

Christl, Michael; Köppl-Turyna, Monika

Working Paper

Gender wage gap and the role of skills: evidence from PIAAC dataset

Agenda Austria Working Paper, No. 05

Provided in Cooperation with:

Agenda Austria, Vienna

Suggested Citation: Christl, Michael; Köppl-Turyna, Monika (2017) : Gender wage gap and the role of skills: evidence from PIAAC dataset, Agenda Austria Working Paper, No. 05, Agenda Austria, Wien

This Version is available at:

<https://hdl.handle.net/10419/163562>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Gender wage gap and the role of skills: evidence from PIAAC dataset

Working Paper – No. 05

Michael Christl,
Monika Köppl-Turyna,
03/2017

Abstract:

Our paper makes a first attempt to address the impact of skills and skill use in the analysis of the gender wage gap using the PIAAC dataset. Using the case of Austria, we show that skill use as well as the skill match play an important role with regard to wage regressions of men as well as women. When we take skills into account in the gender wage gap analysis, the unexplained part of the gender wage gap is reduced by almost 4 percentage points along the whole wage distribution. Our results suggest that skill use and match play a crucial role in explaining the gender wage gap. Additionally, we show, that the self-selection problem biases the results, in particular in the lower and middle parts of the wage distribution and that we should control for it, although the effect is small. When we additionally consider discretionary bonus payments, we find that the unexplained part in the gender wage gap increases, especially in the upper part of the wage distribution.

Keywords:

Gender wage gap, skills, Austria

Gender wage gap and the role of skills: evidence from PIAAC dataset*

Michael Christl^{a,*}, Monika Köppl–Turyna^{a,b}

^a*Agenda Austria, Schottengasse 1/3, 1010 Vienna, Austria*

^b*Vienna University of Economics and Business, Institute for Analytical Economics, Welthandelsplatz 1, 1020 Vienna, Austria*

Abstract

Our paper makes a first attempt to address the impact of skills and skill use in the analysis of the gender wage gap using the PIAAC dataset. Using the case of Austria, we show that skill use as well as the skill match play an important role with regard to wage regressions of men as well as women. When we take skills into account in the gender wage gap analysis, the unexplained part of the gender wage gap is reduced by almost 4 percentage points along the whole wage distribution. Our results suggest that skill use and match play a crucial role in explaining the gender wage gap. Additionally, we show, that the self-selection problem biases the results, in particular in the lower and middle parts of the wage distribution and that we should control for it, although the effect is small. When we additionally consider discretionary bonus payments, we find that the unexplained part in the gender wage gap increases, especially in the upper part of the wage distribution.

JEL Classification: J31, J71

Keywords: gender wage gap, skills, Austria

*Corresponding author. Email: michael.christl@agenda-austria.at

*The authors are thankful to Anna Stürgkh for her excellent research assistance and the participants of the 7th Ifo Dresden Workshop on Labour Economics and Social Policy for their helpful comments.

1. Introduction

The gender wage gap is one of the most vividly discussed topics regarding possibility of wage discrimination. Even though there exist many forms of discrimination with respect to gender, we focus on the potential for discrimination on the labor market. It is well-established that discrimination with regard to other aspects of the society, such as e.g., in education, might additionally lead to discrimination on the labor market. But this is not the topic of our analysis.

Some scholars, e.g. Kunze (2008) warn against interpreting the unexplained gender wage gap as discrimination, as typically one can not measure productivity differences precisely enough, or assume that those differences themselves are not an outcome of discriminatory behavior. Nevertheless, to assure comparability of our results with previous studies, we look at the gender wage gap using a decomposition method. However, we remain cautious when interpreting our results.

In an overview, Weichselbaumer and Winter-Ebmer (2005) show that there is a significant reduction in the gender wage gap over the past 30 years in the OECD countries, and that most of the gender-wage gap stems from differences in characteristics between men and women. This indicates that discrimination on the labor market is probably smaller compared to discrimination in other areas of the society. Moreover, access to more reliable data and more sophisticated econometric techniques allows us to look more closely at the gap than 30 years ago. Indeed, it might be the case, that some part of the conclusion that the gender wage gap is closing is due to ever-better estimation techniques.

Still, the question of labor market discrimination and determinants of differences between the genders is not an easy one. Good-quality data is needed to ensure that all valid characteristics of workers can be covered. That is why, in our study, we look at some factors affecting wages, which have not been covered in previous analyses using a new dataset.

It is often argued that occupational segregation by gender is a main driver of the difference in the pay of men and women, as it has been argued by Blau and Kahn (2000). Even though occupational segregation fell in most developed countries over the last decades according to Blau et al. (2013) and the beneficial effect of technology on the difference between male and

female earnings (see e.g. Black and Spitz-Oener (2010)), the gender wage gap is still persistent. It is of special interest, that according to Goldin (2014), the gender wage gap is especially driven within narrow defined occupations and not across different occupations. Autor and Handel (2013) show that tasks in the US vary substantially within certain occupations between gender and race. If that is the case, controlling for tasks or skill might explain part of the gender wage gap. This question is exactly the motivation of our paper.

It is particularly difficult to find data regarding the skills of workers, over- or under-qualification, and some other characteristics that are obviously decisive for productivity and in consequence for the earned wage. Not including such variables in the gender-wage-gap analysis leads to a bias in the unexplained part of the gender wage gap and automatically results in a wrong estimate of potential wage discrimination.

Additionally, many studies have shown that the unexplained gender wage gap is not equal over the wage distribution. Most papers find that it increases along the wage distribution, indicating that if there exists discrimination against women, it is higher for higher wages than for lower wages. In Austria and Germany, for instance, this finding is hypothesized to stem from binding collective bargaining contracts at the lower end of the wage distribution. On the other hand, an important issue of selection into the workforce affects the estimation of the gender gap at the lower end of the distribution - and it does that in the opposite direction. Without controlling for sample selection, we would compare a representative sample of males with the most productive females, which leads to an underestimation of the size of the wage gap.

In light of the above observations, we make a first attempt to include variables such as the skills of workers, the over- or under-qualification as well as the flexibility of work and presence of small children to account for otherwise not included characteristics. In fact, those characteristics have not been covered in studies of the gender-wage gap in Austria. Taking into account these factors, we show that the estimated wage gap is much smaller than in the previous studies, once specific skills and their utilization, as well as work flexibility are controlled for. Moreover, we control for selection of individuals into workforce and show that wage gap at the lower end of the wage distribution is higher than if we do not control for this

factor. Finally, the our dataset allows us to consider the wage gap also with regards to bonus payments. We exploit the dataset for a single country, as cross-country differences in labor market participation could bias the results - an issue which we discuss in more detail further on.

More generally, further research regarding the gender gap based on the Programme for the International Assessment of Adult Competencies (PIAAC) dataset is highly desirable, as it allows direct control for adult skills and their utilization. To the best of our knowledge, this is the first study which exploits this dataset.

This work is structured as follows: the next section provides a brief overview of the gender gap literature with a focus on previous estimates of the gender gap in Austria. Section 3 presents the dataset and the empirical approach. Section 4 presents the results. Section 5 concludes the paper.

2. Literature overview

While studies concentrate on different features with regard to the gender wage gap, there are several aspects that are more or less common to all of them. The first is the Blinder (1973) and Oaxaca (1973) (type) decomposition, which can be described as the method of splitting the unadjusted gender wage gap into two parts, where one is described as the explained part and the remainder as the unexplained part. The "explained part" implies as much as the part of the gap which can be positively ascribed to differences in certain characteristics, and the other is commonly interpreted as discrimination.

The Blinder and Oxaca approach is based on human capital theory, stating that wages are tied to productivity and an observed male-female wage gap should be explained by differences in productivity between men and women. As a result, most studies mainly consider factors affecting productivity, such as education, work experience, tenure etc.

Any discussion of previous studies can by no means be definitive or exhaustive. Nevertheless, several trends and recent results can be summarized. Different scholars and institutions choose different factors to add to the basic human capital and productivity characteristics that appear in almost all studies. The choice of factors depends on the specific inquiry and

can increase the part that can be explained through regression analyses. Recently, more detailed studies, therefore, tend to exhibit smaller unexplained parts than earlier works and all-encompassing international analyses.

In the first part of this section we focus on recent international studies that are related to our research, while in the second part we discuss the existing literature for Austria.

2.1. International Literature

Boll and Leppin (2015) show that in Germany, the unadjusted gender wage gap of (up to) 23,9% leaves an unexplained part of 2,3% when controlled for various characteristics. Differences in experience, working hours, work status, sector as well as the migration background between men and women are those five factors that explain the most of the gender wage gap in Germany. The authors further break the gap down across the wage distribution, showing that while women in the bottom quantiles experience positive discrimination of 9% (probably through collective agreements), women at the top of the distribution experience an unexplained wage gap of 8%. Blau and Kahn (2016) draw a similar picture for the US.

In a cross-country study, Boll et al. (2016) add variables for over-education, perceived health, country of birth, firm size, existence of supervisory position, marital status and information on the partners labor market characteristics (if available) to the usual characteristics. They find an unadjusted cross-country gap of 18,4%, which they break down into an 11,1% unexplained gap and a 7,2% explained gap. While on this cross-country basis the unexplained part still makes up the majority of the gap, several countries, namely Austria, France, Norway, Serbia and Switzerland, exhibit an unexplained gap of less than 5%.

Cross-country analyses of the gender wage gap exhibit specific problems. Boll et al. (2016) themselves, and other studies, such as Tjstens et al. (2012) draw attention to the fact that low job-market participation rates among females correlate with a small gender wage gap. This is explained by the selection process that results in lower job-market participation for females. In countries with low female participation rates only the most educated and qualified women participate, and find themselves in relatively well-paid jobs.

The question of female labor force participation often hinges on national family policies. A large part of gender gap literature focuses on the question of the impact of children on

parents wages, which is often called the family wage gap. Meurs et al. (2010) examine the impact of child-related career interruptions on women's wages. Studies such as Angelov et al. (2013) have shown no direct impact of the factor "children" on the wages of mothers. Meurs et al. (2010) similarly conclude that it is not the mere presence of a child that has an impact on women's wages, but the child-related career interruptions. In numbers, the authors show a 2.5% penalty for each year out of work for childcare. The 'interruption' gap between men and women lies at 19%. Interestingly, they also find that women, who did not take child-related time out would be paid more than men if they were getting the same returns to their productive characteristics. This study, therefore, seems to confirm what other authors, such as Weeden et al. (2016) have argued, namely that much of what appears to be a gender wage gap is a gender-specific family gap in pay and that most of it could be explained by factors directly or indirectly related to motherhood.

Since having a child might require more flexibility of the job, especially in countries where public childcare is weak, it is often argued that this factor influences the wages. Goldin (2014) looks at BA graduates working full-time and full-year. She shows that most of the gender gap (68%) is due to differences within occupations. She continues to demonstrate that occupations that show nonlinearity in earnings with respect to the time worked, exhibit the highest gender pay gap. This nonlinearity leads to a discrimination with respect to jobs that offer flexibility and employees that look for it. In this case, flexibility of working hours is an important factor for determining the size of the wage gap, and may be taken to discriminate against people, who have to choose a flexible job or are at least strongly encouraged to do so for whatever reason.

2.2. Literature for Austria

The empirical studies with regard to the gender wage gap in Austria are manifold. Typically, quantile regressions are used, but especially in older works such as Böheim et al. (2005) and Böheim et al. (2007), quantile regressions are applied in combination with Oaxaca-Blinder decomposition, which technically speaking is an incorrect approach and typically leads to an

unexplained residual of unknown size.² Böheim et al. (2005) estimate that discrimination caused women to earn between 17% and 18% less on average. During the period between 1983 and 1997, the average difference in wages between men and women that cannot be explained decreased from 18% to 15.5% (using men's wages as reference). Böheim et al. (2007) present similar results.

Böheim et al. (2013a) use a similar approach to the Machado and Mata (2005) decomposition, that can solve the problems arising from the Oaxaca-Blinder decomposition of quantile regressions. This paper shows that women earn about 14 % less than men controlling for certain characteristics and about 50% of the gender wage gap cannot be explained by observed characteristics. They additionally show that the unexplained part increases along the wage distribution.

Grandner and Gstach (2015) use the decomposition method of Machado-Mata on quantile regressions to compare differences in the gender wage gaps in Austria and Germany. They show that for Austria, the gender wage gap is about 20% across all income groups. They also show that the difference in observable characteristics does not explain any of the gender wage gap. This implies that there are no differences in characteristics. This is exactly the opposite of what Böheim et al. (2013a) have found.

Böheim et al. (2013b) use the Oaxaca-Blinder decomposition to analyze the wage gap between 2002 and 2007. They show that, first, about 50% of the gender wage gap can be explained by observable characteristics. Women have on average higher formal education but less work experience (probably due to child-bearing and child-care). Second, they show that the gender wage gap decreased from 21% in 2002 to 16% in 2007.

Grünberger et al. (2009) use a Mincer wage regression method and show that for employees working full time, women earn on average 22% less than men. Controlling for observable characteristics brings the figure down to 12%. They additionally show that the explained part of the gender wage gap is higher in the lower part of the wage distribution.

The most recent papers and results for Austria are summarized in Table 1. All of these

²see e. g. Grandner and Gstach (2015) and Fortin et al. (2011)

studies face some data problems regarding unobserved characteristics such as the skills of workers, the existence of (small) kids, as well as the flexibility of the job. They, except for Grünberger et al. (2009),³ do not control for sample selection.

³Grünberger et al. (2009) use Mincer wage regressions and add in one specification the inverse-Mills ratio to the regressions; they do not control for sample selection for different quantiles of the distribution. In fact, inverse-Mills ratios from a parametric first-stage would be an improper measure in a quantile-regression context.

Table 1: Literature overview for Austria

	Dep. variable	Controls	Unexplained gap	Method	Remark
Böheim et al. (2005)	net earnings per month	education, experience, position, white-collar, hours worked, occupation, segregation, family status, nationality	15.5 percent in 1997	quantile regression, Oaxaca-Blinder	increase over the wage distribution
Böheim et al. (2007)	net earnings per month (FT equivalent)	education, experience, position, career interruptions, white-collar, industry, family status, nationality, city size, region	14 percent in 1997	quantile regression, Oaxaca-Blinder	decrease of discrimination over time
Grünberger et al. (2009)	gross hourly wages (FT)	education, experience, occupation, worker status, industry, family status, nationality, region	12 percent 2006	Mincer wage regression	increase over the wage distribution
Böheim et al. (2013a)	gross hourly wages (FT)	education, experience, occupation, career interruptions, length of maternity leave, worker status, industry, firm size, family status, nationality, city size, region	14 percent in 2007	quantile regression (Melly (2006))	increase over the wage distribution
Böheim et al. (2013b)	gross hourly wage	education, experience, tenure, length of interruptions, family status, citizenship, population density, worker status, firm size, firm characteristics (age, female/male, turnover)	16 percent in 2007	Oaxaca-Blinder	reduction between 2002 and 2007
Grandner and Gstach (2015)	gross hourly wage	education, experience, country of birth, age, firm size, temp. job, public sector, family status	20 percent in 2008	quantile regression, Machado and Mata (2005)	constant over the wage distribution

3. Methodolgy and Data

3.1. The empirical approach

We estimate the gender wage gap using counterfactual distributions approach by Chernozhukov et al. (2013)⁴. Previous studies of the gender gap in Austria relied on the Machado and Mata (2005) approach to estimating counterfactual distributions. However, Machado and Mata (2005) do not present econometric theory for their results. Chernozhukov et al. (2013) propose a more comprehensive approach to estimation of counterfactual distributions. Improving on the existing literature, it allows us to also study the contribution of specific covariates at different parts of the distribution (see, e.g., Depalo et al. 2015). Following the example of Chernozhukov et al. (2013), let $j = 0, 1$ denote the subpopulation of men ($j = 0$) and the subpopulation of women ($j = 1$). Y_j denotes the wages, and X_j is the vector of job-relevant characteristics affecting the wages. Conditional distribution functions $F_{Y_0|X_0}(y|x)$ and $F_{Y_1|X_1}(y|x)$ describe assignment of wages y to individuals with characteristics x . If $F_{Y<0|0>}$ and $F_{Y<1|1>}$ are the observed distributions for men and women, respectively, we can denote

$$F_{Y<0|1>}(y) \equiv \int_{\chi_1} F_{Y_0|X_0}(y|x) dF_{X_1}(x), \quad (1)$$

the counterfactual distribution function of wages that would have prevailed for women had they faced the mens wage schedule, where χ_1 denotes the support of women's characteristics. This distribution is constructed by integrating the conditional distribution of wages for men with respect to the distribution of characteristics for women (Chernozhukov et al. 2013).

The difference between the wage distributions can be then decomposed as

$$F_{Y<1|1>} - F_{Y<0|0>} = [F_{Y<1|1>} - F_{Y<0|1>}] + [F_{Y<0|1>} - F_{Y<0|0>}], \quad (2)$$

where the first term corresponds to the differences in returns, which is a common measure of discrimination, as opposed to the second term, which corresponds to the differences in characteristics. Under the conditional exogeneity assumption, the counterfactual effect can

⁴Implemented in STATA with *cdeco*.

be interpreted as causal.

The estimator used is the linear quantile regression estimator by Koenker and Bassett Jr (1978). The confidence intervals are bootstrapped, using the asymptotic properties by Chernozhukov et al. (2013).

A second methodological issue is that some often-used control variables are endogenous (such as selection into certain activities, full-time work, or simply a decision to start working) is an important factor. If, for instance, reservation wage of a woman depends on her productivity and, therefore, potentially the same factors as the observed wages, the estimates will be biased due to sample selection. In essence, not correcting for sample selection means comparing the wages of a representative sample of males to a censored sample of females, where the censoring implies that the least productive females are not observed. If this is the case, the uncorrected gender gap will be underestimated.

We therefore correct for selection into employment. Austria faces, among comparable European countries, a fairly low labor force participation among women, which can be explained by several factors:

- Traditional division of tasks in the family, i.e., females are expected to take over the major part of household and childcare obligations;
- Comparative under-provision of public childcare institutions, in particular for children of age three and less;
- Generosity of the social system and long maternity leaves (of up to three years).

In effect, women are more likely to stay out of the labor force, in particular if they have children, as the opportunity cost of working (that is the loss of social benefits and additional costs associated with institutionalized childcare) might be too high. We therefore include factors affecting the reservation wage, as well as childcare obligations, in the selection equation.

To correct for sample selection we apply a three-step procedure, following Buchinsky (1998) and Buchinsky (2002)⁵. In the first step, we estimate the selection equation using

⁵More recently, Arellano and Bonhomme (2017) have proposed a non-additive approach to sample-selection

semiparametric maximum likelihood estimator by Klein and Spady (1993), which is asymptotically efficient in the sense that it attains the semiparametric efficiency bound, and record the generated single index. In the second step we regress the outcome variable on the characteristics and a polynomial of the generated inverse Mill's ratios. This, as shown by Buchinsky (1998) yields a consistent estimator of β , which under some regularity conditions is also asymptotically normal. Finally, as we have obtained consistent estimates of the coefficients, we can decompose the quantile functions in a manner described above. Reported specification tests are based on Chernozhukov et al. (2013).

The quantile selection estimation of Buchinsky (1998) can be shortly summarized as follows. Wage depends linearly on a set of labor market characteristics

$$y_i^* = \beta_0 \mathbf{X}_{i2} + u_i, \quad (3)$$

whereas y is observed only if it exceeds the reservation wage y^R given by

$$y_i^R = \alpha_0 \mathbf{X}_{i1} + v_i, \quad (4)$$

where (dropping the i index) $\mathbf{X}_2 \subset \mathbf{X}_1$.⁶ The observed wage can be written as

$$y = d \cdot y^* = d(\beta_\theta \mathbf{X}_2 + u_\theta), \quad (5)$$

where $d = I(y^* > y^R)$. The conditional quantile of the observed wage is given by

$$Quant_\theta(y|\mathbf{X}_2) = \beta_\theta \mathbf{X}_2 + h_\theta(\mathbf{X}_1, \gamma_0). \quad (6)$$

In the first, step we need to estimate γ_0 , independent of any assumptions about the distribution of v , which we conduct with semiparametric maximum likelihood by Klein and

correction. Moreover, Huber and Melly (2015) have expressed doubt about the assumption of homogenous selection across the quantiles, also visualized by Machado (2017). Since the main focus of this work is on the role of skills, we abstract from these considerations here.

⁶It is also necessary that at least one of the variables not included in \mathbf{X}_2 is continuous.

Spady (1993). In the second step, a consistent estimator for β_θ is obtained from a quantile regression of y on \mathbf{X}_2 and P_S , where P_S is a polynomial approximation of order S of e.g., the inverse of the Mill's ratio.

For identification, it is necessary to include in Equation 4 variables that determine the reservation wage, but do not enter the wage equation. In our case, the first-stage excluded instruments are: age, age squared, dummy for whether a person is living with a spouse/partner, the number of people living in the household and the job status of the partner (full-time work, part-time work, unemployed etc.). Included instruments entering both equations are all other demographic characteristics. We estimate the selection equations separately for males and females. Selection equations results (probit and single index) are reported in Table 8 in the Appendix.

3.2. Data

We use the dataset for Austria provided by the Programme for the International Assessment of Adult Competencies (PIAAC) survey conducted by the OECD in 2011/12. It encompasses 4,810 individual observations, including detailed information about education, skills, income and family background. After filtering out observations with missing data, our sample reduces to about 2,200 observations. In terms of the number of observations, it is smaller than the EU-SILC dataset, used for instance by Grandner and Gstach (2015), or the Austrian social security data used by Böheim et al. (2013a) but reasonably large to conduct an empirical analysis. However, we need to keep in mind that estimation at the tails will be less precise due to a smaller sample size, and caution is needed while forming conclusions, particularly at the upper tail of the distribution.

Dependent variable in all estimations are log-hourly wages⁷ with and without consideration of bonus payments - we are interested in one potential transmission channel for unexplained differences, that is, discretionary payments such as bonuses.

Similarly to previous studies, detailed information about personal and job characteristics is

⁷To see whether the wage distribution of the PIAAC dataset is a representative for the Austrian wage distribution, we compare the wage data of PIAAC with the wage tax statistics of Statistics Austria (see Figure 9 in the Appendix)

provided. The greatest advantage of the new dataset, however, is the possibility of controlling for various skills required and used at work, which typically are not being controlled for using other datasets. Unlike, for instance, Garcia et al. (2001) who uses instrumental variables to control for unobserved skills, we can directly control for skills at workplace, which on the one hand affect productivity, and on the other might be particularly rewarded by the employer.

We control not only for the skill use at work, but also for proficiency of the individual as well as skill match. To control for skill use at work, we follow Perry et al. (2014) and use the job requirements approach. The PIAAC questionnaire asks participants how often they use particular skills at work. Given participants' answers, several indices⁸ are calculated. Those indices are then summarized into quintiles, where the 1 stands for the 20 percent with lowest skill use and 5 for the highest values of skill use at work for a specific skill type. The indices are derived for several skills that are typically used at work: influencing skills, planning skills, reading skills, writing skills, numerical skills and ICT skills.

By simply comparing the average indices for males and females, we can see that there are on averages substantial differences in the skill use between genders. Figure 1 reveals that especially in numeric skills, but also in most of the other skill use variables, males tend to use those skills on average more often than females.

Further we can assess matching of skills and job requirements. We use a self-reported variable of skills' utilization specific to the Austrian edition of the PIAAC survey, which asks participants to assess utilization of their skills and knowledge, and the answer can vary between "not at all" and "to a very high extent".

⁸The mean score and standard errors are standardized with a mean equal to 2 and a standard error equal to 1 across the OECD countries participating in PIAAC.

Figure 1: Skill use at work: gender differences

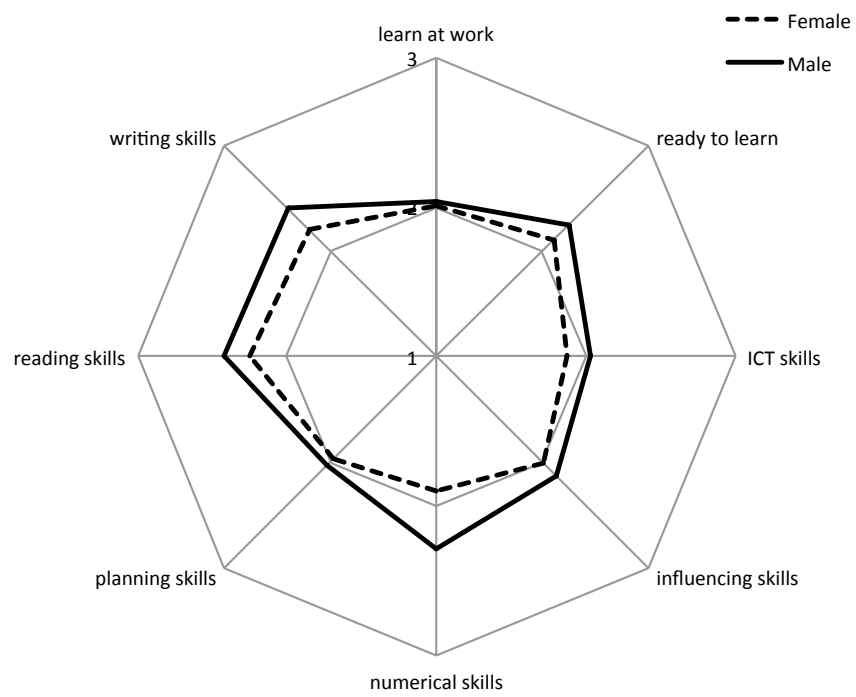


Table 2: Overview of the independent variables

Type	Variable	Categories
Personal characteristics	Education level	ISCED code
	Relevant job experience	years
	Children	number of
	Age of children	categories
	Out of workforce due to childcare	dummy
	Citizenship	country dummy
Job characteristics	Migration background	migrant (1st or 2nd gen.), born in AUT, parents born in AUT
	Employment status	dummy for blue-collar, white-collar, public servant, etc.
	Firm size	1 to 10, 11 to 50, 51 to 250, 251 to 1000 and ≥1000 employees
	Participation in on-the-job training	dummy
	Hours worked per week	hours
	Economic sector	NACE (2-digit) sector
	Type of activity	ISCO (2-digit) activity classification
	Work flexibility	none, little, moderate, high, very high
	Learning at work	measured in PIAAC
	Readiness to learn	measured in PIAAC
Skill use	ICT skills	measured in PIAAC
	Influencing skills	measured in PIAAC
	Numeracy skills	measured in PIAAC
	Planning skills	measured in PIAAC
	Writing skills	measured in PIAAC
	Reading skills	measured in PIAAC
	assessed Scores	measured in PIAAC
	skill match	skill use between "not at all" and "to a very high extent"

A second novelty in the Austrian context is use of information on familiar background. We control for several variables regarding childcare: the number of children and their age, as well as work time pauses due to childcare. The topic of childcare obligations is of high relevance in the Austrian context, as women tend to be more involved in childcare duties and often work part-time or even leave the labor force completely.

Third, we try to take a closer look on job flexibility, which is not often used as an explanatory variable due to the lack of data, but in the PIAAC dataset we are provided information about the flexibility of working hours, which is a good overall measure for the job flexibility in our opinion. Participants are asked, to what extent they can choose or change working hours. The answer can again vary between "not at all" and "to a very high extent".

As a benchmark case, we first estimate the gender gap using a similar set of control variables as Böheim et al. (2013a), without additional control for individual skills, to make sure that the findings derived from the new dataset can be compared to previous studies. Subsequently, we show how the estimation results change, when we consider skills' use. Finally, in the sample-selection model we include personal characteristics affecting the decision to work, as a result of the reservation wage, and at the same time potentially affecting the wages, such as childcare, education level, age, migration status etc. For the sample selection model, we use the whole sample, that is also the observations for which we do not observe the wages since the person is not working; the overall sample size in this case equals 4,700 observations.

Log-wage distributions for genders are presented in Figure 8 in the Appendix. We can see that for the hourly wages in total, the wage distribution of males is shifted to the right, no matter whether we use the whole sample or only those that work full-time. But we can see, that the differences in the wage distributions between females and males is lower when we restrict the sample to full-time workers.

4. Results

In this section, we first give an overview of the quantile regression results for men and women, and determine whether different characteristics have more or less influence on wages of both genders. Detailed results can be found in the Appendix, whereas for coefficients of

interest visualization is provided in Figures 2, 3 and 4. In a second step, we take a closer look on how much of the unexplained part of the gender wage gap is influenced by skills, selection bias and the bonus payments. In a third step, we take a closer look on whether the gender wage gap is constant along the income distribution or whether there are differences depending on the income.

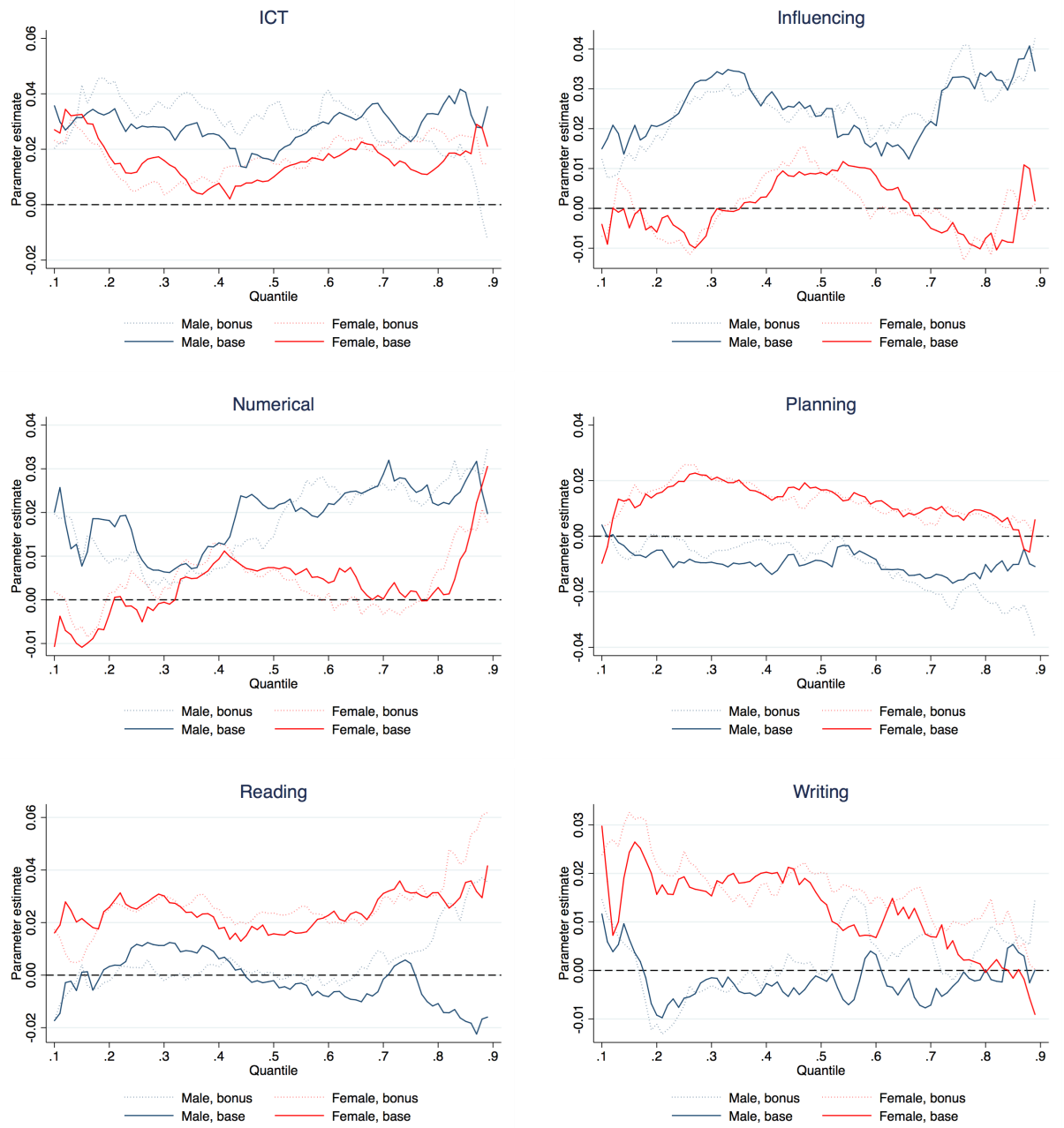
4.1. Wage regression

In this section we want to focus on drivers that influence male and female wages. Quantile regression results for men and women are shown in Tables 9, 10, 11 and 12 in the Appendix. The typical variables such as experience and education show, as expected, a significant positive effect on wages. There seem to be no large differences in the coefficients between both genders for the typical explanatory variables such as experience and education. Additionally, the wage is lower, the lower the firm size, which confirms some previous findings. We can see that especially small firms (one to 10 employees) pay significantly less than the largest companies. This is also driven by the fact that self-employed (one-person entities) usually experience lower wages.

The largest differences between men and women emerge, when we take a closer look on the skills (Figure 2). Our results indicate a wage premium for planing skills, reading skills and writing skills for women. Reading skill premium equals about 2 percent along the wage distribution. For writing and planing skills, the premium equals about 2 percent for the middle income groups, declining towards zero along with increasing income. Women that use these skills at work are better paid than those who use other skills. Numerical skills, for example, only show positive effects at the upper end of the wage distribution for women. This is in strong contrast to male-based skill premiums where especially numerical skills (2 percent) and ICT skills (between 2 and 4 percent) have a positive influence on the wage.

In general, Figure 2 shows substantial differences in returns to specific skills between men and women. While use of writing, reading and planing skills at work shows positive returns for women, there are no or even negative returns for men. On the other hand, numerical skills and influencing skills show high returns for males, while for females returns are about zero.

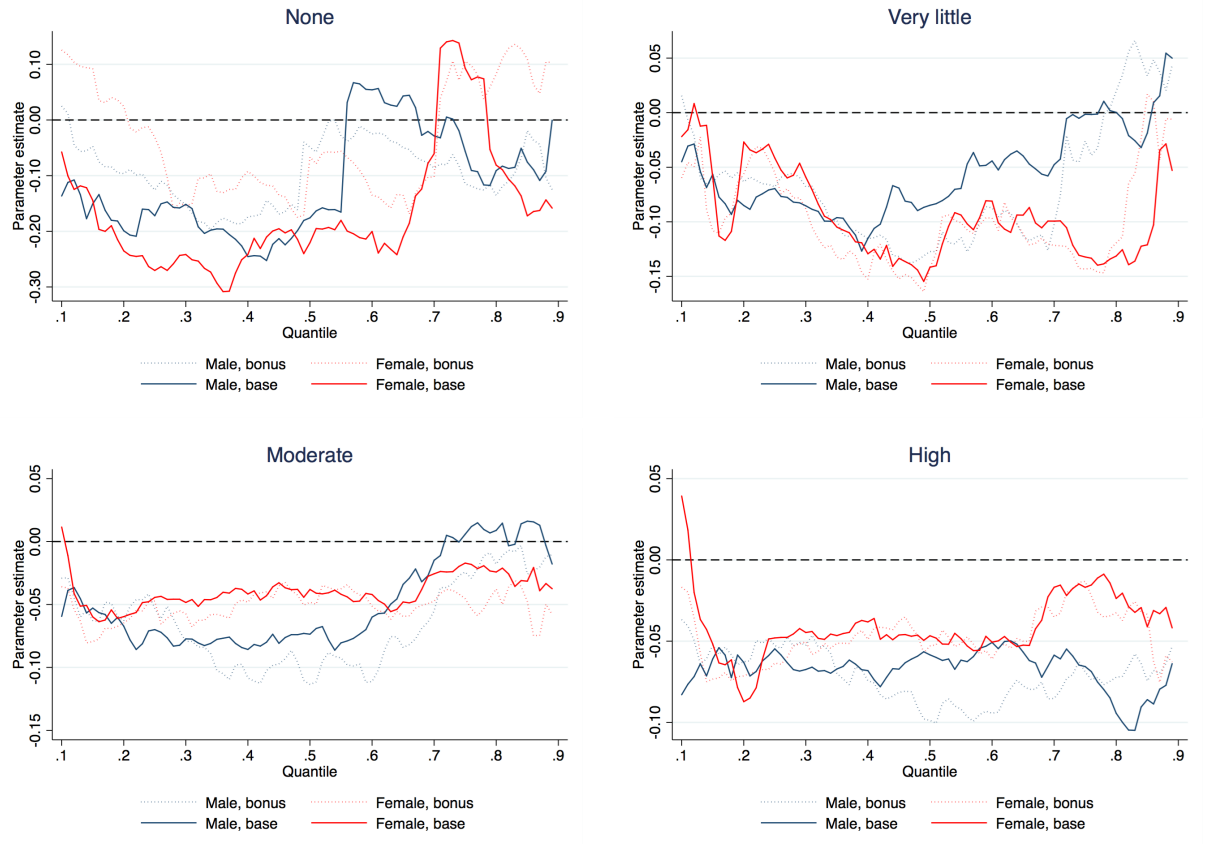
Figure 2: Wage regression coefficients for skills



Positive returns for both males and females can be found with respect to ICT skills; still, the premium for males (2 to 4 percent) is larger than those for females (0 to 2 percent).

Skill utilization also has a positive influence on wages of female workers (Figure 3). Using the skills to a very high extent leads to a wage premium. In the middle of the income distribution, a low skill match leads to 22 percent lower wage compared to a high skill match.

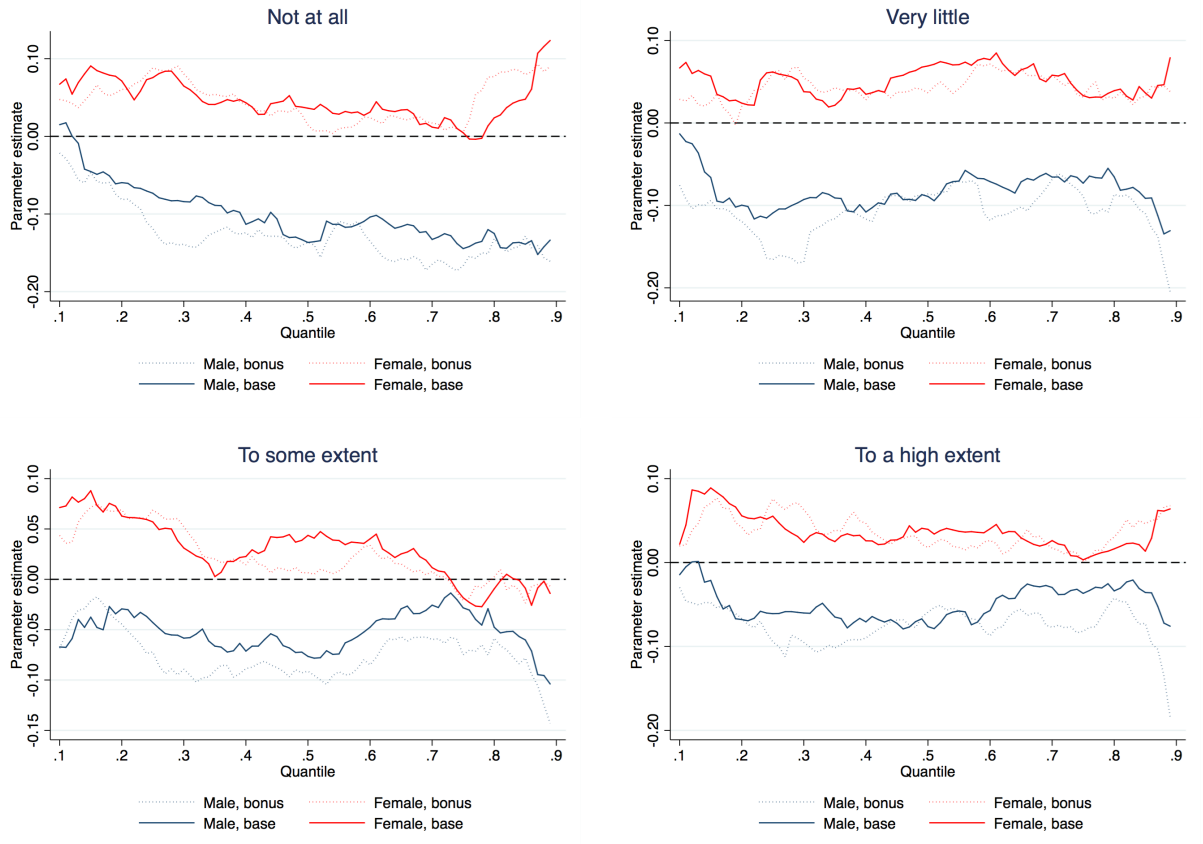
Figure 3: Wage regression coefficients for skills' utilization



In general, there are no large differences in returns to skill match between male and female workers.

Work flexibility shows opposite signs for women and men (Figure 4). While for males, flexible jobs are rewarded (no flexibility leads to a wage decrease of about 10 percent, especially for middle and high incomes compared to high flexibility), for women jobs that are less flexible pay more than those that offer flexible working hours (no flexibility increases the wage by almost 10 percent compared to high flexibility for low and middle incomes part of the income distribution). This fact, jointly with the skill premiums, suggest that, job flexibility for females has a different form than for males. Women might typically work in jobs requiring office-based skills such as reading and planning, and are not associated with flexible working hours. It seems that, these kinds of occupations are generally less rewarded compared to ICT and managerial positions, even considering for similar educational backgrounds. This result

Figure 4: Wage regression coefficients for job flexibility



is in line with Goldin (2014).

Children seem to have a negative effect on wages for women, but only in the middle of the income distribution, this negative impact is significant. This is likely associated with age, as younger females experience lower wages resulting from career breaks due to Austrian seniority wages, whereas at the very top of the distribution, children will typically be already grown up. On the other hand, there seems to be a child bonus for men, especially for low and middle income earners, although the coefficients are not significant.

4.2. Selection equation

In Table 8 in the Appendix, we report the results of the selection equations, where the dependent variable is a dummy for working. Panel (a) presents the results of a probit regression and panel (b) refers to the single-index estimation, which is subsequently used to calculate the Mill's ratios for the wage equations. We can observe, that age is a significant

predictor of being employed, probability of which increases along with age for both genders. The squared term is insignificantly different from zero in the single-index model. Employment status of a partner is relevant for both genders, with significantly negative impact on working if the partner is retired, despite controlling for age. A significant predictor of not working is disability of a partner, although only for males. Similarly, unemployment of a partner is a significant predictor of lower probability of working for males, which suggests clustering of households, with both partners being either unemployed or staying out of the labor force.

As for the variables which enter both the selection and the wage equation, consistently with our predictions and previous studies (Picchio and Mussida 2011), women are working less often if they have children. Unlike the case of Italy analyzed by (Picchio and Mussida 2011), however, lower labor market participation of women persists for a longer time-span, that is also for the cases of children between 6 and 12. This is to a large extent explained by a generally conservative view of the society on the role of females and little cultural acceptance for working mothers⁹. Additionally, women of Turkish origin work less than other nationalities. Finally, controlling for nationalities, first generation migrants work more than Austrian-born persons, which is understandable in the context of labor migrations within the European Union, as well as the laws governing permission to settle in a country for foreigners, which are generally dependent on employment.

4.3. The Gender Wage Gap

In the first part, we compare the results of our study with previous specifications, to show whether using the same set of controls, the wage differences have been changing since Böheim et al. (2013a), who use data from the year 2007. We report decomposition results from the male-based estimations¹⁰.

⁹Compare with European Social Survey (2009) and International Social Survey Programme (2012) data, which generally show a negative attitude towards working mothers in Austria, compared to other European countries.

¹⁰We do not report the female-based results for space-saving purposes. Full results can be obtained upon request, and are broadly consistent with findings of Böheim et al. (2013a) and other literature, i.e. show higher unexplained gender gaps compared to male-based results.

Table 3: Benchmark model (in %) - effects of coefficients - hourly wages without bonus payments

Male-based			
Quantile	Effect	Functional 95% LB	Functional 95% UB
1	6.33	-4.70	17.38
2	8.60	0.71	16.49
3	10.08	3.16	17.00
4	10.97	4.25	17.69
5	11.01	4.35	17.67
6	11.12	4.39	17.85
7	11.33	4.62	18.03
8	12.78	5.91	19.66
9	15.00	5.75	24.25
Test		KS p-val	CMR p-val
Correct Model 0		0.32	0.15
Correct Model 1		0.30	0.06
Differences			
$QE(\tau) = QE(0.5)\forall\tau$		0.04	0.18
Characteristics			
$QE(\tau) = QE(0.5)\forall\tau$		0.65	0.47
Coefficients			
$QE(\tau) = QE(0.5)\forall\tau$		0.57	0.48

Böheim et al. (2013a) shows that the unexplained gender wage gap is on average 13 percent. Using a similar model as a benchmark, our data suggests an unexplained gender wage gap between 6 and 15 percent along the wage distribution. In the middle of the wage distribution, our data reveals a unexplained gap of approximately 11 percent, indicating that our data would suggest a similar but slightly lower gender wage gap¹¹.

The low level of the unexplained gender wage gap at the lower tail of the distribution is consistent with the fact that low income earners in Austria are usually covered by collective bargaining rules. These laws are typically gender-neutral and do not allow for discrimination. Therefore, the result is not surprising, however, as mentioned above, it might also indicate that selection to working is in place, a question which we will approach later on. The unexplained part of the gender wage gap increases along the wage distribution, when we move from collective to individual bargaining.

¹¹Whether the (small) differences between the unexplained parts of the gender wage gap of Böheim et al. (2013a) and our calculations result from different times of data collection (2007 vs. 2011/2012), or from general differences in the data can not be answered with certainty.

Table 4: Skills model (in %) - effects of coefficients - hourly wages without bonus payments

Male-based			
Quantile	Effect	Functional 95% LB	Functional 95% UB
1	2.61	-9.43	14.66
2	4.34	-3.88	12.57
3	5.74	-1.81	13.31
4	6.93	-0.44	14.30
5	7.29	0.19	14.39
6	7.96	1.07	14.85
7	8.47	1.71	15.23
8	9.19	2.26	16.13
9	10.81	2.18	19.44
Test		KS p-val	CMR p-val
Correct Model 0		0.53	0.57
Correct Model 1		0.46	0.31
Differences			
$QE(\tau) = QE(0.5)\forall\tau$		0.02	0.00
Characteristics			
$QE(\tau) = QE(0.5)\forall\tau$		0.20	0.19
Coefficients			
$QE(\tau) = QE(0.5)\forall\tau$		0.63	0.40

4.3.1. Results including skills and skills-matching

In the second step, we add skills, skills utilization and work flexibility to our estimations. Figure 5 shows, that especially in the middle and upper parts of the distribution we can explain a larger part of the wage gap simply by adding these characteristics to our model.

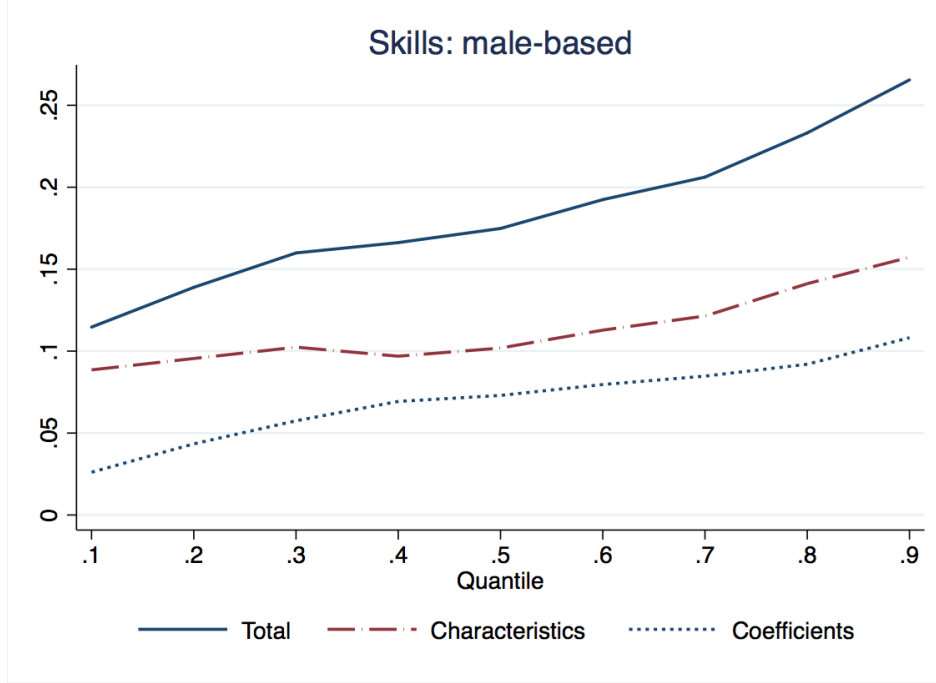
We can see, that especially in the middle part of the distribution, the unexplained part of the gender wage gap lies much lower than in Table 3, indicating that the skill and skill-matching characteristics differ to a great extent between men and women. Compared to 11% at the median for the benchmark case, the unexplained part reduces to 7.29% for the skills' model. That already shows the vast influence of skills and skill match wage differences between men and women, a result which we have already observed while looking at the wage regressions.

Analyzing the unexplained part of the gender wage gap (blue dashed line in Figure 5), we see, that especially in the lower part of the distribution, the unexplained part becomes close to zero, but it still increases along with increasing incomes.

4.3.2. Sample-selection correction

Figure 6 shows the results of the decomposition along the wage distribution when we correct for sample selection. Especially in the lower part of the distribution, we can see a

Figure 5: Wage differences (without bonus payments) considering skills



substantial decrease in the explained part of the the gender wage gap (red line) compared to the results without the sample selection correction (see Figure 5). As expected, sample selection is of particular relevance for lower incomes.

The sample selection bias is especially high at the lower end of the distribution, where the unexplained GWG increases from 2.63 percent to 4.04 percent (see table 5). In the middle of the income distribution, the selection bias becomes smaller and tends towards zero while moving further up in the wage distribution. This is in line with the finding, that especially low productive female tend to stay out of the workforce.

4.3.3. Bonus payments

In the next step, we will take a closer look on the bonus payments in our preferred specification, which is the one where we add skills and control for sample selection. Figure 7 shows the decomposition results along the wage distribution.

We can see that the wage gap (blue line) gets larger, especially in the upper part of the distribution. Since bonuses are typically paid to workers in higher positions, who are also

Table 5: Skills and sample selection model (in %) - effects of coefficients - hourly wages without bonus payments

Male-based			
Quantile	Effect	Functional 95% LB	Functional 95% UB
1	4.04	-9.60	17.68
2	5.42	-4.99	15.84
3	6.77	-2.15	15.69
4	7.01	-1.42	15.46
5	7.10	-1.39	15.60
6	7.76	-0.89	16.43
7	8.53	-0.31	17.37
8	8.75	-0.06	18.17
9	10.75	0.06	22.12
Test		KS p-val	CMR p-val
Correct Model 0		0.38	0.41
Correct Model 1		0.48	0.31
Differences			
$QE(\tau) = QE(0.5)\forall\tau$		0.05	0.05
Characteristics			
$QE(\tau) = QE(0.5)\forall\tau$		0.30	0.36
Coefficients			
$QE(\tau) = QE(0.5)\forall\tau$		0.77	0.71

Figure 6: Wage differences (without bonus payments) with sample selection

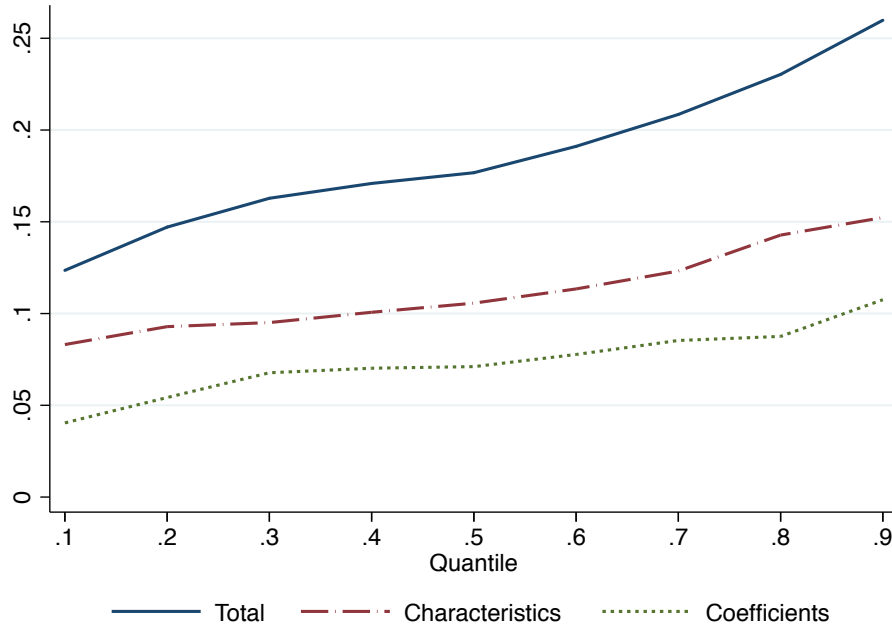


Table 6: Skills and sample selection model (in %) - effects of coefficients - hourly wages with bonus payments

Male-based			
Quantile	Effect	Functional 95% LB	Functional 95% UB
1	2.54	-10.61	15.70
2	4.92	-4.43	14.29
3	6.62	-1.89	15.15
4	6.71	-1.69	15.10
5	7.06	-1.22	15.35
6	8.42	0.78	16.05
7	9.36	1.62	17.10
8	10.26	2.74	17.77
9	13.43	3.42	23.43
Test		KS p-val	CMR p-val
Correct Model 0		0.29	0.23
Correct Model 1		0.13	0.26
Differences			
$QE(\tau) = QE(0.5)\forall\tau$		0.01	0.01
Characteristics			
$QE(\tau) = QE(0.5)\forall\tau$		0.18	0.19
Coefficients			
$QE(\tau) = QE(0.5)\forall\tau$		0.51	0.36

better paid in general, this seems to be straightforward. However, we can also see that the unexplained part of the wage gap increases in the upper part of the distribution (compared to Figure 6). Whether this is due to discrimination, individual bargaining process for bonuses, or, for instance, due to risk-aversion or cultural reasons cannot be answered with certainty.

4.4. The unexplained part of the gender gap - comparison

In this section, we compare our results with those of Böheim et al. (2013b). As a benchmark for our decomposition, we use a similar specification as Böheim et al. (2013b). We compare these results with a specification in which we add job-specific skill variables, as well as the skill-matching variable.

Böheim et al. (2013b) found a (male-based) unexplained gender wage gap of 15.2 percent in 2002, which decreased to 13.2 percent in 2007. Our results, which are summarized in Table 7, show that the unexplained gender wage gap in 2012 is close to 11 percent in the middle of the wage distribution, when we use a similar specification. The slightly lower value does not necessarily mean the unexplained gender wage gap has narrowed during this period. The slight difference can also stem from different methodology, as well as different data that is used. Still, the results are quite similar.

The unexplained part of the gender wage gap is increasing along the wage distribution,

Figure 7: Wage differences with bonus payments and sample selection

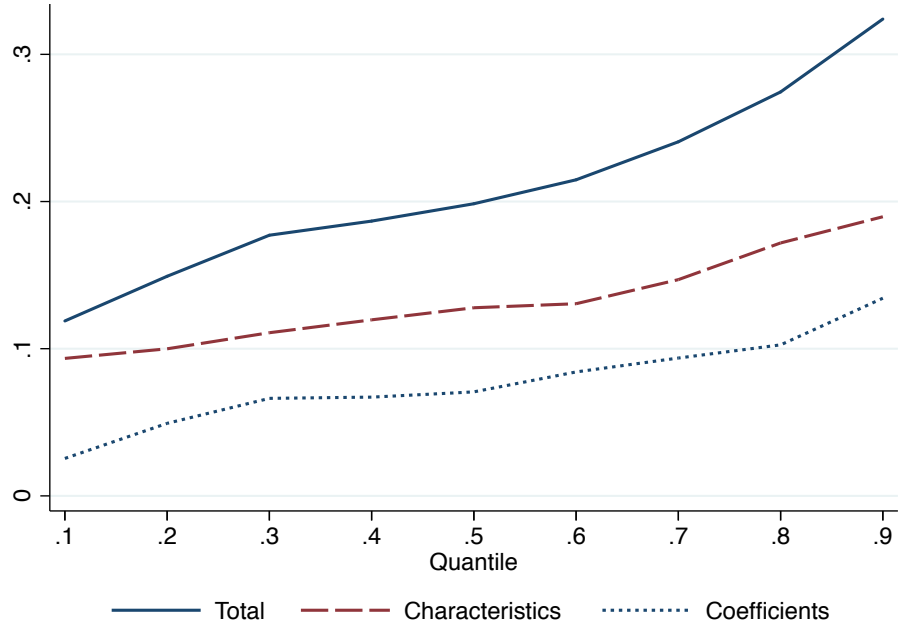


Table 7: Overview of the unexplained differences (in %) - hourly wages without bonus payments

Quantile	Male-based		
	Benchmark	Skills	Skills and sample selection
1	6.33	2.61	4.04
2	8.60	4.34	5.42
3	10.08	5.74	6.77
4	10.97	6.93	7.01
5	11.01	7.29	7.10
6	11.12	7.96	7.76
7	11.33	8.47	8.53
8	12.78	9.19	8.75
9	15.00	10.81	10.75

no matter which specification we use. Therefore, our findings are in contrast to Grandner and Gstach (2015), who argue that the unexplained part of the gender wage gap in Austria is stable over the wage distribution, with a slight exception on the lower tail of the distribution. On the other hand, our results are in line with Böheim et al. (2013a) and Grünberger et al. (2009).

When we add the skill variables to our specification, the unexplained part of the gender wage gap decreases significantly. The new specification reduces the unexplained part by almost 4 percentage points along the distribution. In the lower part of the distribution, the unexplained part is cut to half (compared to the benchmark specification) by adding skills. In the upper part of the wage distribution, more than 25 percent of the unexplained part of the gender wage gap (in the benchmark scenario) can be explained by skills. The unexplained part of the gender wage gap in the middle of the wage distribution is lowered from 11 to 7.3 percent by adding skills in the specification.

As the results in Columns (1) and (2) of Table 7 show, skills play a crucial role in explaining a large part of the gender wage gap. Still, we need to account for self-selection, since it is reasonable to assume that, many less productive women drop out of the labor force, leading to biased results in our model.

In the third specification we account for the self-selection problem. It turns out, that the self-selection biases our results especially in the lower part of the wage distribution, while in the middle and upper parts of the wage distribution, the effect disappears. Still, one can see that the self selection problem is not only a problem at the very lower tale of the wage distribution. Nevertheless, even considering for selection, the model with skills explains about 3 to 4 percentage points more of the gender gap than the benchmark specification.

5. Summary and conclusions

Our paper analyzed the gender wage gap for Austria along the wage distribution. We use a quantile regression approach and combine with a decomposition method that follows Chernozhukov et al. (2013). Additionally, we account for the self-selection problem following Buchinsky (1998) and Buchinsky (2002).

Our results are in line with Böheim et al. (2013b), who show that the unexplained wage gap in Austria is approximately 13 percent in 2007. When we use our data with a similar specification, our results remain fairly close to this result in 2011. However, as a novelty of this work, we use data of the Programme for the International Assessment of Adult Competencies (PIAAC) survey conducted by the OECD in 2011/12. This allows us to control for skill-use at work, as well as for skill-match, which are often argued to affect the wages even within the same occupations.

Adding these variables to the wage regressions for males and females already shows some interesting insights. When we take a closer look on female skill use, we can see that there is a wage premium for planing skills, reading skills and writing skills. Numerical skills only show positive effects at the upper end of the wage distribution. This is in strong contrast to the male-based skill premium where numerical skills and ICT skills have a positive influence on the wage. Skill match has positive influence on wages for both genders.

When we add the skill variables to our model, the unexplained part of the gender wage gap decreases by almost 4 percentage points all along the wage distribution. Our results show that in the lower part of the wage distribution more than 70% of the gender wage gap can be explained by different characteristics and is, therefore, due to "fair" discrimination. The unexplained part is less than 3 percent in the lower part of the distribution. When we move up along the wage distribution, the percentage of the explained gender wage gap is decreasing to less than 60 percent in the top quantile of the wage distribution. The unexplained gap at the upper end is about 11%.

The low unexplained part of the gender wage gap in the lower part of the income distribution is mostly due to the fact that most employees are covered by collective minimum wages and collective bargaining, which do not allow for gender discrimination. When we move further up in the wage distribution, the collective bargaining moves to individual bargaining. So it seems that individual bargaining has a substantial impact on the differences in wages between men and women. As we show, this also seems to hold true when it comes to bonus payments. Whether this is due to differences in individual bargaining behaviour or purely to discrimination cannot be answered within our methodology.

Controlling for sample selection is an important question, when we want to state the size of the unexplained gender wage gap, particularly at the lower end of the wage distribution. Our results show that, as expected, the self-selection problem is especially important for low-income earners. For middle-wage earners, the effects are small and not significant. At the upper end of the wage distribution the effect seems to disappear. In general the effect of self selection is small in size - approximately 1.5 percentage points for low income earners, but not deniable.

Many works analyzing the gender wage gap conclude that the unexplained part is decreasing over time. This is good news, but we would argue, that this is only half of the good news. We found that additional information on the job and workers' characteristics (which are often difficult to measure) such as skill-use or skill-match, reduce the unexplained gender wage gap even further, and suggests that part of the decrease over time can be attributed to better data access and more sophisticated econometric techniques. This does not mean, that there is no discrimination between men and women on the labor market, as is argued by many papers that question the possibility of capturing all individual characteristics. But it means that we might be closer to the absence of labor-market discrimination than we thought before, especially with regard to low-wage earners.

References

- Angelov, N., P. Johansson, and E. Lindahl (2013). Is the persistent gender gap in income and wages due to unequal family responsibilities?
- Arellano, M. and S. Bonhomme (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica* 85(1), 1–28.
- Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics* 31(S1), S59–S96.
- Black, S. E. and A. Spitz-Oener (2010). Explaining women’s success: technological change and the skill content of women’s work. *The Review of Economics and Statistics* 92(1), 187–194.
- Blau, F. D., P. Brummund, and A. Y.-H. Liu (2013). Trends in occupational segregation by gender 1970–2009: Adjusting for the impact of changes in the occupational coding system. *Demography* 50(2), 471–492.
- Blau, F. D. and L. M. Kahn (2000). Gender differences in pay. Technical report, National bureau of economic research.
- Blau, F. D. and L. M. Kahn (2016). The gender wage gap: Extent, trends, and explanations. Technical report, National Bureau of Economic Research.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.
- Böheim, R., K. Himpele, H. Mahringer, and C. Zulehner (2013a). The distribution of the gender wage gap in Austria: evidence from matched employer-employee data and tax records. *Journal for Labour Market Research* 46(1), 19–34.
- Böheim, R., K. Himpele, H. Mahringer, and C. Zulehner (2013b). The gender wage gap in Austria: eppur si muove! *Empirica* 40(4), 585–606.

- Böheim, R., H. Hofer, and C. Zulehner (2005). Wage differences between men and women in Austria: Evidence from 1983 and 1997.
- Böheim, R., H. Hofer, and C. Zulehner (2007). Wage differences between Austrian men and women: semper idem? *Empirica* 34(3), 213–229.
- Boll, C., J. Leppin, A. Rossen, and A. Wolf (2016). Magnitude and impact factors of the gender pay gap in eu countries. *Report prepared for and financed by the European Commission–Directorate-General for Justice, European Union (ed.)*.
- Boll, C. and J. S. Leppin (2015). Die geschlechtsspezifische lohnücke in deutschland: Umfang, ursachen und interpretation. *Wirtschaftsdienst* 95(4), 249–254.
- Buchinsky, M. (1998). Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of human resources*, 88–126.
- Buchinsky, M. (2002). Quantile regression with sample selection: Estimating women’s return to education in the US. In *Economic Applications of Quantile Regression*, pp. 87–113. Springer.
- Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica* 81(6), 2205–2268.
- Depalo, D., R. Giordano, and E. Papapetrou (2015). Public–private wage differentials in euro-area countries: evidence from quantile decomposition analysis. *Empirical Economics* 49(3), 985–1015.
- Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of labor economics* 4, 1–102.
- Garcia, J., P. J. Hernández, and A. Lopez-Nicolas (2001). How wide is the gap? an investigation of gender wage differences using quantile regression. *Empirical economics* 26(1), 149–167.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *The American Economic Review* 104(4), 1091–1119.

- Grandner, T. and D. Gstach (2015). Decomposing wage discrimination in Germany and Austria with counterfactual densities. *Empirica* 42(1), 49–76.
- Grünberger, K., C. Zulehner, et al. (2009). Gender-specific wage gaps in Austria. *WIFO Monatsberichte (monthly reports)* 82(2), 139–150.
- Huber, M. and B. Melly (2015). A test of the conditional independence assumption in sample selection models. *Journal of Applied Econometrics* 30(7), 1144–1168. jae.2431.
- Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 387–421.
- Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica*, 33–50.
- Kunze, A. (2008). Gender wage gap studies: consistency and decomposition. *Empirical Economics* 35(1), 63–76.
- Machado, C. (2017). Unobserved selection heterogeneity and the gender wage gap. *Journal of Applied Econometrics*, n/a–n/a. jae.2561.
- Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics* 20(4), 445–465.
- Melly, B. (2006). Estimation of counterfactual distributions using quantile regression.
- Meurs, D., A. Pailhé, and S. Ponthieux (2010). Child-related career interruptions and the gender wage gap in france. *Annals of Economics and Statistics/Annales d’Économie et de Statistique*, 15–46.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693–709.
- Perry, A., S. Wiederhold, and D. Ackermann-Piek (2014). How can skill mismatch be measured? new approaches with PIAAC. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology* 8(2), 137–174.

- Picchio, M. and C. Mussida (2011). Gender wage gap: A semi-parametric approach with sample selection correction. *Labour Economics* 18(5), 564–578.
- Tijdens, K., M. Klaveren, et al. (2012). Frozen in time: Gender pay gap unchanged for 10 years.
- Weeden, K. A., Y. Cha, and M. Bucca (2016). Long work hours, part-time work, and trends in the gender gap in pay, the motherhood wage penalty, and the fatherhood wage premium. *RSF*.
- Weichselbaumer, D. and R. Winter-Ebmer (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys* 19(3), 479–511.

Appendix

Figure 8: Log hourly wages without bonus payments: all workers (left panel), full-time workers (right panel)

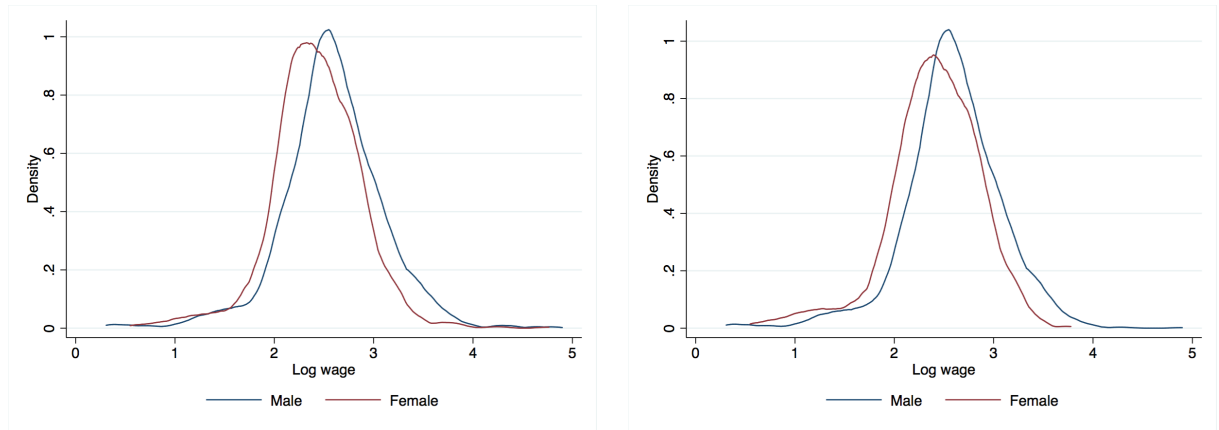
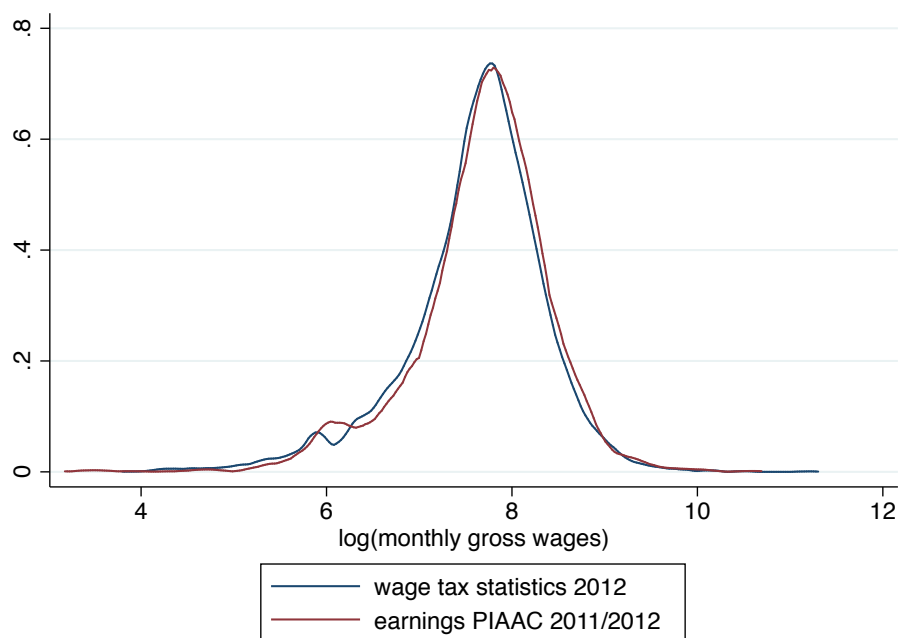


Figure 9: Wage dsitribution - PIAAC vs. wage statistics



Comment: To make the both datasets comparable, we adjust the wage statistics 2012 for age-groups below 65 and include only taxpayers that were employed during the whole year, since PIAAC data is based on hourly wages, whereas the tax statistics is on a yearly basis. For the PIAAC data 2011/2012, we leave out self-employed, who are not included in the tax statistics.

Table 8: Selection equation, dependent variable: Working or not working; probit and single index

(a) Probit			(b) Single Index		
	Male	Female		Male	Female
Children	Normalized to 1		Children	Normalized to 1	
ChildCare	0.00	-1.91***	ChildCare	0.04	-0.51***
(.)	(.)	(-6.15)	(0.25)	(0.25)	(5.34)
2 or younger	-0.45*	-1.53***	2 or younger	0.07	-0.10
(-1.65)	(-1.65)	(-8.94)	(0.98)	(0.98)	(1.46)
aged 3 to 5	0.00	-1.47***	aged 3 to 5	0.15	-0.18**
(.)	(.)	(-7.93)	(1.47)	(1.47)	(2.47)
aged 6 to 12	-0.46*	-1.70***	aged 6 to 12	0.00	-0.16**
(-1.88)	(-1.88)	(-9.84)	(0.05)	(0.05)	(2.34)
ISCED 1	-0.35**	-0.36**	ISCED 1	-0.11*	-0.05
(-1.96)	(-1.96)	(-2.21)	(1.67)	(1.67)	(0.70)
ISCED 2	0.21	-0.19	ISCED 2	0.01	-0.04
(1.32)	(1.32)	(-1.22)	(0.19)	(0.19)	(0.63)
ISCED 3	0.27	-0.09	ISCED 3	0.02	-0.07
(1.31)	(1.31)	(-0.56)	(0.34)	(0.34)	(1.25)
ISCED 4	-0.29*	-0.25	ISCED 4	-0.12**	-0.02
(-1.65)	(-1.65)	(-1.46)	(2.22)	(2.22)	(0.27)
ISCED 5+6	0.19	0.05	ISCED 5-6	0.06	0.01
(1.01)	(1.01)	(0.25)	(1.01)	(1.01)	(0.12)
Austrian	-2.18**	-4.46***	Austrian	0.06	0.08
(-2.48)	(-2.48)	(-5.40)	(0.85)	(0.85)	(1.09)
German	-2.07**	-5.11***	German	0.11	-0.01
(-2.15)	(-2.15)	(-5.75)	(1.16)	(1.16)	(0.09)
Serbian	-2.81***	-3.78***	Serbian	-0.03	0.19
(-2.93)	(-2.93)	(-3.72)	(0.29)	(0.29)	(1.50)
Turkish	-2.73***	-6.06***	Turkish	-0.03	-0.19*
(-2.88)	(-2.88)	(-6.63)	(0.35)	(0.35)	(1.95)
Bosnian	-2.85***	-5.12***	Bosnian	-0.05	-0.09
(-2.86)	(-2.86)	(-5.76)	(0.50)	(0.50)	(1.02)
Born in country	0.39	0.57	Born in country	0.12	0.11
(0.52)	(0.52)	(0.82)	(1.30)	(1.30)	(1.22)
Parents born in country	0.03	-0.08	Parents born in country	-0.09	-0.02
(0.16)	(0.16)	(-0.40)	(1.62)	(1.62)	(0.28)
1st generation migrant	0.28	0.48	1st generation migrant	0.10*	0.11**
(0.35)	(0.35)	(0.66)	(1.90)	(1.90)	(1.97)
Age	0.19***	0.30***	Age	0.06***	0.07***
(8.31)	(8.31)	(12.42)	(3.90)	(3.90)	(4.35)
Age sq.	-0.00***	-0.00***	Age sq.	0.00	0.00
(-9.74)	(-9.74)	(-13.56)	(0.75)	(0.75)	(0.82)
Living with partner	0.32	-0.19	Living with partner	0.09	-0.04
(0.67)	(0.67)	(-0.38)	(0.78)	(0.78)	(0.34)
Empl. status of partner			Empl. status of partner		
Full-time employed	0.22	0.44	Full-time employed	-0.02	0.03
(0.47)	(0.47)	(0.90)	(0.35)	(0.35)	(0.66)
Part-time employed	0.27	0.42	Part-time employed	0.00	0.14
(0.55)	(0.55)	(0.77)	(0.08)	(0.08)	(1.39)
Unemployed	-0.66	0.03	Unemployed	-0.20**	0.10
(-1.20)	(-1.20)	(0.05)	(2.42)	(2.42)	(1.24)
Pupil, student	-0.13	-0.57	Pupil, student	-0.07	-0.15
(-0.24)	(-0.24)	(-1.00)	(0.89)	(0.89)	(1.58)
Apprentice, internship	0.00	-0.04	Apprentice, internship	0.10	-0.06
(.)	(.)	(-0.04)	(0.41)	(0.41)	(0.24)
Retired	-0.62	-0.12	Retired	-0.19***	-0.11**
(-1.26)	(-1.26)	(-0.24)	(3.42)	(3.42)	(2.02)
Disabled	-1.47**	0.08	Disabled	-0.43***	0.07
(-2.25)	(-2.25)	(0.12)	(4.29)	(4.29)	(0.53)
Domestic tasks	-0.23	-0.22	Domestic tasks	-0.09*	-0.06
(-0.47)	(-0.47)	(-0.34)	(1.64)	(1.64)	(0.55)
Number living in household			Number living in household		
1			1		
2	-0.26	0.38**	2	-0.04	0.05
(-1.28)	(-1.28)	(2.29)	(1.05)	(1.05)	(1.12)
3	-0.28	0.33**	3	-0.05	0.04
(-1.42)	(-1.42)	(2.03)	(1.26)	(1.26)	(0.82)
4	-0.14	0.31*	4	-0.02	0.06
(-0.70)	(-0.70)	(1.95)	(0.35)	(0.35)	(1.21)
5 or more	-0.06	0.24	5 or more	0.05	0.08
(-0.27)	(-0.27)	(1.28)	(0.93)	(0.93)	(1.15)
Observations	1749	1944	Observations	1749	1944

Significance * 0.1, ** 0.05, *** 0.01
t-Statistics in brackets.

Significance * 0.1, ** 0.05, *** 0.01
t-Statistics in brackets.

Table 9: Quantile regressions with skills - hourly wages without bonus payments - results for males

	(1) 10%	(2) 25%	(3) 50%	(4) 75%	(5) 90%
Children	-0.02 (-0.47)	-0.03 (-0.95)	-0.03 (-0.90)	0.02 (0.58)	0.01 (0.21)
Age of children					
2 or younger	0.05 (0.90)	0.05 (1.62)	0.04 (0.93)	-0.02 (-0.56)	-0.07 (-1.07)
aged 3 to 5	-0.04 (-0.62)	0.02 (0.46)	0.08 (1.38)	0.02 (0.38)	0.10 (1.22)
aged 6 to 12	0.04 (0.76)	0.06** (2.40)	0.03 (0.91)	-0.04 (-1.16)	-0.05 (-0.80)
13 or older			base		
Skills					
Learning at work	0.00 (0.33)	-0.00 (-0.21)	-0.01 (-0.97)	-0.00 (-0.37)	-0.01 (-0.88)
Readiness to learn	0.01 (0.51)	-0.00 (-0.57)	-0.01 (-0.72)	-0.01 (-0.66)	0.01 (0.88)
ICT	0.04** (2.20)	0.03*** (3.24)	0.02 (1.61)	0.02** (2.45)	0.01 (0.97)
Influencing	0.01 (1.16)	0.03*** (4.19)	0.02** (2.37)	0.03*** (3.10)	0.03** (2.15)
Numerical	0.02 (1.51)	0.01* (1.86)	0.02** (2.29)	0.03*** (3.36)	0.02 (1.43)
Planning	0.00 (0.33)	-0.01* (-1.78)	-0.01 (-1.13)	-0.02** (-2.02)	-0.01 (-0.55)
Reading	-0.02 (-1.02)	0.01 (1.26)	-0.00 (-0.20)	0.00 (0.41)	-0.01 (-0.63)
Writing	0.01 (0.84)	-0.01 (-0.81)	-0.00 (-0.19)	-0.00 (-0.23)	0.01 (0.54)
Work flexibility					
Not at all	0.02 (0.25)	-0.07*** (-3.08)	-0.14*** (-4.43)	-0.14*** (-4.01)	-0.13** (-2.24)
Very little	-0.01 (-0.21)	-0.11*** (-3.18)	-0.09** (-2.11)	-0.07** (-2.15)	-0.15** (-2.43)
To some extent	-0.07 (-1.53)	-0.04** (-2.06)	-0.08*** (-2.69)	-0.03 (-1.05)	-0.13** (-2.29)
To a high extent	-0.01 (-0.31)	-0.06*** (-2.63)	-0.08** (-2.49)	-0.04 (-1.31)	-0.10 (-1.57)
To a very high extent			base		
Education					
ISCED 1			base		
ISCED 2	0.07 (1.41)	-0.02 (-0.69)	-0.02 (-0.72)	-0.01 (-0.51)	0.01 (0.26)
ISCED 3	0.07 (0.72)	0.04 (1.09)	-0.01 (-0.24)	-0.02 (-0.64)	-0.02 (-0.55)
ISCED 4	0.23*** (3.64)	0.08** (2.31)	0.07* (1.94)	0.08*** (2.99)	0.16*** (3.55)
ISCED 5+6	0.23*** (3.06)	0.12*** (3.05)	0.12*** (2.81)	0.21*** (5.31)	0.21*** (3.77)
Work Experience	0.05*** (6.89)	0.03*** (11.25)	0.03*** (7.35)	0.02*** (6.77)	0.01** (2.11)
Work Experience sq.	-0.00*** (-5.05)	-0.00*** (-6.52)	-0.00*** (-4.28)	-0.00*** (-3.56)	-0.00 (-0.29)
On the job training					
Yes	0.07** (2.09)	0.01 (0.47)	0.01 (0.72)	-0.02 (-1.15)	-0.06* (-1.80)
No			base		
Public Sector	0.02 (0.22)	0.06 (1.59)	0.09* (1.94)	0.00 (0.06)	0.06 (0.76)

In all regressions job (ISCO) and industry (NACE) classifications included; observations weighted with PIAAC poststratification weights
Significance * 0.1, ** 0.05, *** 0.01, t-Statistics in brackets.

Table 10: Quantile regressions with skills - hourly wages without bonus payments - results for males cont'd

	(1) 10%	(2) 25%	(3) 50%	(4) 75%	(5) 90%
Job status					
White-collar	0.23** (2.13)	0.14*** (2.77)	0.21*** (3.45)	0.10 (1.48)	0.10 (0.99)
Blue-collar	0.12 (1.03)	0.05 (1.03)	0.11* (1.69)	0.09 (1.22)	0.09 (0.80)
Civil Servant	0.22** (2.11)	0.14*** (3.14)	0.20*** (3.16)	0.13** (2.06)	0.09 (0.99)
Contracted Public Servant			base		
Firm size					
1 to 10 people	-0.25*** (-2.67)	-0.26*** (-6.30)	-0.24*** (-5.41)	-0.21*** (-4.41)	-0.21*** (-2.70)
11 to 50 people	-0.07 (-0.77)	-0.13*** (-4.96)	-0.15*** (-4.25)	-0.14*** (-3.37)	-0.22*** (-3.15)
51 to 250 people	-0.01 (-0.15)	-0.07*** (-2.79)	-0.13*** (-3.32)	-0.10** (-2.21)	-0.12** (-1.99)
251 to 1000 people	0.01 (0.07)	-0.02 (-0.59)	-0.04 (-0.97)	-0.04 (-0.78)	-0.07 (-1.12)
more than 1000 people			base		
Skills utilization					
None	-0.14 (-0.84)	-0.17 (-1.58)	-0.18 (-0.66)	-0.06 (-0.75)	-0.06 (-0.14)
Very little	-0.04 (-0.51)	-0.07 (-1.49)	-0.08 (-1.33)	-0.00 (-0.02)	0.00 (0.01)
Moderate	-0.06 (-1.21)	-0.07*** (-3.15)	-0.07** (-2.49)	0.01 (0.17)	-0.07 (-1.22)
High	-0.08** (-1.99)	-0.05*** (-2.95)	-0.06** (-2.45)	-0.07*** (-2.82)	-0.09** (-2.02)
Very high			base		
Citizenship & Background					
Austrian	-0.16 (-1.14)	-0.19*** (-3.10)	-0.19 (-1.48)	-0.05 (-0.64)	0.25 (1.50)
German	0.02 (0.12)	-0.10 (-1.07)	-0.09 (-0.73)	0.03 (0.13)	0.15 (0.93)
Serbian	-0.16 (.)	-0.11 (-0.40)	-0.18 (-0.68)	-0.00 (-0.02)	0.43 (1.18)
Turkish	-0.18 (-0.45)	-0.10 (-0.19)	0.08 (0.18)	0.08 (0.29)	0.11 (0.16)
Bosnian	0.01 (0.01)	-0.04 (-0.08)	-0.24 (-1.00)	-0.00 (-0.03)	-0.21 (-0.22)
Other			base		
Parents born in country	0.06 (0.29)	0.13*** (3.30)	0.04 (0.60)	-0.02 (-0.36)	-0.02 (-0.14)
1st generation migrant	-0.04 (-0.01)	-0.21 (-0.94)	-0.14 (-1.05)	0.10 (0.07)	0.55 (1.54)
Born in country	0.07 (0.01)	0.13 (0.66)	0.04 (.)	0.14 (0.11)	0.38 (1.29)
HoursWorked	-0.01*** (-4.04)	-0.01*** (-5.77)	-0.00*** (-2.70)	-0.00*** (-3.34)	-0.00* (-1.75)

In all regressions job (ISCO) and industry (NACE) classifications included; observations weighted with PIAAC poststratification weights
Significance * 0.1, ** 0.05, *** 0.01, t-Statistics in brackets.

Table 11: Quantile regressions with skills - hourly wages without bonus payments - results for females

	(1)	(2)	(3)	(4)	(5)
	10%	25%	50%	75%	90%
Children	0.04 (0.63)	-0.02 (-0.89)	-0.06*** (-2.86)	-0.01 (-0.32)	-0.03 (-0.60)
Age of children					
2 or younger	-0.05 (-0.59)	-0.00 (-0.02)	-0.02 (-0.38)	0.00 (0.02)	-0.01 (-0.13)
aged 3 to 5	0.01 (0.10)	0.06 (1.62)	0.03 (0.78)	0.00 (0.13)	-0.08 (-0.98)
aged 6 to 12	-0.09 (-1.18)	0.00 (0.11)	0.02 (0.50)	-0.03 (-1.11)	-0.05 (-0.97)
13 or older			base		
Skills					
Learning at work	-0.02 (-1.16)	-0.02*** (-2.93)	-0.02*** (-3.09)	-0.00 (-0.43)	-0.01 (-0.75)
Readiness to learn	-0.01 (-0.33)	-0.03*** (-3.49)	-0.01 (-1.47)	0.00 (0.36)	-0.00 (-0.05)
ICT	0.03 (1.29)	0.01 (1.29)	0.01 (1.30)	0.01 (1.34)	0.01 (0.75)
Influencing	-0.00 (-0.23)	-0.01 (-0.77)	0.01 (1.34)	-0.01 (-0.73)	0.00 (0.11)
Numerical	-0.01 (-0.64)	-0.00 (-0.29)	0.01 (1.32)	0.00 (0.30)	0.03** (2.32)
Planning	-0.01 (-0.60)	0.02** (2.49)	0.02*** (2.60)	0.01 (1.06)	-0.00 (-0.01)
Reading	0.02 (0.69)	0.03** (2.57)	0.02* (1.93)	0.03*** (3.37)	0.05*** (2.73)
Writing	0.03 (1.61)	0.02** (2.12)	0.01* (1.94)	0.00 (0.42)	-0.01 (-0.41)
Work flexibility					
Not at all	0.07 (0.87)	0.08** (2.37)	0.04 (1.48)	0.00 (0.16)	0.11** (1.99)
Very little	0.07 (0.84)	0.06* (1.77)	0.07** (2.16)	0.03 (1.12)	0.06 (0.97)
To some extent	0.07 (1.03)	0.06** (2.03)	0.04* (1.81)	-0.02 (-0.66)	-0.00 (-0.06)
To a high extent	0.02 (0.29)	0.06 (1.64)	0.04 (1.45)	0.00 (0.09)	0.07 (1.22)
To a very high extent			base		
Education					
ISCED 1			base		
ISCED 2	0.02 (0.32)	-0.06 (-1.27)	-0.08*** (-3.04)	-0.04 (-1.17)	-0.06 (-0.83)
ISCED 3	0.04 (0.88)	-0.02 (-0.53)	-0.05* (-1.82)	0.01 (0.38)	-0.03 (-0.56)
ISCED 4	0.06 (1.03)	0.02 (0.57)	-0.01 (-0.51)	-0.03 (-1.00)	-0.08 (-1.38)
ISCED 5+6	0.24*** (4.39)	0.12** (2.51)	0.09*** (3.07)	0.11*** (2.66)	0.09 (1.49)
Work Experience	0.04*** (4.42)	0.03*** (6.80)	0.03*** (9.77)	0.02*** (5.24)	0.02*** (3.54)
Work Experience sq.	-0.00*** (-3.31)	-0.00*** (-4.20)	-0.00*** (-4.98)	-0.00* (-1.74)	-0.00* (-1.76)
On the job training					
Yes	0.04 (0.99)	0.03 (1.40)	-0.02 (-1.16)	-0.03* (-1.86)	-0.04 (-1.00)
Public Sector	0.08 (0.87)	0.01 (0.47)	0.06** (1.97)	0.01 (0.19)	-0.01 (-0.08)

In all regressions job (ISCO) and industry (NACE) classifications included; observations weighted with PIAAC poststratification weights
Significance * 0.1, ** 0.05, *** 0.01, t-Statistics in brackets.

Table 12: Quantile regressions with skills - hourly wages without bonus payments - results for females cont'd

	(1) 10%	(2) 25%	(3) 50%	(4) 75%	(5) 90%
Job status					
White-collar	0.10 (1.15)	0.07 (1.42)	0.07** (2.27)	0.03 (0.86)	0.06 (0.89)
Blue-collar	0.07 (0.61)	-0.02 (-0.22)	0.03 (0.62)	0.00 (0.05)	0.04 (0.28)
Civil Servant	0.22** (2.51)	0.19*** (4.12)	0.10*** (3.45)	0.01 (0.29)	0.09 (1.42)
Contracted Public Servant			base		
Firm size					
1 to 10 people	-0.18 (-1.59)	-0.15*** (-4.09)	-0.11** (-2.52)	-0.14*** (-3.62)	-0.14* (-1.90)
11 to 50 people	-0.06 (-0.54)	-0.06* (-1.88)	-0.05 (-1.13)	-0.09** (-2.53)	-0.08 (-1.05)
51 to 250 people	-0.06 (-0.59)	-0.07** (-2.19)	-0.08* (-1.84)	-0.08** (-2.28)	-0.03 (-0.35)
251 to 1000 people	0.09 (0.86)	0.01 (0.36)	0.02 (0.36)	-0.02 (-0.49)	-0.02 (-0.31)
more than 1000 people			base		
Skills utilization					
None	-0.06 (-0.09)	-0.27 (-1.46)	-0.22 (-1.18)	0.09 (0.35)	-0.13 (-0.23)
Very little	-0.02 (-0.11)	-0.04 (-0.49)	-0.14*** (-3.06)	-0.13** (-2.08)	-0.06 (-0.54)
Moderate	0.01 (0.20)	-0.04* (-1.80)	-0.04 (-1.64)	-0.02 (-0.71)	-0.04 (-0.82)
High	0.04 (0.76)	-0.05** (-2.32)	-0.05** (-2.26)	-0.02 (-0.82)	-0.05 (-0.93)
Very high			base		
Citizenship & Background					
Austrian	-0.15 (-0.81)	-0.07 (-0.47)	-0.11 (-1.31)	-0.15 (-1.13)	-0.18 (-0.41)
German	-0.13 (-0.69)	-0.12 (-0.61)	-0.07 (-0.54)	-0.07 (-0.69)	0.00 (0.01)
Serbian	0.13 (0.05)	0.04 (0.17)	-0.16 (.)	-0.43 (.)	-0.62 (.)
Turkish	0.14 (.)	0.29 (.)	0.06 (.)	0.10 (.)	-0.29 (.)
Bosnian	0.44 (0.55)	0.13 (0.28)	-0.04 (-0.28)	-0.01 (-0.01)	-0.03 (-0.03)
Other			base		
Parents born in country	0.13 (0.64)	0.00 (0.00)	-0.06 (-1.35)	-0.09* (-1.69)	-0.13 (-0.81)
1st generation migrant	-0.02 (-0.04)	0.28 (0.16)	0.13 (0.23)	0.20 (0.13)	-0.78 (-1.13)
Born in country	0.10 (0.30)	0.22 (0.12)	0.02 (0.04)	0.11 (0.07)	-0.92 (-1.42)
HoursWorked	-0.00 (-1.31)	-0.00*** (-4.10)	-0.01*** (-6.89)	-0.01*** (-6.23)	-0.01*** (-4.12)

In all regressions job (ISCO) and industry (NACE) classifications included; observations weighted with PIAAC poststratification weights
Significance * 0.1, ** 0.05, *** 0.01, t-Statistics in brackets.