ISSN 2279-9362

Collegio Carlo Alberto

# Occupational Licensing and the Gender Wage Gap

Maria Koumenta
Mario Pagliero
Davud Rostam-Afschar

No. 631
December 2020

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

# Occupational Licensing and the Gender Wage Gap*

Maria Koumenta†
Mario Pagliero‡
Davud Rostam-Afschar§

October 1, 2020

**Abstract**

*We use a unique survey of the EU labor force to investigate the relationship between occupational licensing and the gender wage gap. We find that the gender wage gap is canceled for licensed self-employed workers. However, this closure of the gender wage gap is not mirrored by significant changes in the gender gap in hours worked. Our results are robust using decomposition methods, quantile regressions, different datasets, and selection correction.*

**Keywords** Licensing · Gender gap · Wages · Female Labour Supply · Quantile regression · Selection

**JEL Classification** J16 · J31 · J44 · J71

## 1. Introduction

Gender wage gaps have been a prominent theme within academic and policy debates for many decades. Despite some recent convergence in the earnings of men and women, the question remains as to what kind of policies are effective in closing the gap (Blau and Kahn 2017; Goldin 2014a; Olivetti and Petrongolo 2008; Blau and Kahn 2000). The role of the institutional context, including equal pay legislation, national minimum wages, unionisation and childcare subsidies, has been shown to go some way in correcting gender pay disparities. Goldin (2014b) has also suggested that the credentialisation of certain occupations through occupational licensing could reduce wage discrimination, but empirical evidence has been lacking so far.

Occupational licensing regulates the legal requirements for legally working in specific occupations. It is one of the most common forms of regulation in the labor market. In the European Union, occupational licensing directly affects 22 percent of workers (Koumenta and Pagliero 2019). Common examples

*We gratefully acknowledged comments and suggestions of conference participants at the AEA Annual Meeting 2020 and the EALE/SOLE/AASLE World Conference 2020. The usual disclaimer applies. Declarations of interest: none.

†Queen Mary, University of London (e-mail: m.koumenta@qmul.ac.uk),

‡University of Turin, Collegio Carlo Alberto, CEPR (e-mail: mario.pagliero@unito.it),

§University of Mannheim, 68131 Mannheim, Germany, University of Hohenheim (e-mail: rostam-afschar@uni-mannheim.de),

of licensed occupations include medical doctors, nurses, pharmacists, lawyers, architects, engineers, hairdressers, electricians and plumbers. The prevalence of licensing is similarly high in the US where, depending on the data source, it is estimated that between 20 and 29 percent of workers are licensed (Kleiner and Krueger 2010, 2013; Kleiner and Vorotnikov 2017).

There are three main channels through which occupational licensing may affect the gender wage gap. First, licensing typically imposes minimum human capital requirements for legally working in a profession. If the cost of meeting these requirements differs by gender, human capital accumulation may be differentially affected and gender disparities in labor market outcomes may increase (or be reduced). Second, by standardizing the human capital requirements for entry into a profession, licensing can correct for the information asymmetry associated with statistical discrimination. Third, occupational licensing generally requires passing a professional exam administered by a professional association or licensing board. If the process of obtaining a license is discriminatory, then occupational licensing regulations may provide more opportunities to one group of workers and contribute to the creation of gender differences. In spite of the potential importance of licensing in determining gender differences, there is currently no evidence on the role played by occupational licensing in determining gender differences, let alone on the relevance of these three specific channels.

In this paper, we provide the first systematic evidence on the relation between occupational licensing and gender wage gaps in the labor market. We also compare our results on licensing with those on certification, unionization, type of employment (employee versus self-employed), and other individual characteristics, such as educational attainment and type of occupation, which have been found to partially explain the wage gender gap. Our analysis is based on the European Union Survey of Occupational Regulation, a unique dataset covering the EU workforce that allows us to use a self-reported measure of occupational licensing, certification and unionization, as well as wages, hours worked, and other individual characteristics commonly observable in labor force surveys.

Our main results are as follows. First, using classic wage regressions, we find that the gender gap is about 16 percent on average for all workers, whereas it is completely canceled for self-employed licensed workers. For unionized employees we find a partial reduction in the wage gap, which remains at about 10 percent. Second, using Oaxaca-Blinder decompositions, we find that the explained component of the gap is small.[1] Still, gender differences in the proportion of licensed workers explain about 0.3 percentage points (2%) of the wage gap.[2] This is because there is a wage premium associated with licensing, and licensing is more prevalent among men. However, the wage structure effect (the unexplained component)

[1]This is consistent with the literature. For instance, in a meta-analysis, Weichselbaumer and Winter-Ebmer (2005) show that more than 2/3 of the wage gap is attributed to the unexplained component in many of the more than 200 studies with data from the 1980s and 1990s.

[2]In relative terms, this is about 27% of the component of the wage gap explained by occupation fixed effects.

of licensing amounts to -0.5 percentage points (-3%) of the wage gap. This negative effect reflects the closure of the gender wage gap for licensed self-employed workers. Although the total effect is negative, the decomposition results point towards an ambiguous role of licensing in explaining overall gender wage differences.

With respect to other institutions, we find that certification explains 0.1 percentage points, while unionization -0.2 percentage points. Third, using quantile regressions, we find that our main results are robust throughout the wage distribution. However, the licensing premium increases when moving towards higher quantiles in the wage distribution. A similar effect occurs for the certification premium, whereas unionization seems to be more beneficial at lower wages. Fourth, the closure of the gender wage gap observed for self-employed licensed workers is not mirrored by significant changes in the gender gap in hours worked.

Finally, we find that our results are robust to sensitivity checks. Using two different datasets, namely the Skills and Employment Survey for the UK and the Socio-Economic Panel for Germany, we are able to replicate our main findings. This suggests that our results are not driven by specific features of how our data are collected or how occupational licensing is measured in our main dataset. We also show that our results are robust when accounting for selection into self-employment.

Although our results are descriptive and cannot be interpreted as causal, they are important for three main reasons. First, they highlight occupational licensing as a possible new determinant of the wage gap, neglected until now in the literature on gender wage differentials. Second, they provide the first evidence that licensing might have an important quantitative effect, possibly together with other variables describing labor market institutions and the type of work arrangement (employee versus self-employed). Finally, they suggest that licensing operates very differently from other labor market institutions such as unions.

The study is structured as follows. Section 2 reviews the related literature. Section 3 describes the data. Section 4 discusses the empirical strategy and results. We discuss sensitivity of the results to different data sources and measures of licensing as well as selection on unobservables in Section 5. Section 6 concludes.

## 2. Related literature

Licensing defines the minimum human capital requirements for legally working in a profession. These may be in the form of educational qualifications, work experience and the passing of an exam. Certification is less restrictive in that it imposes no restrictions on the right to practice a profession, but allows individuals to voluntarily apply to be certified as meeting the standards of a professional or regulatory body. The wage effects of licensing are well-documented in the literature. Studies have consistently shown that licensing is associated with a wage premium ranging from 3-18% in the US depending on the data source (Kleiner and Krueger 2010, 2013; Gittleman and Kleiner 2016; Kleiner and Vorotnikov 2017), 4% in the European context (Koumenta and Pagliero 2019), between 8-11% in Germany (Bol and Weeden 2014) and between

9-19% in the UK depending on the occupational group (Koumenta et al. 2014). The precise mechanism that may account for the observed wage premium is less clear. From a human capital perspective, if licensing were not mandatory, the wage premium associated with it would reflect returns to education. However, if licensing increases the skills equilibrium within the occupation and keeps less competent practitioners out of the market, it may also reflect monopoly rents accruing to insiders as a result of labor supply restrictions into the occupation.

Turning to gender wage gaps, economists have commonly sought to explain gender wage disparities by examining gender differences in human capital and occupational choices (Mincer and Polachek 1974; Polachek 1981), with the unexplained residual being attributed to discrimination. Within this literature, human capital theory attributes the wage penalty to women acquiring less formal education and training and choosing occupations that require low occupation-specific investments in skills. Labor market discrimination approaches instead locate the female disadvantage in models of statistical discrimination assumed to be the outcome of imperfect means of observing productivity (Blau and Kahn 2017; Altonji and Blank 1999).

Against the backdrop of this literature, our starting point is that the economic and legal arrangements governing occupations affect occupational wage differentials and the structure of earnings (Koumenta and Pagliero 2019; Kleiner and Krueger 2010; Card 2001; Gittleman et al. 2018). Any rents or higher returns to education accruing to licensed occupations may thus impact the relative wages of men and women if these groups differ in their capacity to access these occupations. Such barriers to entry could be artificial, for example in the form of discriminatory practices by those controlling the licensing process, or they could arise from differences in ability between the two groups.

Our interest in whether licensing affects the gender wage gap is further motivated by its signaling properties. Becoming licensed is a costly endeavor (Carpenter et al. 2018). Typical requirements include education, work experience, and written and practical examinations. Additional costs include application, processing and licensing fees, and of course the opportunity cost in terms of forgone income while in training. From a human capital perspective, it follows that a woman entering a licensed occupation is signaling not only levels of competency analogous to those of her male counterparts, but also similar labour market commitment as evidenced by these investments. Further, in standard models of statistical discrimination, pay disparities between men and women arise from the average differences in the expected value of the productivity of these two groups or in the accuracy with which productivity can be predicted (Arrow 1973; Phelps 1972). By standardizing the human capital requirements for entering the occupation, the signaling properties of licensing in relation to both skills and ability can correct for the information asymmetry problem associated with statistical discrimination. Goldin (2014b), for example, argues that credentialisation of occupations might eliminate the negative signal provided by hiring a woman, particularly if the credential were well-known and verifiable. Some scant evidence of this proposition already exists, though not for the gender wage gap. Law and Marks (2009) find that licensing reduced the share of women in teaching but increased it in other occupations (engineers, pharmacists). Blair and Chung (2018) show that licensed African-American men have higher wages than their non-licensed counterparts and attribute this to the signaling effect of licensing in relation to non-felony status. Certification can have a similar effect, while the union's egalitarian wage policies and pay scales attached to jobs rather than individuals are also expected not only to put an upward pressure on female wages, but also to limit the scope for employers to discriminate (Metcalf et al. 2001; Machin 1999). In addition to employer discrimination, licensing may reduce incomplete information about self-employed workers and hence consumer discrimination. For instance, Borjas and Bronars (1989) develop a model of incomplete information that explains why minorities are negatively selected into self-employment, whereas whites exhibit positive selection.

The licensing literature is largely silent about the effect of licensing on the wages of those in self-employment both in aggregate and by gender. This is despite self-employment accounting for a sizable proportion of licensed workers and evidence that the removal of licensing requirements in 53 professions in Germany increased the propensity to work in self-employment Rostam-Afschar (2014). Therefore, we put particular focus on how licensing differentially affects the gender wage gap of employees and the self-employed. In the broader literature on gender pay, there is consensus that the gender wage gap observed for employees is also evident among the self-employed (Moore 1983; Haber et al. 1987; Sexton and Robinson 1989). Moore (1983) formulates testable predictions for employer discrimination: self-employment as a method of avoiding racist (or sexist) employment practices should result in a higher female/male earnings ratio among self-employed workers than their wage and salary counterparts but cannot find evidence for it. Hundley (2000), shows that a large part of the gap is due to women being under-represented in more rewarding areas of self-employed work, such as professional services and trades, and attributes this to women avoiding entering markets where licensing requirements impose additional hurdles for entry. However, whether the earnings gap between the two sexes is different when women enter licensed occupations remains an open question.

## 3. Data and summary statistics

The European Union Survey of Occupational Regulation (EU-SOR) is specifically designed to capture occupational regulation. It covers the individuals residing in the 28 EU member states, aged 15 and above. The survey was carried out by TNS (a market research company) in March and April 2015 by means of telephone interviews (Computer Assisted Telephone Interviews). A total of 26,640 individuals were interviewed providing data on their licensing, certification and trade union membership status. Licensing is defined as having obtained a license or passed an exam required in addition to education to legally practice a profession. Certification is defined as having obtained a license or passed an exam, in addition to education, which is not required to legally practice a profession. We define unionization as being a member of a trade union.[3] For details, see Table A.1 and Appendix E. Detailed information on a variety of individual characteristics, similar to those commonly included in labor force surveys, was also collected. These include net wages, hours worked, age, educational attainment, occupation, type of work arrangement (employee versus self-employed), country of residence, industry in which the firm or organization operates, and its size.

[TABLE 1 ABOUT HERE]

Table 1 provides descriptive statistics of the differences between females and males across the 28 European countries in our sample. Female hourly wages are on average substantially lower than those of their male counterparts. This difference is statistically significant according to t-tests reported in the right-most columns. The number of hours worked per month also differs significantly: women work on average over 20 hours less than men. There are also disparities regarding the type of employment. The fraction of self-employed women is only 11%, while 18% of men are self-employed. The self-employed, who are well represented in our sample, constitute an important portion of the overall labor force in the EU.[4] Female and male workers are roughly equally attached to the three labor market institutions we study. About 25% of females and slightly fewer men report being a union member. Licensing covers 21% of women and 22% of men. Certification is slightly less common, covering 19% of women and 22% of men. A total of 35% of women and 33% of men are neither licensed, certified, nor union members.

Figure 1a and 1b describe the wage distributions for male and female non-licensed workers. In each, the distribution for employees and self-employed workers is indicated separately. Figure 1c and 1d show the wage distributions for licensed workers. The vertical bars represent the mean wage for each group of workers. In general, licensing is associated with higher wages (the licensing premium, see Koumenta and Pagliero (2019)). This can be clearly seen comparing Figures 1a and 1b for non-licensed workers with the corresponding figures for licensed workers (figures 1c and 1d). Self-employment is also associated with higher wages, which can be seen comparing the vertical bars representing mean wages within each figure. While this self-employment premium is similar for licensed and non-licensed males (figures 1b and 1d), it is substantially larger for licensed than for non-licensed females. The large wage premium associated with female self-employment in licensed occupations is an important and robust empirical regularity that has a large impact on our results on the gender wage gap described in the following sections.

[FIGURE 1 ABOUT HERE]

[3] Since our focus is on licensing, we do not address here the many issues related to differences in union coverage across occupations and countries.

[4] Overall, we observe 1,263 self-employed female who represent 11,810,015 individuals (using the sample weights provided by EU-SOR).

## 4. The Gender Wage Gap

### 4.1 Wage regressions

In order to describe the mean gender wage gap across labor market institutions (licensing, certification, and unionization) and type of employment (employees and self-employed workers), we estimate a classic Mincer (1958) wage regression model,

$$\begin{aligned}
\log(w_{ionc}) &= \beta_0 + \beta_1 X_{ionc} + \beta_2 \text{Female}_{ionc} \\
&+ \beta_3 \text{Self-Employed}_{ionc} + \beta_4 \text{Female}_{ionc} \times \text{Self-Employed}_{ionc} \\
&+ \sum_{j=1}^{3} \gamma_j \text{Institution(j)}_{ionc} + \sum_{j=1}^{3} \delta_j \text{Female}_{ionc} \times \text{Institution(j)}_{ionc} \\
&+ \sum_{j=1}^{3} \lambda_j \text{Self-Employed}_{ionc} \times \text{Institution(j)}_{ionc} \\
&+ \sum_{j=1}^{3} \mu_j \text{Female}_{ionc} \times \text{Self-Employed}_{ionc} \times \text{Institution(j)}_{iocn} \\
&+ \theta_o + \theta_n + \theta_c + \varepsilon_{ionc}
\end{aligned} \tag{1}$$

where the dependent variable $\log(w_{ionc})$ denotes net hourly log-wage of individual $i$, in occupation $o$, industry $n$, and country $c$. The vector $X_{ionc}$ includes individual characteristics (age, age$^2$, indicators for 6 levels of education achieved, and the size of the firm or organization in which the individual is working).

The indicator variables Female$_{ionc}$ and Self-Employed$_{ionc}$ are equal to one if the respondent is female and self-employed respectively. Three indicator variables, Institution(j)$_{ionc}$, j=1,...3, describe whether the worker is licensed, certified, or unionized. The model includes all the direct effects and the three-way interactions between Female$_{ionc}$, Self-Employed$_{ionc}$, and Institution(j)$_{ionc}$. Finally, $\theta_o$, $\theta_n$, and $\theta_c$ are occupation-, industry-, and country-specific fixed effects, and $\varepsilon_{ionc}$ captures unobserved determinants of wages.

The coefficients $\beta_2$, $\beta_4$, $\delta_j$, $\mu_j$ capture the average gender wage differences across labor market institutions and type of employment. Figure 2 graphically describes our main results based on the OLS estimates reported in Table B.2. The gender wage gap is measured on the vertical axes (in percent of average male wage) for employees and for self-employed workers in the two panels, respectively. The figure shows that for employees who are neither licensed, certified, nor unionized, the predicted gender wage gap is 16.4%, after controlling for individual characteristics and fixed effects. The gender gap is not significantly different for licensed and certified workers, while it is substantially lower, about 10%, for unionized workers. Turning to the self-employed, the gender wage gap is about 20% for all groups of workers, with the remarkable exception of licensed workers, for whom the wage gap is completely canceled. The point estimate of the wage gap for self-employed licensed workers in Figure 2 is 2.6% (not significantly different from zero at conventional levels). For unionized workers, the standard errors are too large to produce conclusive results, since few self-employed workers are union members.

[FIGURE 2 ABOUT HERE]

Complementing Figure 2, Table B.2 shows that the closure of the wage gap for self-employed licensed workers is due to significantly higher wages for female workers (not lower wages for males). Using the results in column IX of Table B.2, the gender gap for these workers (2.6%) can be computed as the sum of the coefficients of Female -0.164, Female $\times$ Licensing -0.006, Female $\times$ Self-Employed -0.084, Female $\times$ Self-Employed $\times$ Licensing 0.228. Hence, the higher wages for female self-employed licensed workers compensate for the lower wages that women have on average in the labor market.

Table B.2 also provides five additional insights. First, our estimates of the gender gap are robust to the inclusion of different sets of regressors in model (1). Including occupation and industry fixed effects reduces the gap only slightly, and it remains similar regardless of which institution or interaction terms are included. Overall, our estimates of the gender gap for European workers are in line with those obtained in the literature (see, e.g. Olivetti and Petrongolo 2008; Weichselbaumer and Winter-Ebmer 2005). Second, on average, licensed workers earn about 5% higher wages. This is consistent with a large literature on the wage effects of occupational licensing (Kleiner 2000; Kleiner and Krueger 2013; Koumenta and Pagliero 2019). Third, certified workers also enjoy a wage premium, although it is lower than that of licensed workers (Koumenta and Pagliero 2019). Fourth, there is a wage premium for union members, which is significantly different from zero (about 7%) for female workers alone (Bryson et al. 2020). Fifth, there is a substantial premium for self-employment (cf. Rostam-Afschar and Strohmaier 2019) of about 17% higher wages for men and about 9% for women.

## 4.2 Oaxaca Blinder decompositions

Wage regressions using model (1) constrain some of the coefficients to be the same for male and female workers. We relax this assumption using a standard Oaxaca-Blinder (Blinder 1973; Oaxaca 1973) decomposition, which requires estimating a linear model similar to (1) separately for the sample of male and female workers. The decomposition is also interesting as it provides an estimate of the part of the wage gap that is due to worker characteristics (composition effect or explained component) and the estimated coefficients (wage structure effect or unexplained component).

We decompose the average wage gap as follows:

$$\begin{aligned}\Delta_{M-F} &= [E(X_M - X_F)]'\beta_{\text{Pooled}} \\ &+ [E(X_M)'(\beta_{\text{Pooled}} - \beta_M) - E(X_F)'(\beta_{\text{Pooled}} - \beta_F)],\end{aligned} \tag{2}$$

where $\Delta_{M-F}$ denotes the hourly wage differential, $E(X_M - X_F)$ the average difference in characteristics, which gives the explained component if multiplied by a coefficient vector. The second term is the unexplained component, commonly attributed to discrimination, but which also includes potential effects of unobserved variables. Following Neumark (1988), we estimate the coefficients $\beta_{\mathrm{M}}$, $\beta_{\mathrm{F}}$, and $\beta_{\mathrm{Pooled}}$ from the sample of males, females, and the pooled sample, respectively. The matrix of regressors $X$ in the decomposition includes the same explanatory variables used in the wage regressions.

[TABLE 2 ABOUT HERE]

The results from the decompositions are presented in Table 2. The top panel shows the mean predicted wage by group and their difference. In our sample, the mean of the log wages is 1.963 for men and 1.797 for women. The wage gap is 0.166. For each model specification (columns I-IV), the table reports the total explained and unexplained components and the detailed decomposition for the interaction between the institution (licensing, certification, and unionisation) and the self-employed dummies. The table also reports the two components for the other individual characteristics, and the fixed effects.

Overall, the results show that the explained component is small. Going into more detail, column IV shows that the difference in prevalence of occupational licensing among men and women explains about 0.3 percentage points (2%) of the wage gap.[5] Certification explains 0.1 percentage points, and unionization -0.2 percentage points (hence, it "reduces" the wage gap).[6] In comparison, individual characteristics explain -1.6 percentage points. Country (2.4%) and industry (1.4%) fixed effects contribute positively to the wage gap, while occupation fixed effects contribute negatively (-1.1%). Hence, in relative terms, the part of the wage gap explained by licensing is about 19% of that of the individual characteristics included in the model or about 27% of that of occupation fixed effects.

The unexplained components (due to differences in estimated coefficients) are generally not significantly different from zero. One important exception is the negative unexplained component for the interaction of the Licensing and Self-employed dummies, which accounts for -0.5 percentage points (-3%) of the wage gap. This result captures the fact that self-employed licensed women earn significantly more than self-employed unlicensed women, which cancels the gap for licensed self-employed workers (see Figure 2 and Table B.2). Hence, the decomposition indicates that licensing significantly reduces the unexplained component or wage structure effect. Overall, the role of licensing in explaining the wage gap is ambiguous,

[5] For licensing, the explained component of 0.002 is the share of licensed among men and women, 0.171-0.14, times the coefficient, 0.62, from the pooled regression, ignoring the interaction. The unexplained component is obtained as $0.14 \times (0.062 + 0.006 - 0.052) - 0.171 \times (.062 - 0.052)$.

[6] Unionization has a positive effect on wages and women are, on average, slightly more often union members than men. Without this unionization advantage of women, they would be even worse off and, hence, the overall wage gap would increase.

although in our estimates the negative wage structure effect dominates the positive effect on the explained component, leading to a negative total effect of -0.2 percentage points.

### 4.3 Quantile regressions and quantile decomposition

In this section, we investigate how the gender wage gap and the explanatory power of covariates in wage regressions vary over the wage distribution. We estimate model (1) using quantile regressions (Koenker and Bassett 1978) and report the estimated coeffcients in Table 3.

[TABLE 3 ABOUT HERE]

Overall, our main findings from wage regressions hold throughout the wage distribution. Still, some interesting results emerge. First, the licensing premium increases moving towards higher quantiles in the wage distribution. Second, a similar effect occurs for the certification premium, while unionization is associated with higher wages at the bottom of the wage distribution. Third, the large wage increase associated with female self-employed licensed workers in wage regressions is apparent throughout the wage distribution, possibly slightly larger at lower quantiles.

We also decompose the gender wage gap over the wage distribution using the method developed in (Chernozhukov et al. 2013). This extends the decomposition of the average gender gap (Section 4.2) to different quantiles of the wage distribution. The quantile decomposition is based on a counterfactual wage distribution that is constructed by asking what the distribution of wages for females would be if the conditional distribution of wages was the same as that estimated for males. Any differences from the fitted wage distribution of males resulting from the covariates of females is attributed to the explained component or composition effect. Analogous to the case of the decomposition of the mean, the remaining differences due to differences in the wage functions conditional on characteristics constitute the unexplained component or wage structure effect.

The unconditional distribution of log wages $F_{W_{\mathrm{M,M}}}$ with the male wage function and the male characteristics is contrasted with the corresponding distribution for females, $F_{W_{\mathrm{F,F}}}$,

$$F_{W_{\mathrm{M,M}}} - F_{W_{\mathrm{F,F}}} = \left[F_{W_{\mathrm{M,M}}} - F_{W_{\mathrm{M,F}}}\right] + \left[F_{W_{\mathrm{M,F}}} - F_{W_{\mathrm{F,F}}}\right], \tag{3}$$

where $F_{W_{\mathrm{M,F}}}$ is the counterfactual wage distribution that women would face if they were compensated according to the male wage function. Given a distribution of male wages conditional on their characteristics $F_{W_{\mathrm{M}}|X_{\mathrm{M}}}$, where $F_{X_{\mathrm{M}}}$ is the distribution of male characteristics, the unconditional distribution of log wages with the male wage function and the male characteristics can be computed. The first term in equation (3) then gives the effect of differing distributions of characteristics, while the second term shows the unexplained component or wage structure effect.[7]

We estimate quantile functions for the 10th through the 90th percentile. As before, the dependent variable is log hourly wage including control variables for employment status, labour market institution, their interaction with employment status, education, age, its square, firm size, indicators for occupation, country, and industry. 95% confidence intervals are based on bootstrapped standard errors obtained with 100 replications.[8]

[FIGURE 3 ABOUT HERE]

The quantile decomposition provides some new insights. Figure 3a shows the wage gap (sum of both components) measured at different quantiles of the wage distribution in comparison to the average wage gap from the Oaxaca-Blinder procedure (along with Eicker-Huber-White 95% confidence intervals). It is apparent that the gender wage gap varies between 20% and 14% over the wage distribution, a variation that cannot be captured by the mean decomposition.

Figure 3b compares the explained and unexplained components of the gender wage gap at different quantiles of the wage distribution. As in the mean decomposition, most of the difference is attributed to the unexplained component. Although covariates explain little of the wage gap on average, they explain about -0.05 percentage points at the bottom and 0.05 at the top of the wage distribution. Hence, the Oaxaca-Blinder decomposition masks some of the explanatory power of the covariates. The larger explained component at the top of the wage distribution is consistent with a lower representation of women in highly paid occupations and differences in individual characteristics required in highly paid jobs.

The quantile decomposition is consistent with our previous results from mean decompositions and quantile regressions. For example, the quantile regression results for the median in Table 3 imply a predicted wage gap of 0.136. This corresponds to the unexplained component at the median resulting from the quantile decomposition desribed in Figure 3 (also about 0.13).

Finally, Figure 4 illustrates a detailed decomposition of the explained component of the gender wage gap. The model includes the same covariates as in the Oaxaca-Blinder decomposition arranged in three groups

[7] To illustrate how the counterfactual distribution is constructed, consider the unconditional distribution of wages ($w$) as the integral of the wage distribution conditional on characteristics $x$. Then $F_{W_{\mathrm{M,M}}}(w) = \int F_{W_{\mathrm{M}}|X_{\mathrm{M}}}(w|x)dF_{X_{\mathrm{M}}}(x)$ and $F_{W_{\mathrm{F,F}}}(w)$ is obtained analogously for females. Using the distribution of female characteristics $F_{X_{\mathrm{F}}}$ instead of $F_{X_{\mathrm{M}}}$ gives the counterfactual distribution $F_{W_{\mathrm{M,F}}}(w) = \int F_{W_{\mathrm{M}}|X_{\mathrm{M}}}(w|x)dF_{X_{\mathrm{F}}}(x)$.

[8] The estimation method is different from quantile regressions in that proportions, not quantiles, are estimated and then inverted globally back to quantiles, i.e. using the entire distribution and not only a single quantile. The conditional quantile function is monotonized using the re-arrangement method suggested by Chernozhukov et al. (2010). This allows to invert the quantile function to obtain the conditional distribution function.

to ease exposition. The first (licensing) includes the licensing dummies and its interactions, the second (certification and unionization) includes the dummies for the other two institutions and their interactions, and the third includes all the remaining covariates (individual characteristics and fixed effects).

In Section 4.2 we showed that the different prevalence of licensing among men and women explains part of the mean wage gender gap. Figure 4 qualifies this result by showing that the positive explained component at the mean is the result of a larger explained component at the bottom than at the top of the wage distribution. The second group of covariates (certification and unionization) also explains more of the wage gap at the bottom than at the top of the wage distribution. Finally, individual characteristics and fixed effects (for occupation, industry, and country) explain a growing proportion of the wage gap as we move towards higher quantiles of the wage distribution. This large heterogeneity across quantiles of the wage distribution cannot be captured by the decomposition at the mean.

[FIGURE 4 ABOUT HERE]

### 4.4 The Gender Hour Gap

To understand how the different labor market institutions work, it is interesting to study the gender hour gap. For example, if the closure of the gender wage gap for self-employed, licensed women was accompanied by a similar change in hours, then the gender gap in income would be closed as well. It is also interesting to ask whether the premia on wages gained through certification and union membership are associated with more hours of work. To this end, we estimate an OLS hours regression similar to the wage regression used above.

$$\begin{aligned} Hours_{ionc} &= \varphi_0 + \varphi_1 X_{ionc} + \varphi_2 \text{Female}_{ionc} \\ &+ \varphi_3 \text{Self-Employed}_{ionc} + \varphi_4 \text{Female}_{ionc} \times \text{Self-Employed}_{ionc} \\ &+ \sum_{j=1}^{3} \eta_j \text{Institution(j)}_{ionc} + \sum_{j=1}^{3} \zeta_j \text{Female}_{ionc} \times \text{Institution(j)}_{ionc} \\ &+ \sum_{j=1}^{3} \kappa_j \text{Self-Employed}_{ionc} \times \text{Institution(j)}_{ionc} \\ &+ \sum_{j=1}^{3} \nu_j \text{Female}_{ionc} \times \text{Self-Employed}_{ionc} \times \text{Institution(j)}_{iocn} \\ &+ \vartheta_o + \vartheta_n + \vartheta_c + u_{ionc} \end{aligned} \tag{4}$$

where the dependent variable $Hours_{ionc}$ denotes monthly hours worked by individual $i$, in occupation $o$, industry $n$, and country $c$. The regressors $X_{ionc}$ are identical to those in equation (1). $u_{ionc}$ captures unobserved factors in the error term.

Figure 5 shows the predicted gender hour gap for employees (left-hand graph) and the self-employed (right hand graph), while the estimated coefficients are reported in Table B.3.

[FIGURE 5 ABOUT HERE]

Female employees who are not associated to any of the three institutions work on average 14 hours per month less than men (after controlling for individual characteristics and fixed effects). Differences across institutions are rather small. Relative to unregulated workers, the gap increases for licensed and certified employees by 3.6 and 2.7 hours per month respectively, while it decreases by 2.6 hours per month for union members. Differences in the gender hour gap are also relatively small in the right hand side graph, which shows gender gaps for self-employed workers. Interestingly, the canceling of the wage gap for licensed self-employed woman (described in Section 4.1) does not translate into a reduction in the gender hour gap.

Table B.3 shows that the licensing premium and the self-employment premium (documented in the wage regressions) are associated with longer working hours (4.6 and 12 hours respectively in column IX).[9] Also certification, and to a lesser extent unionization, are associated with longer hours (4.2 and 2.1 hours), reflecting the higher wage observed for these workers in wage regressions. However, the higher wages for self-employed licensed women observed in wage regressions do not correspond to longer working hours. The coefficient for the interaction of the Female, Self-employed, and Licensing dummies is not significantly different from zero (the point estimate is -1.87 with a standard error of 6.29).

## 5. Robustness checks

Our main finding from Section 4 suggests a large wage premium for self-employed licensed women. To corroborate the reliability of this result, we described earlier the set of different techniques employed (OLS regressions, quantile regressions, counterfactual distribution regressions) to estimate different specifications. As a further check of whether our main finding survives when using very different data from that of the EU-SOR, we replicate our estimations using datasets commonly used for studying labour markets. Since to our knowledge no dataset provides a self-reported measure on licensing and certification for the EU and classification of licensing on the occupational level is not readily available, we use two countries representing different styles of occupational regulation (of English and German legal origin, see Pagliero 2019): Germany and the UK.

The German Socio-Economic Panel (GER-SOEP), provides detailed information on all the variables of interest (see Wagner et al. 2007), with the exception of whether licensing is legally required to work. We address this by compiling a database of more than 250 licensed 4-digit occupations using information from Rostam-Afschar (2015), the German Federal Employment Agency and the European Commission database on regulated professions. We use all of the annual waves from 1984 through 2018 and assign for our estimations to each occupation code a binary indicator for being licensed. For the UK, we pool comparable

[9] The fact that self-employed work longer hours been attributed to precautionary labor supply (Jessen et al. 2018).

cross-sections from the 1992, 2001, 2006, and 2012 iterations of the Skills and Employment Survey (UK-SES) (see Felstead et al. 2014). As with the SOEP, the survey has no information on the regulation status of the respondent. We address this using the UK Database of Regulated Occupations (Koumenta et al. 2014), a hand-collected database put together through desk research on the regulatory status of the 353 4-digit occupation unit groups defined by the UK Office for National Statistics' Standard Occupational Classification system (SOC 2000).

[TABLE 4 ABOUT HERE]

We estimate the same model specifications for both wages and hours as before (see Section 4.1), with the exception of certification, which is not available in the UK-SES and the GER-SOEP data. We then include further variables not available in the EU-SOR, particularly indicator variables for the presence of dependent children, in a step-wise fashion.

Table 4 presents the main findings from our sensitivity checks (the full set of results is reported in Table B.4). Both in the UK-SES and the GER-SOEP, the large wage effect for self-employed licensed women is present. The coefficients are of similar size and statistically indistinguishable from each other. The gender wage gaps in the UK and Germany are somewhat larger than the EU average (which is reported for comparison in column 1). This might be due to the focus on the more recent period of observation of the EU-SOR.

Comparing the results from hour regressions (columns 4-6), we find that women work about 10 fewer hours per month than men in the UK, and 40 hours less in Germany, and in line with existing research (again recall the different periods of observations). Our results from the EU-SOR (column 4) are somewhat in between. There is a similar ordering of the estimated coefficients for the interaction of Female, Self-Employed, and Licensed dummies, although they are not significantly different from zero. Even though self-employed licensed women have a significantly higher wage, they do not respond by working more hours. We conclude that our results are not sensitive to the specificities of any of our datasets or the methodology used for measuring occupational licensing.

## 5.1 Selection

The decision to become self-employed may be determined by characteristics such as ability, preference for locus of control, or attitudes towards risk unobserved in our dataset. This can be accounted for econometrically with methods like the Heckman (1979) selection correction for the conditional expectation and similar techniques for quantile regressions. We follow the approach of Biewen et al. (2020) who extend the Buchinsky (1998, 2001) control function method to correct for selection of females into full-time employment. In contrast, we model selection into self-employment. We include $Z$, a vector of the standard control variables described above plus a set of additional variables assumed to influence selection into self-employment in this first stage regression. Our choice of the exclusion restriction is similar to those of the literature (Biewen et al. 2020): We include regional unemployment rates obtained from Eurostat (at NUTS1 or NUTS2-level, as recorded in the survey). The variable regional unemployment is an indicator for the tightness of the local labor market.

Consider two groups of workers. The first (All) includes all workers, the second (SE) only comprises self-employed workers. The conditional wage quantiles in the entire population can then be written as

$$Q_\tau(y_{\text{All}}|X_{\text{All}}) \quad = \quad X'_{\text{All}}\chi_{\text{All},\tau} + h_\tau(Z'_{\text{SE}}\mu),$$

where $Q_\tau(y_{\text{All}}|X_{\text{All}})$ denotes the $\tau$-quantile of the distribution of potential log wages $y_{\text{All}}$ for workers with characteristics $X_{\text{All}}$. This is the distribution of wages a randomly-drawn worker with characteristics $X_{\text{All}}$ would face if she decided to be self-employed. The vector $\chi_{\text{All},\tau}$ represents the quantile-specific coefficients for this distribution of potential wages. Since not all workers actually work in self-employment, we can only use the wages of workers who indicate to be self-employed, which implies a non-randomly selected subpopulation. To correct for this selection problem, we include the selection correction term $h_\tau(Z'_{\text{SE}}\mu)$ at each quantile, where $Z'_{\text{SE}}$ is a vector of individual characteristics explaining participation in self-employment. Participation in self-employment (cf. Rostam-Afschar 2014) is described as follows:

$$Pr(\text{SE} = 1|Z) \quad = \quad Pr(Z'\mu + \nu > 0|Z), \tag{5}$$

where $\mu$ is a vector of parameters and $\nu$ is an error term of the selection equation, possibly correlated with the error term in the quantile wage regressions.

[FIGURE C.1 ABOUT HERE]

Figure C.1 presents visual evidence for our selection model for female participation in self-employment. Table C.5 reports the coefficient estimates. Column I displays the probit coefficients and column II the average marginal effects. There is a gender gap with respect to participation in self-employment. Women are less likely to be self-employed by 4.0 percentage points. The probability to be in self-employment is 5.6 percentage points higher for licensed workers, and 12.6 percentage points lower for union members. The coefficient of regional unemployment (the instrument in the selection model) is significantly different from zero.[10]

[10]The Sanderson and Windmeijer (2016) F-test rejects that the first stage is weakly identified (F-Statistic 10.59, p-value 0.001).

Analogous to the Heckman (1979) approach for the conditional expectation, the selection correction term $\hat{h}_\tau$ at quantile $\tau$ involves the inverse Mills ratio. Following (Newey 2009), we approximate the selection correction term from a prediction of the first-stage regression (5) using a power series with parameters $\pi$,

$$\hat{h}_\tau(Z'_{\mathrm{SE}}\mu) = \pi_{0,\tau} + \pi_{1,\tau}\lambda(Z'_{\mathrm{SE}}\mu) + \pi_{2,\tau}\lambda(Z'_{\mathrm{SE}}\mu)^2 + \pi_{3,\tau}\lambda(Z'_{\mathrm{SE}}\mu)^3 \tag{6}$$

where $\lambda$ denotes the inverse Mills ratio.[11]

[TABLE D.6 ABOUT HERE]

Table D.6 reports the quantile regression results. For comparison, the last column shows the estimates of the two-step Heckman selection correction (Heckit) for the conditional expectation. Overall, the results are in line with previous results, showing a closing of the wage gap for self-employed workers. However, while there is complete canceling of the gap at the top and the bottom of the wage distribution, the closure of the gap is only partial in the middle of the wage distribution, where selection seems to play a larger role.[12]

## 6. Conclusion

Although our results are descriptive and do not capture causal effects, they show that occupational licensing has a potentially significant impact on the gender wage gap, possibly in conjunction with other variables capturing labor market regulations, institutions, and type of work arrangements (self-employed versus employees) and in different ways across the wage distribution. In particular, we showed that the gender wage gap disappears for the licensed self-employed. This is a new fact about gender gaps, and one that is quite robust to a number of estimation techniques and the use of different data sets.

Understanding the channel through which licensing might affect the gender gap is beyond the objectives of this paper. The canceling of the wage gap for licensed self-employed workers might be related to

[11] As pointed out by Huber and Melly (2015), the error terms of the wage equation cannot be heteroscedastic in $Z$ but only in the determinants of selection to self-employment $Z'_{\mathrm{SE}}$. In order to estimate a model that does not rely on this restrictive assumption, we follow Biewen et al. (2020) and transform our model by dividing the wage equation by $\eta(Z'_{\mathrm{SE}}, Z'_{\mathrm{SE}}\mu)$. This scaling factor is based on two quantiles $\alpha_1 = 0.15$ and $\alpha_2 = 0.85$ from which we calculate the interquantile spread for self-employed women with characteristics $Z_{\mathrm{SE}}\Delta q(Z_{\mathrm{SE}}) = Q_(\alpha_2)(y_{\mathrm{SE}}|Z_{\mathrm{SE}}) - Q_(\alpha_1)(y_{\mathrm{SE}}|Z_{\mathrm{SE}})$, where $Q_\tau(y_{\mathrm{SE}}|Z_{\mathrm{SE}})$ denotes the $\tau$-quantile of the distribution of potential log wages $y_{\mathrm{SE}}$ for workers with characteristics $Z_{\mathrm{SE}}$. Then, assuming a standard normally distributed unobservable part of the wage equation with cumulative distribution function $\Phi$, we can calculate the interquantile spread as $\Delta v(Z'_{\mathrm{SE}}\mu) = Q_(\alpha_2)(v|Z'_{\mathrm{SE}}\mu) - Q_(\alpha_1)(v|Z'_{\mathrm{SE}}\mu) = \Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)$. The ratio of these two spreads $\eta(Z_{\mathrm{SE}}, Z'_{\mathrm{SE}}\mu) = \Delta q(Z_{\mathrm{SE}})/\Delta v(Z'_{\mathrm{SE}}\mu)$ gives the factor by which we scale all variables in the wage equation.

[12] The selectivity-corrected results can be contrasted with those of Table D.7, which shows the uncorrected results estimated on the sample of self-employed individuals.

the signaling effect of licensing, which corrects for the asymmetric information of consumers. This type of explanation would require that consumers react to licensing differently from employers, who would not be induced to abandon statistical discrimination even in the presence of licensing. However, other explanations are possible (mainly based on selection into licensed professions and self-employment). These would generally require some sort of bias in the licensing process (or differential cost of licensing) that positively selects women simultaneously into licensed professions and self-employment. Exploring these possible mechanisms might provide new insights into the relation between regulations, institutions, and gender differences.

Another interesting fact that we document in this paper is that the absence of the wage gap for the licensed self-employed is not mirrored by a significant reduction in the gender gap in hours worked. This could be due to preferences of women for flexibility, or social and family constraints that limit their labor supply response. Overall, our work suggests that the interaction of occupational licensing regulations with labor supply decisions might be an interesting area for future research.

## References


ALTONJI, J. G., AND R. M. BLANK (1999): "Race and Gender in the Labor Market," vol. 3 of *Handbook of Labor Economics*, 3143 – 3259. Elsevier. Cited on page 4.

ARROW, K. J. (1973): *The Theory of Discrimination* 3–33. Princeton University Press. Cited on page 4.

BIEWEN, M., B. FITZENBERGER, AND M. SECKLER (2020): "Counterfactual Quantile Decompositions with Selection Correction Taking Into Account Huber/Melly (2015): An Application to the German Gender Wage gap," *Labour Economics*, forthcoming. Cited on pages 14, 15, and 16.

BLAIR, P. Q., AND B. W. CHUNG (2018): "Job Market Signaling through Occupational Licensing," Working Paper 24791, National Bureau of Economic Research. Cited on page 5.

BLAU, F. D., AND L. M. KAHN (2000): "Gender Differences in Pay," *Journal of Economic Perspectives*, 14(4), 75–99. Cited on page 1.

——— (2017): "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55(3), 789–865. Cited on pages 1 and 4.

BLINDER, A. S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*, 8(4), 436–455. Cited on page 8.

BOL, T., AND K. A. WEEDEN (2014): "Occupational Closure and Wage Inequality in Germany and the United Kingdom," *European Sociological Review*, 31(3), 354–369. Cited on page 3.

BORJAS, G. J., AND S. G. BRONARS (1989): "Consumer Discrimination and Self-Employment," *Journal of Political Economy*, 97(3), 581–605. Cited on page 5.

BRYSON, A., H. DALE-OLSEN, AND K. NERGAARD (2020): "Gender Differences in the Union Wage Premium? A Comparative Case Study," *European Journal of Industrial Relations*, 26(2), 173–190. Cited on page 8.

BUCHINSKY, M. (1998): "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach," *Journal of Applied Econometrics*, 13(1), 1–30. Cited on page 14.

——— (2001): "Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S.," *Empirical Economics*, 26, 87–113. Cited on page 14.

CARD, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19(1), 22–64. Cited on page 4.

CARPENTER, D. M., L. KNEPPER, K. SWEETLAND, AND J. MCDONALD (2018): "The Continuing Burden of Occupational Licensing in the United States," *Economic Affairs*, 38(3), 380–405. Cited on page 4.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78(3), 1093–1125. Cited on page 11.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205–2268. Cited on page 10.

FELSTEAD, A., D. GALLIE, F. GREEN, AND H.INANC (2014): "Skills and Employment Surveys Series Dataset, 1986, 1992, 1997, 2001, 2006 and 2012," Data collection, UK Data Service, SN: 7467. Cited on page 14.

GITTLEMAN, M., M. A. KLEE, AND M. M. KLEINER (2018): "Analyzing the Labor Market Outcomes of Occupational Licensing," *Industrial Relations: A Journal of Economy and Society*, 57(1), 57–100. Cited on page 4.

GITTLEMAN, M., AND M. M. KLEINER (2016): "Wage Effects of Unionization and Occupational Licensing Coverage in the United States," *ILR Review*, 69(1), 142–172. Cited on page 3.

GOLDIN, C. (2014a): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104(4), 1091–1119. Cited on page 1.

——— (2014b): *A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings* 313–348. University of Chicago Press. Cited on pages 1 and 4.

HABER, S. E., E. J. LAMAS, AND J. H. LICHTENSTEIN (1987): "On Their Own: The Self-Employed and Others in Private Business," *Monthly Labor Review*, 110(5), 17–23. Cited on page 5.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161. Cited on pages 14 and 16.

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. Cited on page 16.

HUNDLEY, G. (2000): "Male/Female Earnings Differences in Self-Employment: The Effects of Marriage, Children, and the Household Division of Labor," *ILR Review*, 54(1), 95–114. Cited on page 5.

JESSEN, R., D. ROSTAM-AFSCHAR, AND S. SCHMITZ (2018): "How Important is Precautionary Labour Supply?," *Oxford Economic Papers*, 70(3), 868–891. Cited on page 13.

KLEINER, M. M. (2000): "Occupational Licensing," *Journal of Economic Perspectives*, 14(4), 189–202. Cited on page 8.

KLEINER, M. M., AND A. B. KRUEGER (2010): "The Prevalence and Effects of Occupational Licensing," *British Journal of Industrial Relations*, 48(4), 676–687. Cited on pages 2, 3, and 4.

——— (2013): "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market," *Journal of Labor Economics*, 31(S1), S173–S202. Cited on pages 2, 3, and 8.

KLEINER, M. M., AND E. VOROTNIKOV (2017): "Analyzing Occupational Licensing Among the States," *Journal of Regulatory Economics*, 52(2), 132–158. Cited on pages 2 and 3.

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50. Cited on page 10.

KOUMENTA, M., A. HUMPHRIS, M. KLEINER, AND M. PAGLIERO (2014): "Occupational Regulation in the EU and UK: Prevalence and Labour Market Impacts," *Final Report, Department for Business, Innovation and Skills, School of Business and Management, Queen Mary University of London, London*. Cited on pages 4 and 14.

KOUMENTA, M., AND M. PAGLIERO (2019): "Occupational Regulation in the European Union: Coverage and Wage Effects," *British Journal of Industrial Relations*, 57(4), 818–849. Cited on pages 1, 3, 4, 6, and 8.

LAW, M. T., AND M. S. MARKS (2009): "Effects of Occupational Licensing Laws on Minorities: Evidence from the Progressive Era," *The Journal of Law and Economics*, 52(2), 351–366. Cited on page 4.

MACHIN, S. (1999): "Wage Inequality in the 1970s, 1980s and 1990s," in *The State of Working Britain*, ed. by P. Gregg, and J. Wadsworth. Manchester University Press. Cited on page 5.

METCALF, D., K. HANSEN, AND A. CHARLWOOD (2001): "Unions and the Sword of Justice: Unions and Pay Systems, Pay Inequality, Pay Discrimination and Low Pay," *National Institute Economic Review*, 176(1), 61–75. Cited on page 5.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302. Cited on page 7.

MINCER, J., AND S. POLACHEK (1974): "Family Investments in Human Capital: Earnings of Women," *Journal of Political Economy*, 82(2), S76–S108. Cited on page 4.

MOORE, R. L. (1983): "Employer Discrimination: Evidence From Self-Employed Workers," *The Review of Economics and Statistics*, 65(3), 496–501. Cited on page 5.

NEUMARK, D. (1988): "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination," *The Journal of Human Resources*, 23(3), 279–295. Cited on page 9.

NEWEY, W. K. (2009): "Two-Step Series Estimation of Sample Selection Models," *Econometrics Journal*, 12(s1), S217–S229. Cited on page 16.

OAXACA, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14(3), 693–709. Cited on page 8.

OLIVETTI, C., AND B. PETRONGOLO (2008): "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps," *Journal of Labor Economics*, 26(4), 621–654. Cited on pages 1 and 8.

PAGLIERO, M. (2019): "Occupational Licensing in the EU: Protecting Consumers or Limiting Competition?," *Review of Industrial Organization*, 55(1), 137–153. Cited on page 13.

PHELPS, E. S. (1972): "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 62(4), 659–661. Cited on page 4.

POLACHEK, S. W. (1981): "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure," *Review of Economics and Statistics*, 63(1), 60–69. Cited on page 4.

ROSTAM-AFSCHAR, D. (2014): "Entry Regulation and Entrepreneurship: A Natural Experiment in German Craftsmanship," *Empirical Economics*, 47(3), 1067–1101. Cited on pages 5 and 15.

——— (2015): "Regulatory Effects of the Amendment to the HwO in 2004 in German Craftsmanship," *Research Report, Directorate General Internal Market and Services, European Commission*. Cited on page 13.

ROSTAM-AFSCHAR, D., AND K. STROHMAIER (2019): "Does Regulation Trade Off Quality against Inequality? The Case of German Architects and Construction Engineers," *British Journal of Industrial Relations*, 57(4), 870–893. Cited on page 8.

SANDERSON, E., AND F. WINDMEIJER (2016): "A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics*, 190(2), 212 – 221, Endogeneity Problems in Econometrics. Cited on page 15.

SEXTON, E. A., AND P. B. ROBINSON (1989): "The Economic and Demographic Determinants of Self-Employment," in *Frontiers of Entrepreneurship Research*, ed. by R. H. Brockhaus, Handbook of Labor Economics, 28–42. Elsevier, Wellesley, Mass.: Centre for Entrepreneurial Studies, Babson College. Cited on page 5.

WAGNER, G. G., J. R. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements," *Journal of Applied Social Science Studies*, 127(1), 139–169. Cited on page 13.

WEICHSELBAUMER, D., AND R. WINTER-EBMER (2005): "A Meta-Analysis of the International Gender Wage Gap," *Journal of Economic Surveys*, 19(3), 479–511. Cited on pages 2 and 8.

## Tables

TABLE 1
Summary statistics

| | Females | | Males | | | |
|---|---|---|---|---|---|---|
| | Mean | Std. dev. | Mean | Std. dev. | Gap | t-stat. |
| Hourly wage (Euro) | 9.82 | 8.65 | 11.27 | 8.58 | -1.45*** | -6.28 |
| Monthly hours (h) | 144.41 | 45.07 | 167.77 | 33.53 | -23.36*** | -24.1 |
| *%-Share of* | | | | | | |
| Self-Employed | 0.11 | | 0.18 | | -0.07*** | -9.08 |
| Employee | 0.89 | | 0.82 | | 0.07*** | 9.08 |
| Licensed | 0.21 | | 0.22 | | -0.00 | -0.42 |
| Certified | 0.19 | | 0.22 | | -0.03*** | -3.21 |
| Unionised | 0.25 | | 0.23 | | 0.02** | 2.01 |
| Observations | | 11,981 | | 11,032 | | |

*Notes:* The table reports averages and standard deviations for the main variables used in the analysis by gender. The last two columns report results of t-tests on the gender differences between means. Figures are weighted by survey weights provided by the EU-SOR. Individuals with missing information on any of the variables reported in the table are excluded. Definitions of licensed, certified, certified, and unionized are described in Table A.1 and Appendix E. Significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$
*Source:* Own calculations based on the EU-SOR 2015.

TABLE 2
Oaxaca-Blinder decomposition of the gender wage gap

| | I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|---|
| Predicted log wage of males | 1.963*** | | 1.963*** | | 1.963*** | | 1.963*** | |
| | (0.011) | | (0.011) | | (0.011) | | (0.011) | |
| Predicted log wage of females | 1.797*** | | 1.797*** | | 1.797*** | | 1.797*** | |
| | (0.010) | | (0.010) | | (0.010) | | (0.010) | |
| Difference | 0.166*** | | 0.166*** | | 0.166*** | | 0.166*** | |
| | (0.015) | | (0.015) | | (0.015) | | (0.015) | |
| | Expl. | Unexpl. | Expl. | Unexpl. | Expl. | Unexpl. | Expl. | Unexpl. |
| Licensed | -0.002** | -0.033*** | -0.003** | -0.037*** | 0.001** | 0.001 | 0.002*** | -0.000 |
| | (0.001) | (0.006) | (0.001) | (0.007) | (0.000) | (0.003) | (0.000) | (0.003) |
| Licensed × Self-Employed | | | 0.002* | 0.001 | 0.001 | -0.005** | 0.001 | -0.005* |
| | | | (0.001) | (0.003) | (0.001) | (0.002) | (0.001) | (0.002) |
| Certified | | | | | | | 0.001* | -0.002 |
| | | | | | | | (0.000) | (0.003) |
| Certified × Self-Employed | | | | | | | -0.000 | 0.002 |
| | | | | | | | (0.001) | (0.002) |
| Unionisation | | | | | | | -0.001** | -0.006 |
| | | | | | | | (0.000) | (0.004) |
| Unionisation × Self-Employed | | | | | | | -0.001 | -0.000 |
| | | | | | | | (0.001) | (0.001) |
| Self-Employed | | | 0.004 | 0.010 | -0.007*** | 0.008 | -0.005* | 0.004 |
| | | | (0.002) | (0.007) | (0.002) | (0.005) | (0.002) | (0.006) |
| Individual controls | | | | | -0.016*** | 0.074 | -0.016*** | 0.090 |
| | | | | | (0.003) | (0.112) | (0.003) | (0.113) |
| Occupation FE | | | | | -0.010** | 0.014 | -0.011** | 0.013 |
| | | | | | (0.004) | (0.026) | (0.004) | (0.025) |
| Country FE | | | | | 0.024* | -0.053 | 0.024* | -0.053 |
| | | | | | (0.011) | (0.044) | (0.011) | (0.044) |
| Industry FE | | | | | 0.013*** | 0.023 | 0.014*** | 0.021 |
| | | | | | (0.003) | (0.058) | (0.003) | (0.058) |
| Constant | | 0.201*** | | 0.188*** | | 0.099 | | 0.096 |
| | | (0.016) | | (0.017) | | (0.136) | | (0.136) |
| Total | -0.002** | 0.168*** | 0.003 | 0.163*** | 0.006 | 0.160*** | 0.007 | 0.159*** |
| | (0.001) | (0.015) | (0.002) | (0.015) | (0.013) | (0.008) | (0.013) | (0.008) |
| Observations | 13,734 | | 13,734 | | 13,734 | | 13,734 | |

*Estimation:* OLS Oaxaca-Blinder decomposition.
*Sample:* Full sample. Decomposition is expressed from the viewpoint of females in contrast to males.
*Control variables:* Self-employment dummy, labour market institution dummies and their interaction with self-employment dummy, individual controls (education, age, age$^2$, firm size dummies), indicators for occupation, country, and industry. The table reports the sum of individual contributions for four groups of variables (individual characteristics, occupation fixed effects, country fixed effects, and industry fixed effects).
*Inference:* Robust standard errors obtained with the Delta method are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

TABLE 3
Quantile regressions of log hourly wages

| | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|
| Quantile | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| Dependent variable | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage |
| Female | -0.110*** | -0.131*** | -0.141*** | -0.147*** | -0.158*** | -0.171*** | -0.163*** | -0.166*** | -0.193*** |
| | (0.019) | (0.013) | (0.012) | (0.012) | (0.012) | (0.012) | (0.013) | (0.016) | (0.022) |
| Licensing | 0.035 | 0.037* | 0.047** | 0.044** | 0.079*** | 0.070*** | 0.076*** | 0.072*** | 0.070** |
| | (0.027) | (0.019) | (0.018) | (0.017) | (0.018) | (0.018) | (0.020) | (0.024) | (0.033) |
| Female × Licensing | 0.009 | 0.022 | 0.014 | 0.008 | -0.041* | -0.034 | -0.032 | -0.032 | 0.004 |
| | (0.038) | (0.026) | (0.025) | (0.024) | (0.025) | (0.025) | (0.028) | (0.033) | (0.046) |
| Certification | 0.032 | 0.040** | 0.039** | 0.039** | 0.040** | 0.047** | 0.064*** | 0.061** | 0.055* |
| | (0.028) | (0.019) | (0.018) | (0.017) | (0.018) | (0.018) | (0.020) | (0.024) | (0.033) |
| Female × Certification | -0.025 | -0.022 | -0.005 | -0.012 | 0.001 | -0.011 | -0.038 | -0.047 | 0.005 |
| | (0.038) | (0.026) | (0.025) | (0.024) | (0.025) | (0.025) | (0.028) | (0.033) | (0.045) |
| Unionisation | 0.029 | 0.024 | 0.028* | 0.013 | 0.002 | -0.011 | -0.001 | -0.015 | -0.040 |
| | (0.025) | (0.018) | (0.017) | (0.016) | (0.017) | (0.017) | (0.018) | (0.022) | (0.030) |
| Female × Unionisation | 0.023 | 0.040* | 0.039* | 0.038* | 0.051** | 0.057*** | 0.049** | 0.055** | 0.072* |
| | (0.031) | (0.022) | (0.021) | (0.020) | (0.020) | (0.021) | (0.023) | (0.027) | (0.038) |
| Self-Employment | -0.072 | -0.036 | 0.056 | 0.088** | 0.109*** | 0.176*** | 0.213*** | 0.353*** | 0.357*** |
| | (0.057) | (0.040) | (0.038) | (0.036) | (0.037) | (0.038) | (0.042) | (0.050) | (0.068) |
| Female × Self-Employment | -0.265*** | -0.143*** | -0.128*** | -0.076** | -0.015 | 0.007 | -0.010 | -0.047 | -0.101* |
| | (0.050) | (0.035) | (0.034) | (0.032) | (0.033) | (0.034) | (0.037) | (0.044) | (0.060) |
| Self-Employment × Licensing | -0.063 | -0.017 | -0.029 | 0.005 | -0.035 | -0.059 | -0.061 | -0.109** | -0.022 |
| | (0.061) | (0.042) | (0.041) | (0.038) | (0.039) | (0.040) | (0.044) | (0.053) | (0.072) |
| Female × Licensing | 0.425*** | 0.203*** | 0.240*** | 0.227*** | 0.197*** | 0.152** | 0.205*** | 0.241*** | 0.091 |
| × Self-Employed | (0.098) | (0.068) | (0.066) | (0.062) | (0.064) | (0.065) | (0.071) | (0.085) | (0.117) |
| Self-Employment × Certification | -0.034 | -0.050 | 0.020 | 0.023 | 0.095** | 0.066 | 0.047 | -0.017 | -0.068 |
| | (0.064) | (0.044) | (0.043) | (0.041) | (0.042) | (0.043) | (0.047) | (0.056) | (0.077) |
| Female × Certification | 0.120 | -0.082 | -0.107 | -0.142** | -0.135* | -0.170** | -0.087 | -0.075 | -0.074 |
| × Self-Employed | (0.108) | (0.075) | (0.072) | (0.068) | (0.070) | (0.072) | (0.079) | (0.094) | (0.129) |
| Self-Employment × Union | -0.325*** | -0.261*** | -0.139** | -0.210*** | -0.152*** | -0.108* | -0.034 | -0.091 | -0.102 |
| | (0.090) | (0.062) | (0.060) | (0.057) | (0.058) | (0.060) | (0.065) | (0.078) | (0.108) |
| Female × Union | -0.062 | 0.173 | 0.070 | 0.143 | 0.074 | 0.023 | -0.144 | -0.082 | -0.138 |
| × Self-Employed | (0.159) | (0.110) | (0.106) | (0.101) | (0.103) | (0.106) | (0.116) | (0.138) | (0.190) |
| Constant | 0.701*** | 0.996*** | 1.349*** | 1.469*** | 1.608*** | 1.768*** | 1.921*** | 2.110*** | 2.354*** |
| | (0.097) | (0.067) | (0.065) | (0.061) | (0.063) | (0.065) | (0.070) | (0.084) | (0.116) |
| Individual controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Occupation FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Country FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Industry FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 |

*Estimation:* Quantile regressions.
*Sample:* Full sample.
*Dependent variable:* Log net hourly wage.
*Individual Controls:* Education dummies, age, its square, firm size dummies.
*Inference:* Standard errors are in parentheses, significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

TABLE 4
Comparison of wage and hour regressions for EU, UK and German samples

| Dependent variable | Log Wages | | | Hours | | |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI |
| Sample | EU-SOR | UK-SES | GER-SOEP | EU-SOR | UK-SES | GER-SOEP |
| Female | -0.164*** | -0.232*** | -0.209*** | -14.017*** | -10.359*** | -39.708*** |
| | (0.015) | (0.017) | (0.010) | (2.000) | (0.400) | (1.985) |
| Female × Self-Employed × Licensed | 0.228*** | 0.259 | 0.265*** | -1.870 | -6.203 | -0.495 |
| | (0.069) | (0.209) | (0.067) | (6.290) | (3.978) | (4.214) |

*Estimation:* OLS regressions. Columns 1 and and 3 (EU-SOR) report results from Table B.2 and B.3 for comparison. Table B.4 reports the full table of results for the UK and German samples.
*Dependent variable:* Log net hourly wage or hours worked (monthly).
*Control variables:* Self-employment dummy, labour market institution dummies (except certification) and their interaction with self-employment dummies, indicators for kids of ages 0-4, 5-15, education dummies, age, its square, firm size dummies, indicators for occupation, country, and industry.
*Inference:* Standard errors (clustered by federal state in the German sample and by region and survey wave in the UK) are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on data for the UK from the Skills and Employment Survey (UK-SES), for Germany from the German Socio-Economic Panel (GER-SOEP), and the EU-SOR.

## Figures

FIGURE 1

Log Wage Distributions by gender and licensing status


(a) Non-licensed females


(b) Non-licensed males


(c) Licensed females

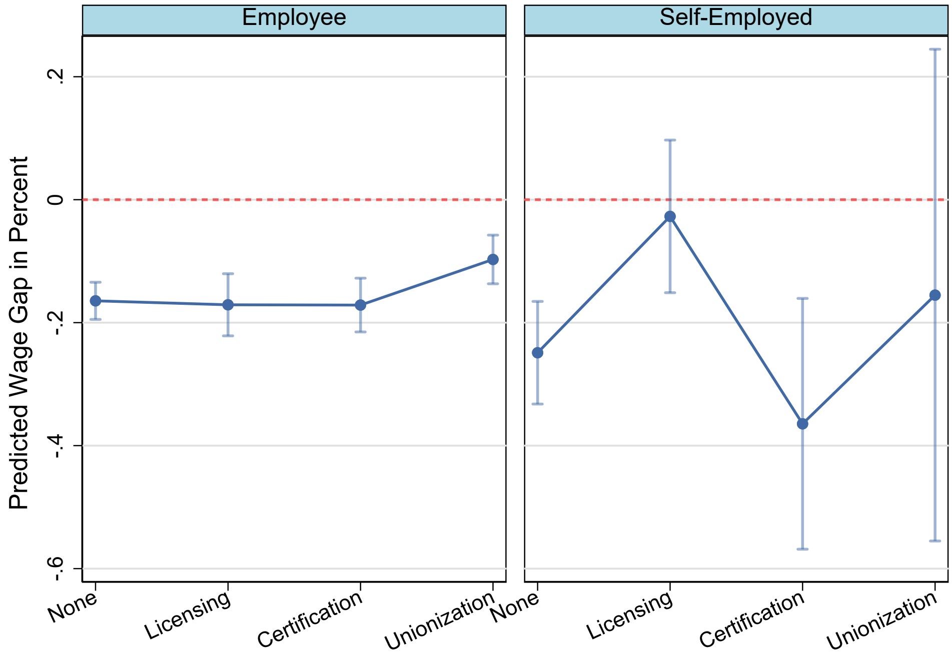

(d) Licensed males

*Notes:* The figure shows weighted kernel density estimates of log hourly wages for the self-employed (solid) and employees (dashed) by gender and licensing status. Estimates have been obtained using the Gaussian kernel function. Predicted mean wage for the self-employed (solid) and employees (dashed) and 95% confidence intervals are plotted as vertical lines and shaded areas, respectively.

*Source:* Own calculations based on the EU-SOR 2015.

FIGURE 2

Gender wage gap and labor market institutions

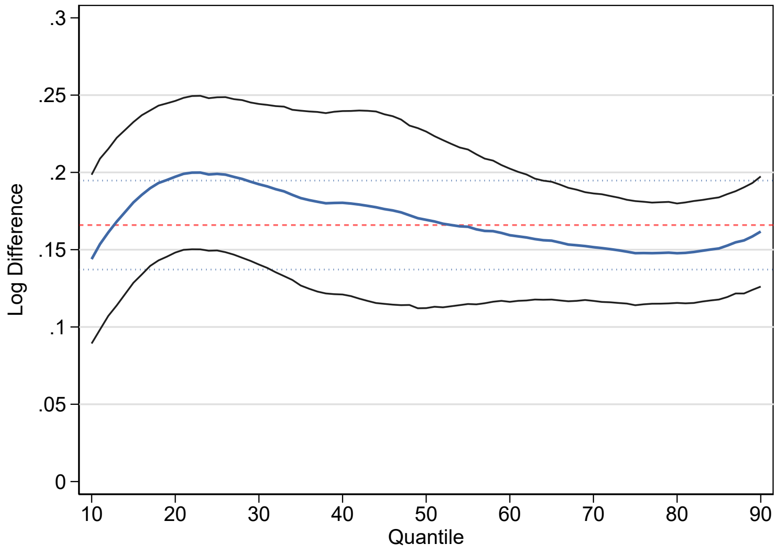



*Notes:* This figure is based on results from wage regressions (2). Estimation coefficients are reported in Table B.2, column IX. 95% confidence intervals are based on standard errors obtained with the Delta-method.

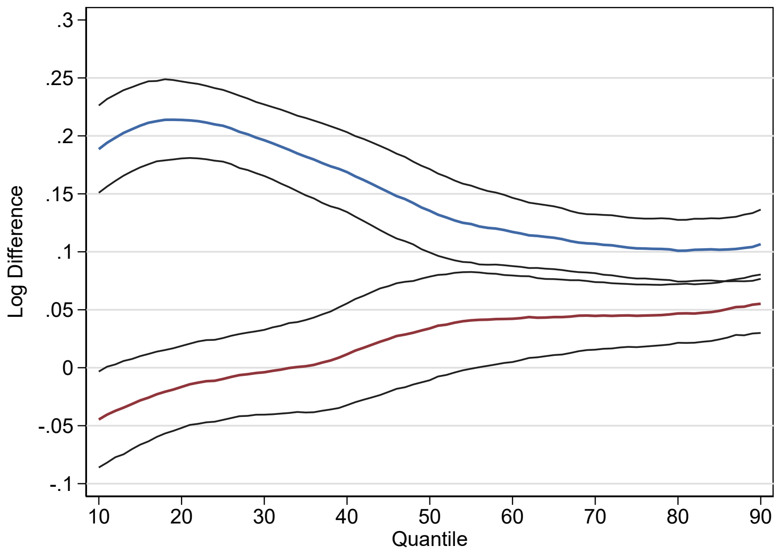
*Source:* Own calculations based on the EU-SOR 2015.

FIGURE 3

The gender wage gap at different quantiles of the wage distribution




(a) Total wage gap



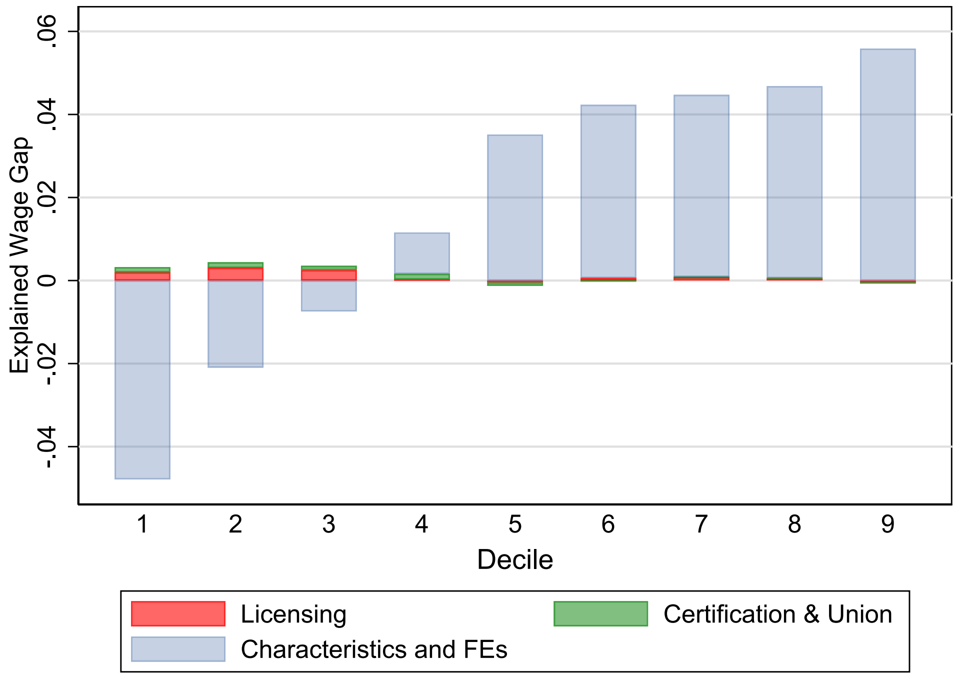

(b) Explained (bottom) and unexplained (top) component of wage gap

*Notes:* The figures are based on 81 quantile functions (with 100 quantile regressions each) from the 10th to the 90th percentile, with dependent variable log hourly wage, including control variables for employment status, labour market institution, their interaction with employment status, education age, its square, firm size, indicators for occupation, country, and industry. 95% confidence intervals are based on bootstrapped standard errors obtained with 100 replications. The dashed and dotted lines indicates the conditional expectation of the gender gap and its Eicker-Huber-White 95% confidence intervals.

*Source:* Own calculations based on the EU-SOR 2015.

FIGURE 4

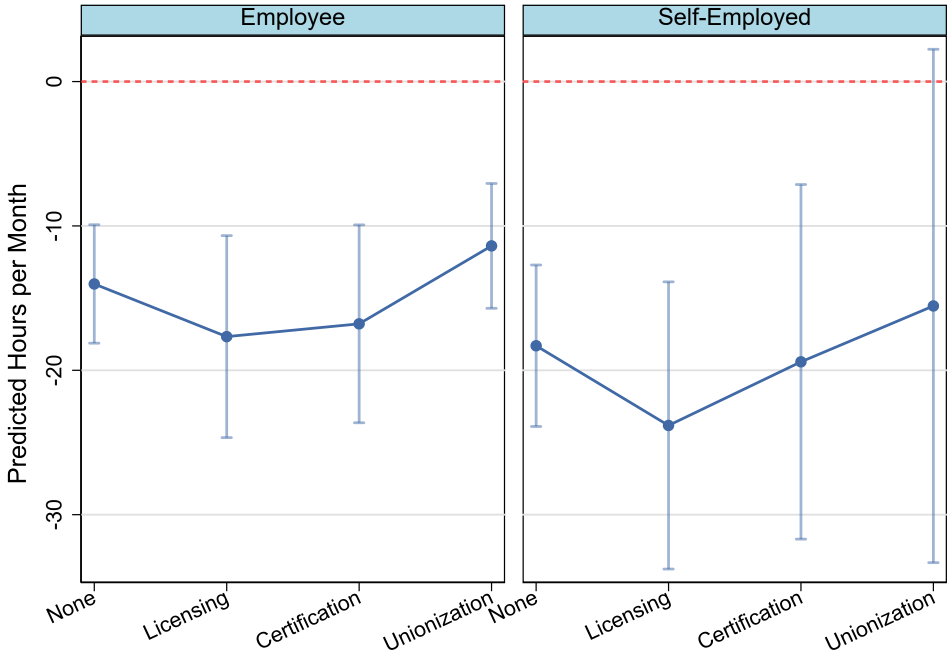
Detailed decomposition of the gender wage gap at different quantiles of the wage distribution




*Source:* Own calculations based on the EU-SOR 2015.

FIGURE 5

Gender hour gap and labor market institutions




*Notes:* The figure is based on results from hour regressions (4). Estimation coefficients are reported in Table B.3, column IX. 95% confidence intervals are based on standard errors obtained with the Delta-method.
*Source:* Own calculations based on the EU-SOR 2015.

## Appendix

## A Details on Key Variables in the EU-SOR

TABLE A.1
Definition of key variables

| Variable | Definition |
|---|---|
| *Female* | Indicator variable equal to one if the respondent reported being female. |
| *Licensing* | Indicator variable based on the question: 'In addition to this education, do you have a professional certification, license or did you have to take an exam which is required to practice your occupation?' and 'Without this professional certification, license or exam would you be legally allowed to practice your occupation?'. An individual is classified as 'licensed' if she answers 'Yes' to the first question and 'No' to the second. We exclude a small number of licensed workers who are in the process of obtaining their qualification and those indicating being both licensed and union members. |
| *Certification* | A worker is classified as 'certified' if she answers 'Yes' to the first question and 'Yes' or 'Don't know/No answer' to the second question. We exclude a small number of certified workers who are in the process of obtaining their qualification and those indicating being both certified and union members. |
| *Unionisation* | Indicator variable equal to one if respondent reports to be member of a trade union. |
| *Self-Employed* | Indicator variable equal to one if the respondent reported working as self-employed |
| *Education* | Set of indicator variables for six classes of highest level of respondent's education. Classes are primary education, lower secondary education, upper secondary education, post-secondary education, university, post-graduate/PhD. |
| *Age* | Respondent's reported age. |
| *Firm size* | Set of indicators for four classes of the number of full and part time workers at firm or organization where the respondent is working. Classes are $< 10, 10-15, 51-250, > 250$. |
| *Occupation* | Indicators for 1-digit ISCO classification. |
| *Industry* | Indicators for 1-digit NACE classification of the industry in which the firm or organization of the worker operates. |

# B Additional Results for the Conditional Expectations

TABLE B.2
Results from wage regressions

| | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|
| Dependent variable | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage |
| Female | -0.181*** | -0.177*** | -0.167*** | -0.160*** | -0.155*** | -0.179*** | -0.177*** | -0.169*** | -0.164*** |
| | (0.017) | (0.016) | (0.015) | (0.014) | (0.015) | (0.014) | (0.015) | (0.015) | (0.015) |
| Licensing | | | | 0.041** | | | 0.044** | 0.042** | 0.052*** |
| | | | | (0.016) | | | (0.018) | (0.019) | (0.018) |
| Female × Licensing | | | | 0.006 | | | 0.024 | 0.025 | -0.006 |
| | | | | (0.025) | | | (0.027) | (0.026) | (0.025) |
| Certification | | | | | 0.038** | | 0.042** | 0.043** | 0.040** |
| | | | | | (0.014) | | (0.018) | (0.018) | (0.018) |
| Female × Certification | | | | | -0.040* | | -0.017 | -0.018 | -0.007 |
| | | | | | (0.022) | | (0.025) | (0.025) | (0.022) |
| Union | | | | | | -0.037** | -0.021 | -0.016 | -0.006 |
| | | | | | | (0.016) | (0.018) | (0.018) | (0.017) |
| Female × Union | | | | | | 0.082*** | 0.082*** | 0.076*** | 0.067*** |
| | | | | | | (0.022) | (0.021) | (0.022) | (0.021) |
| Self-employed | | | | | | | | 0.156*** | 0.171*** |
| | | | | | | | | (0.049) | (0.051) |
| Female × Self-employed | | | | | | | | | -0.084** |
| | | | | | | | | | (0.039) |
| Self-employed × Licensing | | | | | | | | | -0.046 |
| | | | | | | | | | (0.058) |
| Female × Self-employed × Licensing | | | | | | | | | 0.228*** |
| | | | | | | | | | (0.069) |
| Self-employed × Certification | | | | | | | | | 0.016 |
| | | | | | | | | | (0.048) |
| Female × Self-employed × Certification | | | | | | | | | -0.109 |
| | | | | | | | | | (0.105) |
| Self-employed × Unionisation | | | | | | | | | -0.130 |
| | | | | | | | | | (0.081) |
| Female × Self-employed × Unionisation | | | | | | | | | 0.026 |
| | | | | | | | | | (0.189) |
| Individual controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Occupation FE | NO | YES | YES | YES | YES | YES | YES | YES | YES |
| Country FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Industry FE | NO | NO | YES | YES | YES | YES | YES | YES | YES |
| Observations | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 | 13,734 |
| R-squared | 0.737 | 0.754 | 0.758 | 0.761 | 0.760 | 0.761 | 0.761 | 0.762 | 0.762 |

*Estimation:* OLS regressions.
*Sample:* Full sample.
*Control variables:* Self-employment dummy, labour market institution dummies and their interaction with self-employment dummy, individual controls (education, age, age$^2$, firm size dummies), indicators for occupation, country, and industry.
*Inference:* Standard errors clustered by country are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

TABLE B.3
Results from hour regressions

| Dependent variable | I Hours | II Hours | III Hours | IV Hours | V Hours | VI Hours | VII Hours | VIII Hours | IX Hours |
|---|---|---|---|---|---|---|---|---|---|
| Female | -20.444*** | -17.853*** | -15.820*** | -14.569*** | -14.936*** | -16.148*** | -14.828*** | -14.021*** | -14.017*** |
| | (2.361) | (2.209) | (2.019) | (1.808) | (1.870) | (2.140) | (1.844) | (1.981) | (2.000) |
| Licensing | | | | 4.224*** | | | 5.211*** | 5.044*** | 4.618*** |
| | | | | (1.387) | | | (1.576) | (1.586) | (1.536) |
| Female × Licensing | | | | -4.581** | | | -4.182* | -4.100* | -3.650 |
| | | | | (1.970) | | | (2.084) | (2.114) | (2.243) |
| Certification | | | | | 2.365* | | 3.653** | 3.618** | 4.194** |
| | | | | | (1.177) | | (1.394) | (1.392) | (1.582) |
| Female × Certification | | | | | -2.201 | | -2.292 | -2.304 | -2.762 |
| | | | | | (2.384) | | (2.596) | (2.593) | (2.583) |
| Union | | | | | | 0.011 | 1.693 | 2.153* | 2.134* |
| | | | | | | (1.102) | (1.014) | (1.124) | (1.113) |
| Female × Union | | | | | | 4.546** | 3.230 | 2.617 | 2.636 |
| | | | | | | (2.175) | (1.909) | (1.986) | (1.942) |
| Self-employed | | | | | | | | 11.885*** | 11.980*** |
| | | | | | | | | (3.448) | (3.540) |
| Female × Self-employed | | | | | | | | -4.346* | -4.282 |
| | | | | | | | | (2.458) | (3.008) |
| Self-employed × Licensing | | | | | | | | | 1.678 |
| | | | | | | | | | (3.361) |
| Female × Self-employed × Licensing | | | | | | | | | -1.870 |
| | | | | | | | | | (6.290) |
| Self-employed × Certification | | | | | | | | | -2.577 |
| | | | | | | | | | (2.379) |
| Female × Self-employed × Certification | | | | | | | | | 1.650 |
| | | | | | | | | | (4.633) |
| Self-employed × Unionisation | | | | | | | | | 0.424 |
| | | | | | | | | | (3.236) |
| Female × Self-employed × Unionisation | | | | | | | | | 0.121 |
| | | | | | | | | | (8.621) |
| Individual controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Occupation FE | NO | YES | YES | YES | YES | YES | YES | YES | YES |
| Country FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Industry FE | NO | NO | YES | YES | YES | YES | YES | YES | YES |
| Observations | 19,688 | 19,688 | 19,688 | 19,688 | 19,688 | 19,688 | 19,688 | 19,688 | 19,688 |
| R-squared | 0.134 | 0.154 | 0.168 | 0.179 | 0.179 | 0.179 | 0.181 | 0.182 | 0.182 |

*Estimation:* OLS regressions.
*Dependent variable:* Hours worked (monthly).
*Sample:* Full sample.
*Control variables:* Self-employment dummy, labour market institution dummies and their interaction with self-employment dummy, individual controls (education, age, age$^2$, firm size dummies), indicators for occupation, country, and industry.
*Inference:* Standard errors assuming residuals are independent and identically distributed are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

TABLE B.4

Results from wage and hour regressions with UK and German data

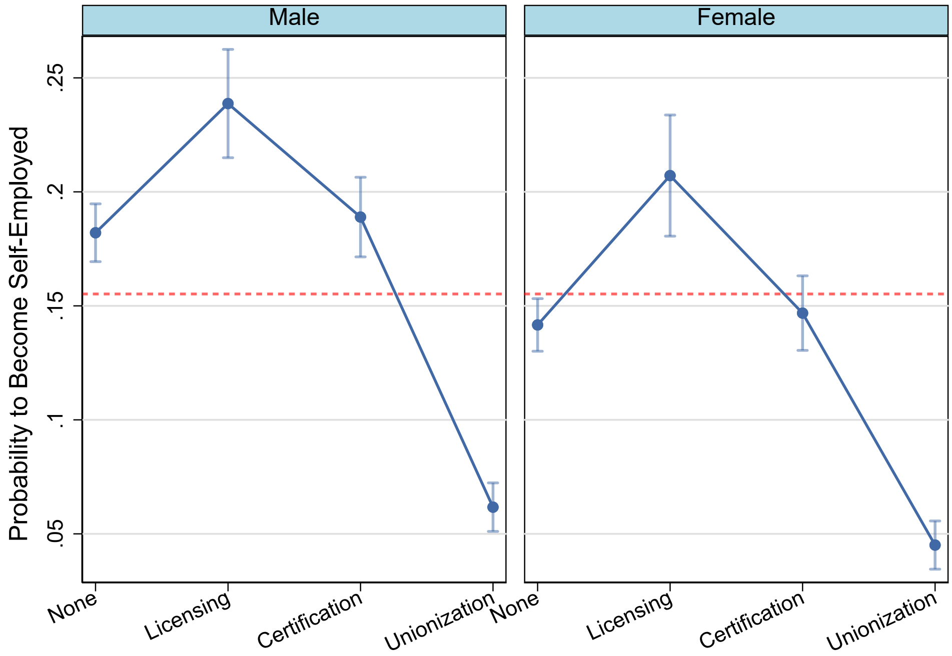
| | UK-SES | | GER-SOEP | |
|---|---|---|---|---|
| | Log Wages | Hours | Log Wages | Hours |
| Female | -0.232*** | -10.359*** | -0.209*** | -39.708*** |
| | (0.017) | (0.400) | (0.010) | (1.985) |
| Self-Employed | 0.071 | 3.235** | 0.166*** | 51.772*** |
| | (0.031) | (0.839) | (0.034) | (1.912) |
| Female × Self-Employed | -0.159** | -2.585* | -0.170*** | -14.539*** |
| | (0.057) | (1.351) | (0.042) | (3.973) |
| Licensed | 0.021 | 2.473 | 0.033*** | -0.487 |
| | (0.040) | (1.256) | (0.007) | (0.681) |
| Female × Licensed | 0.014*** | -0.015 | 0.038** | 5.358*** |
| | (0.042) | (1.593) | (0.013) | (0.900) |
| Self-Employed × Licensed | 0.514*** | 1.287 | 0.051 | -5.225** |
| | (0.112) | (2.116) | (0.033) | (2.233) |
| Female × Self-Employed × Licensed | 0.259 | -6.203 | 0.265*** | -0.495 |
| | (0.209) | (3.978) | (0.067) | (4.214) |
| Union | 0.028*** | -0.526 | 0.041*** | -6.329*** |
| | (0.013) | (0.320) | (0.008) | (0.470) |
| Female × Union | 0.248 | 2.860*** | -0.003 | 15.893*** |
| | (0.017) | (0.475) | (0.006) | (1.561) |
| Selfempl × Union | 0.052 | 1.529 | 0.042 | -33.498*** |
| | (0.112) | (2.346) | (0.123) | (10.976) |
| Female × Self-employed × Union | -0.036 | -0.340 | -0.354 | 20.000 |
| | (0.161) | (3.897) | (0.210) | (14.323) |
| Kids 0 to 4 years | 0.134*** | -1.489*** | 0.133*** | -8.383*** |
| | (0.013) | (0.357) | (0.006) | (0.807) |
| Kids 5 to 15 years | 0.055*** | -2.676*** | 0.049*** | -9.892*** |
| | (0.357) | (0.348) | (0.004) | (0.789) |
| Individual controls | YES | YES | YES | YES |
| Occupation controls | YES | YES | YES | YES |
| Country controls | YES | YES | YES | YES |
| Industry controls | YES | YES | YES | YES |
| Observations | 13,743 | 15,559 | 66,674 | 66,785 |
| R-squared | 0.423 | 0.253 | 0.515 | 0.299 |

*Estimation:* OLS regressions.
*Sample:* Full sample.
*Dependent variable:* Log net hourly wage or hours worked (monthly).
*Control variables:* Self-employment dummy, labour market institution dummies (except certification) and their interaction with self-employment dummies, indicators for kids of ages 0-4, 5-15, education dummies, age, its square, firm size dummies, indicators for occupation, country, and industry.
*Inference:* Standard errors (clustered by federal state in the German sample and by region and survey wave in the UK) are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on data for the UK from the Skills and Employment Survey (UK-SES) and for Germany from the German Socio-Economic Panel (GER-SOEP).

## C Female Participation in Self-Employment

FIGURE C.1

Female participation in self-employment




*Notes:* This figure is based on a probit regression with indicator for self-employment as dependent variable including control variables for education, age, its square, indicators for occupation, regional unemployment rate, and industry. 95% confidence intervals are based on standard errors obtained with the Delta-method. The red dashed line indicates the sample proportion of self-employed without controls.

*Source:* Own calculations based on the EU-SOR 2015.

TABLE C.5
Participation in self-employment

| | Probit | | OLS |
|---|---|---|---|
| | I<br>Coefficients | II<br>Marginal Effects | III<br>Coeff./Marginal Effects |
| Female | -0.203*** | -0.040*** | -0.046*** |
| | (0.036) | (0.007) | (0.007) |
| Licensing | 0.240*** | 0.057*** | 0.056*** |
| | (0.047) | (0.012) | (0.012) |
| Female × Licensing | 0.073 | 0.008 | 0.002 |
| | (0.069) | (0.015) | (0.015) |
| Certification | 0.031 | 0.007 | 0.009 |
| | (0.041) | (0.009) | (0.010) |
| Female × Certification | -0.003 | -0.002 | -0.007 |
| | (0.059) | (0.011) | (0.012) |
| Union | -0.780*** | -0.121*** | -0.126*** |
| | (0.063) | (0.009) | (0.009) |
| Female × Union | 0.010 | 0.023*** | 0.049*** |
| | (0.095) | (0.011) | (0.010) |
| Regional Unemployment | 0.020*** | 0.004*** | 0.004*** |
| | (0.005) | (0.001) | (0.001) |
| Individual controls | YES | YES | YES |
| Occupation controls | YES | YES | YES |
| Industry controls | YES | YES | YES |
| Observations | 19,985 | 19,985 | 19,985 |
| R-squared | | | 0.184 |

*Estimation:* Results from Probit and OLS regressions.
*Sample:* Full sample.
*Dependent variable:* Working as self-employed (binary).
*Control variables:* Labour market institution, their interaction with indicator for female, education, age, its square, indicators for occupation, industry, regional (NUTS1/NUTS2 level) unemployment rate.
*Inference:* Standard errors clustered at regional level (in column II obtained with Delta-method) are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

# D Additional Results on Quantile regressions

TABLE D.6
Quantile Wage Regressions (with selection correction)

| | I | II | III | IV | V | VI | VII | VIII | IX | X |
|---|---|---|---|---|---|---|---|---|---|---|
| Quantile | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | Heckit |
| Dependent variable | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage |
| Female | -0.339*** | -0.260*** | -0.221*** | -0.189*** | -0.148*** | -0.127*** | -0.097** | -0.136** | -0.110 | -0.223** |
| | (0.076) | (0.057) | (0.049) | (0.045) | (0.047) | (0.047) | (0.047) | (0.056) | (0.094) | (0.090) |
| Licensing | 0.001 | 0.070 | 0.069 | 0.018 | 0.048 | 0.022 | 0.026 | -0.030 | -0.055 | -0.058 |
| | (0.082) | (0.061) | (0.052) | (0.048) | (0.051) | (0.051) | (0.051) | (0.060) | (0.101) | (0.104) |
| Female × Licensing | 0.391*** | 0.254** | 0.132 | 0.131 | 0.060 | -0.010 | -0.015 | 0.172* | 0.167 | 0.199** |
| | (0.137) | (0.102) | (0.087) | (0.080) | (0.085) | (0.085) | (0.085) | (0.101) | (0.169) | (0.084) |
| Certification | 0.009 | 0.013 | 0.131** | 0.112** | 0.083 | 0.075 | 0.102* | 0.101 | 0.224** | 0.042 |
| | (0.085) | (0.063) | (0.054) | (0.050) | (0.053) | (0.053) | (0.053) | (0.062) | (0.105) | (0.050) |
| Female × Certification | 0.027 | 0.144 | -0.050 | -0.094 | -0.085 | -0.117 | -0.178* | -0.166 | -0.330* | -0.074 |
| | (0.148) | (0.110) | (0.095) | (0.087) | (0.092) | (0.092) | (0.092) | (0.109) | (0.183) | (0.123) |
| Unionisation | -0.067 | -0.041 | -0.161 | -0.127 | -0.226** | -0.128 | -0.018 | 0.040 | 0.052 | -0.006 |
| | (0.172) | (0.129) | (0.110) | (0.101) | (0.107) | (0.107) | (0.107) | (0.127) | (0.214) | (0.276) |
| Female × Unionisation | 0.062 | 0.216 | 0.255 | 0.224 | 0.219 | 0.109 | 0.075 | 0.030 | -0.070 | 0.148 |
| | (0.300) | (0.224) | (0.192) | (0.176) | (0.187) | (0.186) | (0.187) | (0.221) | (0.372) | (0.177) |
| Inverse Mills ratio | 1.843** | 0.992 | 0.490 | 0.385 | 0.588 | 0.787 | 0.407 | 0.200 | 0.024 | 0.083 |
| | (0.889) | (0.664) | (0.568) | (0.521) | (0.552) | (0.551) | (0.553) | (0.654) | (1.102) | (0.737) |
| Inverse Mills ratio Squared | -0.713 | -0.400 | -0.225 | -0.175 | -0.301 | -0.387 | -0.273 | -0.197 | -0.168 | -0.170 |
| | (0.594) | (0.443) | (0.380) | (0.348) | (0.369) | (0.368) | (0.369) | (0.437) | (0.736) | (0.312) |
| Inverse Mills ratio Cubic | 0.094 | 0.039 | 0.040 | 0.030 | 0.053 | 0.054 | 0.052 | 0.038 | 0.046 | 0.030 |
| | (0.130) | (0.097) | (0.083) | (0.076) | (0.081) | (0.080) | (0.081) | (0.096) | (0.161) | (0.065) |
| Selection Correction Terms | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Individual controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Occupation controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Country controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Industry controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 | 1,502 |

*Estimation:* Estimates of quantile regressions model with transformation, with selectivity correction. Heckit two-step estimation.
*Sample:* Self-employed sample.
*Dependent variable:* Log net hourly wage or hours worked (monthly).
*Control variables:* Employment status, labour market institution, their interaction with employment status, education, age, its square, firm size, indicators for occupation, country, and industry. Exclusion restriction is regional unemployment rate, see Table C.5.
*Inference:* Standard errors are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

TABLE D.7
Quantile Wage Regressions on Sample of Self-employed

| | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|
| Quantile | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| Dependent variable | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage | Log wage |
| Female | -0.292*** | -0.309*** | -0.258*** | -0.183*** | -0.161*** | -0.142*** | -0.118** | -0.143** | -0.240** |
| | (0.094) | (0.065) | (0.056) | (0.055) | (0.052) | (0.054) | (0.058) | (0.071) | (0.101) |
| Licensing | -0.012 | -0.080 | 0.001 | 0.001 | 0.033 | -0.012 | -0.007 | -0.056 | 0.012 |
| | (0.106) | (0.073) | (0.063) | (0.061) | (0.058) | (0.060) | (0.065) | (0.079) | (0.113) |
| Female × Licensing | 0.296* | 0.353*** | 0.186* | 0.199** | 0.120 | 0.062 | 0.045 | 0.181 | -0.008 |
| | (0.174) | (0.120) | (0.103) | (0.101) | (0.095) | (0.099) | (0.107) | (0.130) | (0.185) |
| Certification | 0.006 | 0.009 | 0.023 | 0.089 | 0.083 | 0.058 | 0.081 | 0.075 | 0.056 |
| | (0.112) | (0.077) | (0.067) | (0.065) | (0.062) | (0.064) | (0.069) | (0.084) | (0.119) |
| Female × Certification | 0.020 | 0.012 | 0.002 | -0.170 | -0.156 | -0.173 | -0.209* | -0.222 | -0.211 |
| | (0.193) | (0.134) | (0.115) | (0.112) | (0.106) | (0.110) | (0.119) | (0.145) | (0.206) |
| Unionisation | -0.148 | -0.108 | -0.098 | -0.073 | -0.026 | 0.006 | 0.042 | 0.052 | 0.013 |
| | (0.170) | (0.117) | (0.101) | (0.099) | (0.093) | (0.097) | (0.105) | (0.127) | (0.181) |
| Female × Unionisation | 0.141 | 0.225 | 0.210 | 0.178 | 0.080 | -0.120 | -0.142 | -0.162 | -0.104 |
| | (0.297) | (0.205) | (0.177) | (0.172) | (0.163) | (0.169) | (0.183) | (0.222) | (0.316) |
| Constant | 0.839 | 1.435*** | 1.776*** | 1.829*** | 2.006*** | 2.156*** | 2.465*** | 2.790*** | 3.612*** |
| | (0.516) | (0.357) | (0.307) | (0.300) | (0.284) | (0.293) | (0.318) | (0.387) | (0.550) |
| Selection Correction Terms | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Individual controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Occupation controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Country controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Industry controls | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Observations | 1,601 | 1,601 | 1,601 | 1,601 | 1,601 | 1,601 | 1,601 | 1,601 | 1,601 |

*Estimation:* Estimates of quantile regressions model.
*Sample:* Self-employed sample.
*Dependent variable:* Log net hourly wage or hours worked (monthly).
*Control variables:* Employment status, labour market institution, their interaction with employment status, education, age, its square, firm size, indicators for occupation, country, and industry.
*Inference:* Standard errors are in parentheses, significance levels are $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
*Source:* Own calculations based on the EU-SOR 2015.

## E Key Survey Questions

Two questions were asked after asking for the highest level of education in the EU-SOR to classify workers into two groups subject to (i) licensing and (ii) certification (or accreditation):

**Question:** *"In addition to this education, do you have a professional certification, licence or did you have to take an exam which is required to practice your occupation?"*
**Instructions to interviewees:** *"A professional certification or licence shows you are qualified to perform a specific job and may give you the right to enter a regulated profession or professional association."*
**Instructions to interviewers:** *"Only include certifications or licences obtained by the respondent as an individual. Examples include 'licensed medical doctor' and 'licensed taxi driver [...]"'.*

1. Yes
2. No – but currently in process of obtaining one
3. No
4. Don't know/No answer

To distinguish between licensing and certification, those who answer 1. or 2. to the above question were then asked:

**Question:** *"Without this professional certification, licence or exam would you be legally allowed to practice your occupation?"*
**Instructions to interviewers:** *"Refer to the respondent's specific occupation and personal circumstances. Refer to the current laws and regulations affecting the respondent's occupation (current main paid job).".*

1. Yes
2. No
3. Don't know/No answer

A worker is classified as 'licensed' if she answers 1. in the first question and 2. in the second. A worker is classified as 'certified' if she answers 1. in the first question and 1. or 3. in the second, and 'unregulated' otherwise.