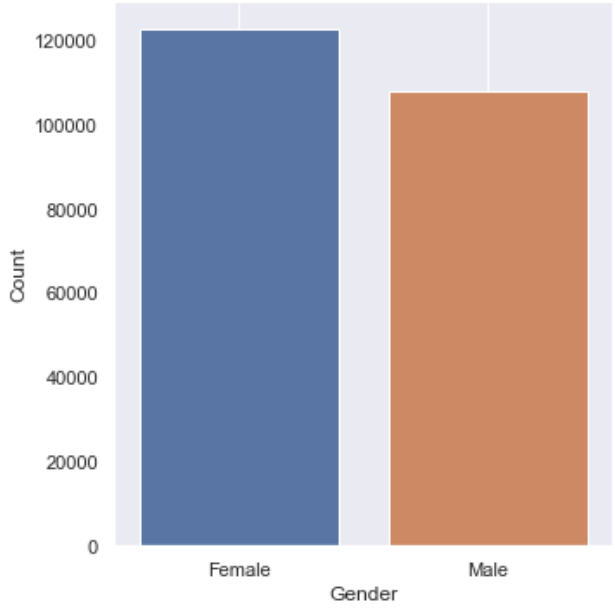


Figure 2: Distribution of Age

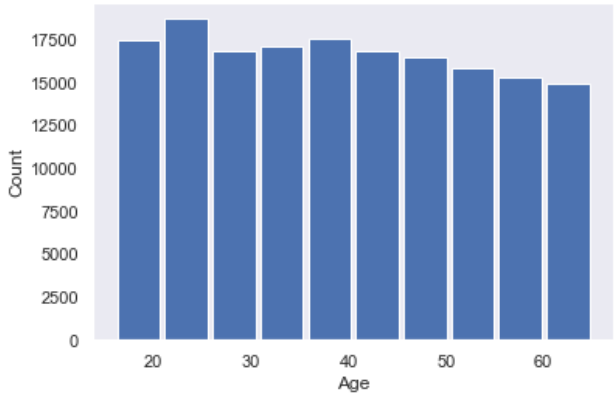
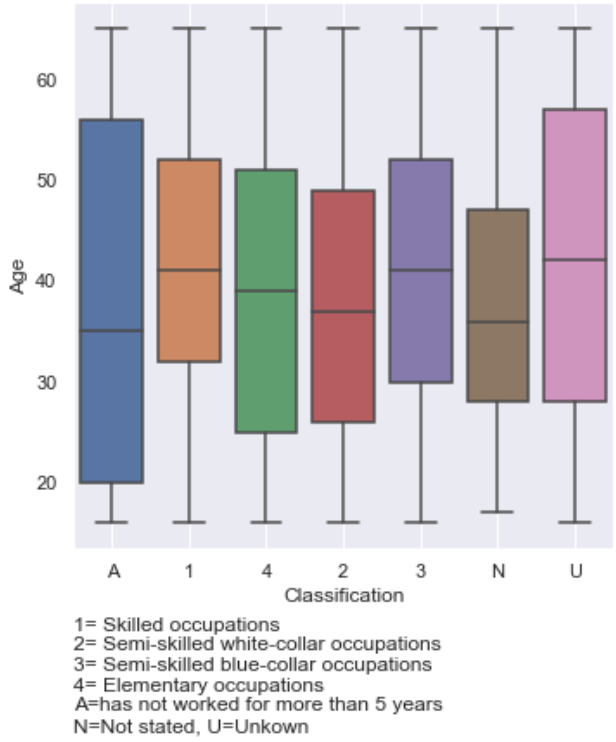


Figure 3: Age and Classification of jobs



- Definition and characteristics of key variables.
- Limitations and potential biases.
- Provide the data and the software code (replication).
- Plot the main empirical associations you want to study!
- Do NOT assume the reader knows anything about these data

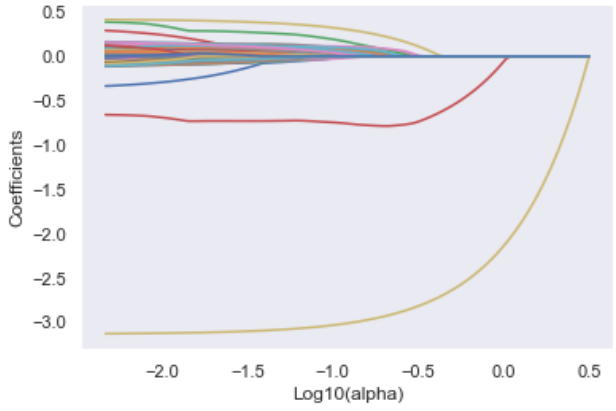
Model and Methods

In order to select the right set of covariates from our set of ..., we apply the Least absolute shrinkage and selection operator (Lasso) that was first proposed by Tibshirani (1996). 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. We add the Lasso penalty equal to $\sum_k |\beta_k|$. The Lasso estimator $\hat{\beta}$ is then given by

$$\hat{\beta} = argmin \sum_i (y_i - x_i^T \beta)^2 + \lambda \sum_k |\beta_k| \tag{1}$$

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. $\lambda > 0$ is the penalty weight or the *tuning parameter*. Figure 4 shows the lasso regularization path of candidate models $\hat{\beta}_1 \ldots \hat{\beta}_t$ that we obtained by minimizing Equation 1 over a sequence of tuning parameters $\lambda_1 < \lambda_2 < ... < \lambda_T$. 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 becomes more complex. To find the optimal value for λ we use *k-fold cross validation*. We split the data in k random evenly sized subset and derive the lasso paths $\hat{\beta}_1^k \ldots \hat{\beta}_T^k$ on each of the folds but the k-th fold to train the models. Then we use the k-th fold to obtain the out-of-sample error for each candidate model. The best $\hat{\lambda}_t$ minimizes the out-of-sample error.

Figure 4: Lasso path for on job training



- Describe how the hypothesis is linked to your estimation.
- Describe the estimation using equations.
- Discuss the parameters and variables.
- What are the identifying assumptions, what are (possible) violations and their consequences?
- What will you do about this?

Results

- Tell a story!
- Guide the reader.
- Focus on the key points, not the details.
- Discuss quality and quantity.
- Discuss problems.
- Compare to the literature.

Further steps

Which results can be expected? What is new? Where lies the progress for science? In what way can scientific discussion proceed / be stimulated by the thesis?

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.

Autor, D.H., Levy, F., Murnane, R.J., 2003. The skill content of recent technological change: An empirical exploration. The Quarterly journal of economics 118, 1279–1333.

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.

Dietz, D., Zwick, T., 2020. The retention effect of training: Portability, visibility, and credibility1. The International Journal of Human Resource Management 1–32.

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

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.

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

Zweimüller, J., Winter-Ebmer, R., others, 2003. On-the-job-training, job search and job mobility. REVUE SUISSE D ECONOMIE ET DE STATISTIQUE 139, 563–576.