GENDER PAY GAP IN THE CZECH REPUBLIC – ITS EVOLUTION AND MAIN DRIVERS*

Drahomíra Zajíčková[a], Miroslav Zajíček[a]

Abstract

The study estimates the size of the gender pay gap (GPG) for the Czech Republic in the years 2006–2017 using data from the EU-SILC survey. The size of the GPG (and the related variables) remains relatively time-invariant with a statistically weak relation to the business cycle. Using the Oaxaca-Blinder decomposition, we found out that the unexplained part of the GPG amounts to 50% of the whole GPG (on average) and only one third of the GPG is caused by an endowment effect or an interaction between the endowment effect and the coefficient effect. Selection bias plays a statistically insignificant role in terms of the GPG formation and explanation. Parenthood is the most important driver of the GPG. For parents, the GPG is about 30 percentage points higher than the one for non-parents. Women are able to narrow the GPG created by the effect of motherhood and reach original unexplained levels of approximately 15% after reaching the age of 50 and higher. Besides parenthood, there is no other demographic characteristic that has any substantial impact on the formation and persistence of the GPG. The GPG is most pronounced for the lowest- and the highest-earning quantiles, indicating the existence of a glass ceiling and a sticky floor on the Czech labour market.

Keywords: Gender, gender pay gap, parenthood, labour market, EU-SILC, quantile regression

JEL Classification: J24, J30, D10

* This article was supported by IGA PEF MENDELU project PEF_DP_2020010 entitled "Fatherhood premium: differences in renumeration between women and men".

This paper is based on data from Eurostat, EU Statistics on Income and Living Conditions 2006–2017. The responsibility for all conclusions drawn from the data lies entirely with the authors.

a Department of Law and Social Sciences, Faculty of Business and Economics, Mendel University in Brno, Czech Republic

Emails: drahomira.zajickova@mendelu.cz; zajicek100@gmail.com

1. Introduction

The gender pay gap (GPG), *i.e.*, the disparity in income between male and female workers, has been the focus of economic research for decades. It is also subject to legislative initiative or executive actions on the European as well as national levels[1]. There is an extensive body of literature on this topic seeking to measure and explain the GPG, its trajectory over time, its causes and also its purported remedies. The literature dealing with the Czech GPG is also extensive. However, as we will argue, it lacks cohesiveness and thoroughness.

This study aims to fill the gaps that we believe are present in studies dealing with the Czech data on the GPG. The study rests on four basic pillars: (a) using representative standardized national data covering the whole working population with a large number of respondents (EU-SILC) over a substantial period of time, thus being able to explore the development of the GPG over time, (b) monitoring the difference between gross GPG and the explained part of the GPG, (c) using the same method for every year and (d) a thorough decomposition of the resulting GPG.

Based on the aforementioned pillars, the paper offers a contribution to the empirical literature on the GPG in the Czech Republic in the following areas. Firstly, we provide a consistent estimate of various measures of the GPG for each year from the period 2006–2017 using the same set of data and the same estimation techniques and models. Secondly, we provide a proof that parenthood is the most important factor in explaining the existing gender pay gap by showing how the GPG evolves over the working life of men and women depending whether they are parents or not. Thirdly, we measure the GPG in the Czech Republic across various demographic groups. We also provide an evaluation of trends over time. Fourthly, we analyse the GPG and its development for deciles in order to investigate the existence of glass ceilings or sticky floors.

The paper is structured as follows. In the next section, we provide a short overview of the theoretical concepts related to the existence of the GPG and also an extensive overview of the Czech literature on the topic. In Section 3, we describe the data and the adjustments we made in order to use them. In Section 4, we introduce our estimation strategy and the following Section 5 provides the main results, Section 6 concludes.

2. Theoretical Concepts and Literature Overview

There are several sources of the existing GPG that can be summarized as follows: (1) economic explanation based on human capital or productivity factors such as education, skills,

1 The same applies to other areas such as the US, Canada and, in recent decades, practically any area in the world.

workforce experience, amount of work, managerial position, overtime exposition, hardship rent, *etc.* (Becker, 1965; Becker, 1981); (2) sociological explanation working mainly through industry or occupational segregation by self-selection as well as dislike among women of taking part in tournaments, accepting variable pay schemes, *etc.* (Gneezy *et al.*, 2003; Dohmen and Falk, 2010; Niederle and Vesterlund, 2007); (3) institutional explanations based on the differences in gender-specific flexibility constraints which can affect promotions and remuneration (Doeringer and Piore, 1971); and finally, (4) taste-based or statistical discrimination operating in hiring, promotion, task assignment, and/or compensation (Phelps, 1972; Donohue, 2007). Motherhood penalty is a part of such a differential discrimination story (Waldfogel, 1997; Zajíčková and Zajíček, 2020).

The exhaustive empirical literature tries to use, test and create new theoretical concepts. As the literature on the GPG is overwhelming, we will focus solely on papers dealing with Czech data to provide an overview of what has been done in relation to estimating the size of the GPG in the Czech Republic. As the other post-communist countries, former Czechoslovakia had a tradition of proclaimed equality between genders as well as a policy that all individuals had to work under the communist government. Wages were set centrally and were based on difficulty of jobs performed, their ideologically perceived "usefulness" (together with privileges for certain types of jobs such as miners, military personnel, high-ranking state administrative and party officials) and also achieved level of education. Although gender did not play a role in setting the wages, there was a substantial GPG under the communist system. For Czechoslovakia the estimates amount to 30% (Ham *et al.*, 1995). The end of communism brought the end of wage grids for non-public sector (thus a feasibility of potential discrimination for non-public industries), substantial wage dispersion and a sort of "antifeminism" promoting a return to traditional family roles of men and women. Several studies attempted to disentangle the impact of transition on the GPG (Večerník, 1995; Brainerd, 2000; Newell and Reily, 2001). All these studies typically use one (or very few) cross-sectional surveys of employees and do not control for any selection bias into employment or employment segregation. However, there is a general agreement on GPG somewhat narrowing as a result of labour reforms in the course of the transition in the Czech Republic with the magnitude of change around 5 percentage points of the gross GPG. According to some studies (Filer *et al.*, 1999), the transition converged to a relatively stable wage structure in the Czech Republic.

Jurajda (2003) and Jurajda (2005) used the ISPV[2] data to estimate GPG for public and non-public sectors in the Czech and Slovak Republics. He finds a strong evidence

2 ISPV (Informační systém o půměrném výdělku; Average Earnings Information System). ISPV is a national employer survey reporting hourly wages of their employees.

for a segregation effect on the size of GPG and also no substantial changes between 1998 and 2002 (the years for which he used data) with gross GPG amounting to around 30% with two thirds of this gap remaining unexplained. Křížková *et al.* (2008) used the same dataset to estimate gender pay gap for the years 1998, 2002 and 2004. The gross GPG amounted to 27% for 1998, 27% for 2002 and 25% for 2004. Once controlled for firm, type of employment and working position, the GPG narrows substantially to 12% (for all years).

However, there are also several limitations regarding the quality of ISPV data at a time[3]; thus, it became helpful that EU-SILC (European Union Statistics on Income and Living Conditions) surveys started to be conducted from 2005 in the Czech Republic. Mysíková (2007) applied the Heckman selection correction method and the Oaxaca-Blinder decomposition on the SILC 2005 data to decompose the gender wage differences. The study shows that the endowment effect for men and women in the Czech Republic is minimal, *i.e.*, the individual characteristics of working men and women are similar. The unexplained wage difference amounts to 21%. The same method is used in a comparative study of the Czech Republic, Slovakia, Hungary and Poland with the EU-SILC 2008 data (Mysíková, 2012). As for the Czech Republic, the unexplained GPG reached 25.6%, with lower values in the other Visegrad countries. The effect of the selection bias is negative in the Czech Republic but relatively modest. Balcar *et al.* (2012) conducted a representative survey of employees. Their task, however, was not to estimate of the size of GPG, it merely reported the gross wage difference in wages among surveyed men and women (22.2%).

The foreign paper worth mentioning is Christofides *et al.* (2013). They use the SILC 2007 data to compare gender pay gaps in the EU member states using the Oaxaca and Ransom methods and using Heckman's selection bias correction. For the Czech Republic, they find a wage difference of 27%, where 7.2% of the difference is explained by the differences in the human capital of men and women and 19.9% remains unexplained. Boll *et al.* (2016) used two different sources of data (EU SES[4] as well as EU-SILC) and the same method. When using EU SES from 2010 data, the Czech Republic shows an above-average

3 Data include industry and ownership type. Only firms over 10 employees were represented (firms with less than 10 employees became included in 2011 and once in four years only). Data on employees cover gender, education, age, *etc.* Top management is not included. Data on education are missing for 25% of workers so education information must be imputed from other sources when using ISPV. Moreover, composition of ISPV data is weighted towards large establishments and manufacturing industries. In order to use such data, they must be re-weighted in order to recover correct shares in population. There is also a questionable quality of data from firms with less than 100 employees.

4 EU SES = EU Structure of Earnings Survey

unadjusted wage gap of 16.5% (3.4% is explained and 13.1% is left unexplained). While using EU-SILC data from 2013, the unadjusted GPG increases to 27.13%, with 22.18% unexplained.

There are also several studies by Hedija dealing with the GPG, some of them using standardized data (mostly EU-SILC), some of them using data from individual companies (we shall not cover such case studies). Hedija (2014) estimated the size of the GPG using "average treatment effect on the treated" (ATT) estimation method on EU-SILC data for 2010. The size of unexplained GPG amounted to 19.5%, varying substantially depending on the NACE classification of the sector. Hedija (2017) examines whether the unexplained differences in the remuneration of women and men in different industries within the EU member states differ and identifies the possible causes of such differences. She uses the EU-SILC 2011 data. The analysis was conducted for 24 EU countries. For the Czech Republic, the ATT estimate for all sectors amounts to 23.9%. GPGs are also calculated for individual sectors. In the Czech Republic, coefficients range from 15.9% to 41.1%. Hedija (2018) uses the SILC data (2010–2012) to examine the sources of the unexplained GPG among 25 EU countries and to assess the impact of the legal environment of these countries, *i.e.*, whether the existing differences in the unexplained GPG in the EU countries can be explained by the differences in the quality of legislation and enforcement. The GPG in the Czech Republic amounts to 22.3%.

Křížková *et al.* (2018) is the only study that uses the same data set (ISPV) over a long period (2002–2016). The gross GPG proves to be very stable, oscillating for the whole period between 24 and 26%, with the unexplained part being about 2/3, *i.e.*, 15–16%. The study investigates the GPG in great detail; however, it does not take into account the one factor that affects the position of women on a labour market most crucially – the effect of parenthood on labour market outcomes (such as GPG), which is also an omission of all the studies described above. That omission stems from the fact that ISPV data do not contain information on parenthood (as opposed to the EU-SILC data used in this article). The only study taking into account the impact of parenthood on the GPG is Pytliková (2015). By using EU-SILC for 2012, she estimates the gross GPG at 38.4%, the unexplained part remaining at 27%, albeit with a limited number of controls.

Table 1 provides an overview of literature on the Czech GPG by various authors within the last 20 years.

Table 1: Overview of studies on GPG using Czech data

Study	Descrip-tion	Dataset used	Method	Sample	Gross GPG	Unexplained GPG
Jurajda (2003)	GPG Czech Republic and Slovakia	Trexima / ISPV 1998	Oaxaca-Ransom	Firms with more than 100 employees	24.1% for public sector 29.7% for non-public sector	9.2% for public sector 18.9% for non-public sector
Jurajda (2005)	GPG Czech Republic and Slovakia	Trexima / ISPV 2002	Oaxaca-Ransom	Firms with more than 100 employees	28.2% for non-public sector	16.5% for non-public sector
Mysíková (2007)	GPG Czech Republic	EU-SILC 2005	Heckman, Oaxaca decomp., Cotton	Employees 15–64 years	–	21%
Křížková *et al.* (2008)	GPG Czech Republic	Trexima / ISPV 1998, 2002, 2004	OLS	Firms with more than 10 employees	27% (1998) 27% (2002) 25% (2004)	12% (1998) 12% (2002) 12% (2004)
Balcar *et al.* (2012)	Descriptive statistic	Own rep-resentative survey	Sample average	1984 employees 25–54 years	22.2%	–
Mysíková (2012)	GPG V4 Countries	EU-SILC 2007	Heckman, Oaxaca decomp.	Employees	–	25.6%
Christofi-des *et al.* (2013)	GPG EU countries	EU-SILC 2007		Employees	26.6%	19.9%
Hedija (2014)	GPG Czech Republic	SILC 2010	ATT	Employees	–	19.5%
Pytliková (2015)	GPG controlled by number of children	SILC 2012	OLS	Employees, 20–49 years	38.4%	27% (only a limited number of controls)
Boll *et al.* (2016)	GPG EU countries	EU SES 2010, EU-SILC 2013	Heckman, Oaxaca decomp.	Employees	16.5% (SES) 27.3% (SILC)	13.1% (SES) 22.18% (SILC)
Hedija (2017)	GPG EU, inc. industries	SILC 2011	ATT	Full time employees	–	23.9%
Hedija (2018)	GPG EU countries	SILC 2010–2012	ATT	Full time employees	–	22.3%
Křížková *et al.* (2018)	GPG Czech Republic	Trexima / ISPV 2002–2016	OLS, Oaxaca decomp.	Firms with more than 10	24–26%	15–16%

Source: Balcar *et al.* (2012); Boll *et al.* (2016); Hedija (2014, 2017, 2018); Christofides *et al.* (2013); Jurajda (2003, 2005); Křížková *et al.* (2008, 2018); Mysíková (2007, 2012); Pytliková (2015)

To sum up, all the studies use relatively old data sets – 2013 as the newest ones – with the exception of Křížková (2018), who used ISPV data. All the studies (with the exception of Křížková, 2018, and Hedija, 2018) use only one year for the analysis, making the comparison quite complicated given the different methods employed. The methods and their application used to estimate the GPG differ. The only analysis of the impact of parenthood on the GPG so far has been Pytliková (2015), who used EU-SILC data only for one year (2012) and indirectly Jurajda (2003), providing estimates for age cohorts without any reference to parenthood. Other analysis and articles either ignore the role parenthood of the GPG or use data sets that do not include information on parenthood at all (ISPV).

In this paper, we try to overcome all the problems mentioned – we use the same data for each year from 2006 to 2017 and we employ the same methods to estimate the GPG in order not only to investigate the size of the GPG but also to capture the development of the GPG over time and we pay utmost attention to the role of parenthood in the formation of the GPG.

The GPG has been traditionally measured using an aggregate index, *i.e.*, around the mean of an income distribution. Much less is known about the GPG at the lower and upper ends of the income distribution. By analysing the GPG for different income quantiles, we can test the presence of a glass ceiling or sticky floors. The literature has identified the existence of a glass ceiling when the pay gap is significantly larger at the top of the distribution and a sticky floor when the wage gap is larger at the bottom. Knowledge of the GPG among households of various levels of income distribution is important if we are to address policies to mitigate the existing GPG – whether such policies are to be directed to the lower or upper part of the income ladder. In order to analyse differences among the different parts of the income distribution, we employ the unconditional quantile regression analysis.

3. Data

The data used in this study come from the EU SILC survey (European Union Statistics on Income and Living Conditions). The EU-SILC includes variables crucial to the analysis such as income and its various sources, education, employment rate of family members, data representing family members' earnings, family residence, and others. Some of these data have been collected for households collectively, some for individuals living in individual households. In particular, data on working experience, family members and children and size of enterprises are of utmost importance as, according to other studies, tenure (Boll and Leppin, 2015), employer size (Boll *et al.*, 2016) and in particular parenthood and other family ties (Light and Ureta, 1995) can explain a substantial part of the gender gap. As we shall show below, the gender gap widens with parenthood when a substantial part of family efforts (effort of mothers in the Czech environment) is devoted to childcare. The other advantage of the EU-SILC data is the inclusion of data for small enterprises (less than 10 employees), where one finds higher proportion of employed women and lower wages.

As the data used in this study are comparable for each particular year, it is possible to measure the changes of a variable over time. The data we used can be characterized as cross-sectional and containing a complete EC-SILC time series for the Czech Republic from 2006[5] to 2017 with over 8,000 households and 20,000 individuals included each year.

Only economically active persons − employees and self-employed persons − are considered in the calculations in order to capture the widest actual economic situation possible. In the case of women, we consider data for the unemployed ones and data for those on maternity leave in order to estimate the selection effects as well. Respondents discarded from the data file were those retired for the whole reference period, students, persons disabled by their health condition, homemakers and other economically inactive persons. We also limit our investigation to persons under 60 years of age in order to form a homogenous sample of a working population in which non-employment is a result of a "voluntary" decision on the labour market and to avoid complications stemming from retirement or premature retirement.

Appendix 1 contains the descriptive characteristics of the variables in the cross-sectional data files for the years 2006−2017.

4. Measuring GPG – Estimation Strategy

Empirical studies estimating the size of the GPG usually use various regression models including various control variables that, given the theoretical assumptions, should affect the pay regardless of gender. The variable used to measure the GPG is thus a dummy for a gender characteristic. Without any control variable, such an estimate produces a gross gender pay gap not taking into account the different characteristics of men and women on the labour market. As a rule, the models also take into account such quantities as age, education attained, work experience to represent the human capital, job type, employment type (full-time versus part-time), sector, job characteristics, industry branch, working conditions and modifications of such conditions. The outcome of such inclusion of various

5 The EU-SILC 2005 results for when the Czech Republic was first included in the surveys were discarded because, after the necessary modifications, the file did not contain a sufficient number of observations.

control variables thus provides a division of the gross GPG into the "explained" part and the "unexplained" part. The explained part is the one accounted for by the group differences in the productivity characteristics. The remaining part not accounted for by the group characteristics is thus "unexplained"[6]. Such a procedure is known in literature as an Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973). The analysis frequently accounts for a selection bias, most often following Heckman (1979). In our estimation strategy, we proceed with the following steps: we build three regression models in order to estimate the gross GPG, the GPG adjusted for various control variables and then we produce an estimate taking into account selection bias.

In order to estimate the GPG, we employ three versions of a regression model.

1. An estimate of the gross GPG without including any control variables.

$$\ln Y_i = \propto + \beta_0 SEX_i + \varepsilon_i \,, \tag{1}$$

where ln$Y$ refers to income in Czech crowns before taxes from primary employment, primary self-employment, secondary employment and secondary self-employment as reported by each respondent. In the EU-SILC enquiry, a respondent's income is given as the salary over the past 12 months. To calculate income per month from the salary, transformation is needed dividing it by the number of months spent in employment or self-employment using the methodology of Berger and Schaffner (2012). The regression models use the logarithm of gross income as this approach makes the interpretation easier. *SEX* is a dummy variable that equals 1 if the person is a woman and 0 if the person is a man. β_0 is a coefficient showing the gross GPG not taking into account any controlling factor. The economic interpretation of such a measure is an average difference in pay between men and women regardless of the source of such a difference.

2. An estimate of the GPG taking into account control variables.

$$\ln Y_i = \propto + \beta_X SEX_i + \beta X_i + \varepsilon_i \,, \tag{2}$$

where notation remains the same as in model (1). Moreover, there is a set of controls represented by the vector X_i. Control variables are explained in detail below. The coefficient β_X that provides an estimate of the GPG adjusted for the control variables.

3. An estimate of the GPG taking into account a correction of the selection bias in the case of women in order to account for a possible systemic difference between women active

6 The unexplained part has often been (albeit not completely accurately) used as a measure of discrimination on the labour market for various groups; however, it is not possible to use this measure this way without other qualifications (Del Rio *et al.*, 2011).

on the labour market and those choosing to stay out of it. Women with low expected income might to choose not to participate on the labour market (and have children), thus lowering the gap between men and women that appears in data (Olivetti and Petrongolo, 2008). Our selection model is based on the seminal work of Heckman (1979). The estimate is produced in a two-stage procedure with a Probit selection model using data for all the women in the sample and serving as the first step followed by the computation of the inverse mill ratio (*imr*) that serves as a missing variable in the OLS of a second stage taking into account only working women.

Regression equation:

$$\ln Y_i = \propto + \beta_X SEX_i + \beta X_i + \beta_{imr} imr_i + \varepsilon_i \ , \tag{3}$$

where imr_i stands for an inverse mill ratio coming from the selection equation and β_{imr} represents the impact of the self-selection of women on the size of the resulting GPG.

$$imr_i\left(V\gamma\right) = \varphi\left(V\gamma\right) / \phi\left(V\gamma\right), \tag{4}$$

where φ is the standard normal probability distribution function and Φ is the standard normal cumulative distribution function, V stands for a vector of the explanatory variables in the selection variables and γ stands for the estimated values of the regression coefficients from the selection equation.

The selection equation is a probit model using all N observations for women with $lpfW_i$ denoting the labour participation function ($lpfW$ = 1 for working women, $lpfW$ = 0 for non-working women) as a dependent variable and a vector V_i of the explanatory variables:

$$lpfW_i \ = \ \gamma V_i + v_i \ . \tag{5}$$

For all the regression models, the vector X consists of the following explanatory variables:

EDUHIGH and *EDUMIDDLE* are dummy variables denoting the highest educational level attained. We distinguish three levels of attained education − primary, secondary or tertiary. Primary education (*EDUBASE*) is the reference group[7] here. *ODPRAC_LET* stands for tenure and *ODPRAC_LET2* stands for its square. *PRAHA* and *STRCECHY*

7 Via inclusion of binary variables, dividing a described group into several subgroups can (if one of those subgroups is relatively less numerous) incur a high level of correlation among the remaining subgroups. It is in fact the situation of the categories *EDUHIGH* and *EDUMIDDLE*. The baseline is the variable *EDUBASE*, which is less numerous. It would be possible to dispose of such multicollinearity by simply choosing as the binary variables *EDUHIGH* and *EDUBASE* or alternatively *EDUMIDDLE* and *EDUBASE*. However, it would not have any impact on the regression model coefficients, statistical inference and power. Using *EDUHIGH* and *EDUMIDLLE* as dummies has an advantage of being more intuitive.

are location dummies equalling 1 if the person lives in Prague or in Central Bohemia, respectively, and 0 otherwise. Such a differentiation is based on the relative economic homogeneity of the rest of the Czech Republic. Thus, the other Czech regions serve as the reference group. *HUSBAND* and *WIFE* are dummy variables denoting whether the man or woman lives with a partner in one household (in that case, it equals 1) or not (0). *MUNIBIG* is a dummy variable that equals 1 if the person lives in a municipality with more than 100,000 inhabitants and 0 otherwise. *MUNIMIDDLE* equals 1 if the person lives in a municipality with more than 50,000 inhabitants and less than 100,000 and 0 otherwise. *SIZEBIG* and *SIZEMIDDLE* are binary variables corresponding to workplace size in terms of the number of employees. *SIZEBIG* equals to 1 if the number of employees exceeds 50, 0 otherwise. *SIZEMIDDLE* equals 1 if the number of employees ranges between 11 and 49, zero otherwise. The baseline is a workplace with fewer than 10 employees. *CONTRACT* is a dummy variable representing the fact that the employee has an unlimited contract, 0 otherwise. *SUPERVISOR* is a dummy that equals 1 if the employee is in a managerial position, 0 otherwise. *NACEI* is a set of dummies depicting the occupational groups, where *I* equals *A* through *U*. *NACEG* stands for the base as the most numerous gender-balanced group. The vector *V* consists of the following variables: *lnYOW* − logarithm of a woman's non-labour income; *PARTNER* and *PARTNERW* − dummy variables for having a partner or a working partner, while living without a partner equalizes this variable to 0; *AGE30W* and *AGE31_45W* − age dummies that equal 1 for women of the corresponding age, 0 otherwise; *CHILD2W*, *CHILD3_5W*, *CHILD6_15W* are dummy variables showing the presence of a child of the respective age in the household; *EDUHIGH* and *EDUMIDDLE* are the educational dummies defined above.

Appendix 2 contains the coefficients estimated for selection equations for each year from 2006 to 2017. We can see that all the regression coefficients have a sign corresponding to general economic intuition. A positive non-labour income decreases the probability of the woman being employed as well as having a partner, children younger than 5, and (less so) having children between 6 and 15. On the other hand, higher and middle education increases the probability of being active on the labour market. The age between 31 and 45 is statistically insignificant for most years as well as having a working partner.

In order to investigate the relation between the GPG and age, we also regressed models (1) and (3) separately for age groups up to 25 years, 26−30 years, 31−35 years, 36−40 years, 41−45 years, 46−50 years, 51−55 years and 56−60 years. Equations (1) and (3) were also adjusted in order to estimate the GPG for various demographic groups. The groups we analysed were:

- employees in the public sector versus others[8] (equations (1public), (1private), (3public) and (3private));
- different levels of education attained (equations (1edh), (1edm), (1edb), (3edh), (3edm) and (3edb));
- Prague citizens versus non-Prague dwellers (equations (1prg), (1nprg), (3prg) and (3nprg));
- parents versus non-parents (equations (1par), (1npar), (3par) and (3npar));
- employees in various sizes of workplaces (equations (1small), (1middle), (1big), (3small), (3middle) and (3big)); and
- employees in typically male versus female employment (equations (1typf), (1mixed), (1tym), (3typf), (3mixed) and (3tym)).

Table 2: Composition of vector *X* for various regression equations

Equation	X_i
(3public), (3private)	*HUSBAND WIFE PRAHA STRCECHY EDUHIGH EDUMIDDLE ODPRAC_LET ODPRAC_LET2 SIZEBIG SIZEMIDDLE MUNIBIG MUNIMIDDLE CONTRACT SUPERVISOR NACEI*
(3edh), (3edm), (3edb)	*HUSBAND WIFE PRAHA STRCECHY ODPRAC_LET ODPRAC_LET2 SIZEBIG SIZEMIDDLE MUNIBIG MUNIMIDDLE CONTRACT SUPERVISOR NACEI*
(3prg), (3nprg)	*HUSBAND WIFE EDUHIGH EDUMIDDLE ODPRAC_LET ODPRAC_LET2 SIZEBIG SIZEMIDDLE CONTRACT SUPERVISOR NACEI*
(3par), (3npar)	*HUSBAND WIFE PRAHA STRCECHY EDUHIGH EDUMIDDLE ODPRAC_LET SIZEBIG SIZEMIDDLE MUNIBIG MUNIMIDDLE CONTRACT SUPERVISOR NACEI*
(3smallmiddle), (3big)	*HUSBAND WIFE PRAHA STRCECHY EDUHIGH EDUMIDDLE ODPRAC_LET ODPRAC_LET2 MUNIBIG MUNIMIDDLE CONTRACT SUPERVISOR NACEI*
(3typf), (3mixed), (3typm)	*HUSBAND WIFE PRAHA STRCECHY EDUHIGH EDUMIDDLE ODPRAC_LET ODPRAC_LET2 SIZEBIG SIZEMIDDLE MUNIBIG MUNIMIDDLE CONTRACT SUPERVISOR NACEI*

Source: Authors‘ own coding

In the case of parents and non-parents, we limited our data set by age from above. The EU-SILC data only contain information on the numbers and ages of children in a household rather than on all the children born. For this reason, it is necessary to discard parents who no longer live with their children in one household but whose

8 Public employee is defined as an employee whose employment falls into sectors 84 and 85 in the CZ-NACE categorization.

income could be affected by their parenthood. In accordance with the literature, we chose an approach that defines the data set by a specific age range to eliminate older parents who, although having no children in a common home at present (regarded as "non-parents" in the EU-SILC data), but have had some. Therefore, the group of parents is truncated from above by an age limit of 40 years.

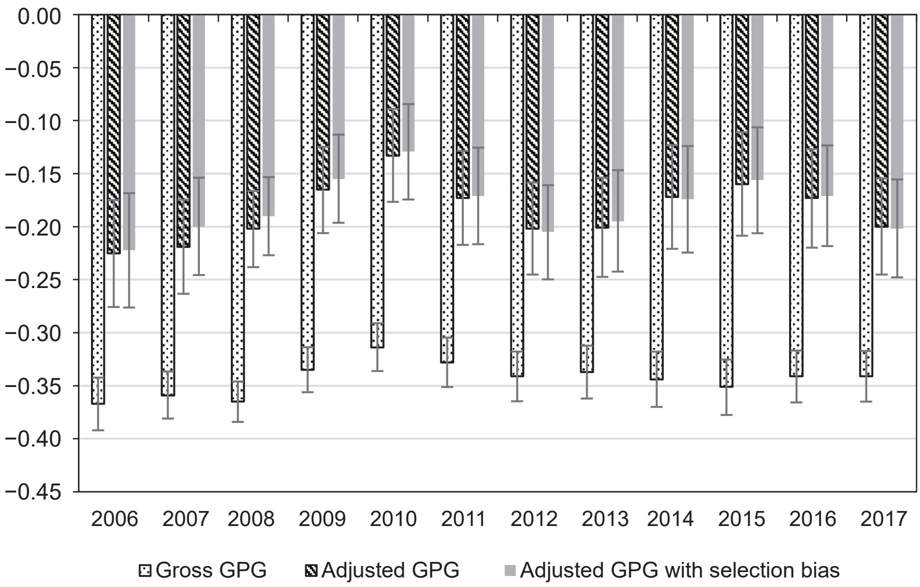
Control variables have to be adjusted accordingly for different regression equations. Table 2 gives an overview of the X_i composition for different regression equations.

In order to analyse income gaps across the different quantiles of the income distribution, we employ a conditional quantile regression to all our models (1) through (3).

5. Main Results

The estimates of the GPG based on models (1), (2) and (3) are shown in Figure 1.
The estimates of various types of GPG for all the years are all statistically significant[9].

Figure 1: Comparison of gross GPG, GPG with control variables and GPG controlling for selection bias



Source: Authors' own computation based on EU-SILC data

9 Detailed results are reported in Appendix 3 for the gross GPG, Appendix 4 for the GPG taking into account the control variables and Appendix 5 for the GPG accounting for a selection bias.

The gross GPG (model 1) amounts to approximately 35% in 2006–2009 (36.7%, 35.9% and 36.5%, respectively). It decreases a little in 2009, 2010 and 2011 (33.5%, 31.4%, 32.8%, in respective years) before returning almost to the original levels around 34%, with the peak value in 2015 (35.1%). The development of the gross GPG over time seems to be partially driven by the business cycle with GPG rising along with economic activity (regression coefficient when regressing GPG deviations from their mean on GDP changes over the years is weakly statistically significant at the 10% level of significance[10]). Apart from that, the level of the gross GPG remains stable over the whole period despite several important changes in legislation enacted in the meantime with the most significant one being the Antidiscrimination Act passed and enforced in 2009 (Act no. 198/2009 Coll., on equal treatment and legal protection against discrimination and on changes in some acts). The Antidiscrimination Act seems to have an impact on the size of the motherhood penalty (Zajíčková *et al.*, 2021). However, it does not seem to have an impact on the GPG as a whole. Another legal change is the shortening of the period in which women did not participate on the labour market as a result of the 2008 reform of the parental allowance, giving parents more leeway to choose the allowance drawing length. The average time after which the mother returned to work dropped from 40 months (3.3 years) to 34 months (2.8 years) after the birth of the youngest child (Pertold-Gebicka, 2018), which substantially increases the mothers' rates of pay. However, it does not seem to affect the overall GPG substantially.

There is a very similar GPG development over the time when including control variables in the regression equation (model 2) as well as taking into account a selection bias (model 3). With the exception of 2007 and 2008, the selection bias does not seem to constitute a problem as the *imr* is not statistically significant. That also means that regression coefficients representing the GPG do not substantially differ in models (2) and (3). Therefore, in what follows, we shall only refer to model (3) results as conclusions stemming from it also apply to model (2) outcomes. The model (3) results shall be called the "adjusted GPG".

5.1 Oaxaca-Blinder decomposition

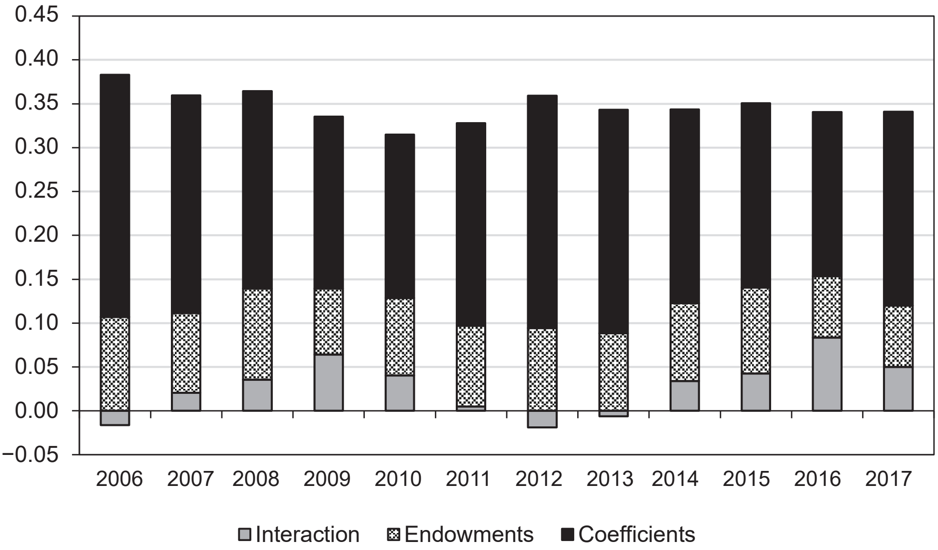
Although control variables can explain a substantial part of the gross GPG, the larger part of the GPG still remains unexplained. However, the unexplained part (for the Oaxaca-Blinder decomposition) varies significantly over time with the average value being 52.51%

10 Regression results when regressing GPG deviations from the mean on GDP changes.

VARIABLES	Gross GPG	SE	Constant	SE	Observations	R-squared
GDP growth	−0.245*	(0.128)	0.573	(0.496)	12	0.267

of the gross GPG, the minimum being 41.08% in 2010 and the maximum being 60.49% in 2006. The basic two-fold Oaxaca-Blinder decomposition (explained versus unexplained part) as well as the three-fold Oaxaca-Blinder decomposition are described in detail in Appendix 6. The first one shows the average increase in women's pay if they have the same characteristics as men (endowment effect). The second part shows the change in women's pay if the men's regression coefficients are applied to women (effect of coefficients). The third part is called the interaction term and provides the measure of a simultaneous effect of both − the effect of endowment and the effect of coefficients. A general overview of the decomposition of the gross GPG according to the endowment versus coefficient effects is shown in Figure 2.

Figure 2: Oaxaca-Blinder decomposition of gross GPG



Source: Authors' own computation based on EU-SILC data

If the characteristics of men and women were the same, the endowment effect would amount to 0. Alternatively, if there is some endowment effect, the labour market char-acteristics of men and women are not the same. It is clear from such a decomposition that the differences in terms of characteristics between the genders in the Czech Republic are relatively small and play a minor role in explaining the existence and persistence of the GPG. The contribution of different characteristics between genders explains around a third of the total wage difference on average if we include the interaction term as a part of it with peaks in 2009 (41.5%) and 2016 (45.0%) and lows in 2006 (24.75%) and 2012 (22.15%).

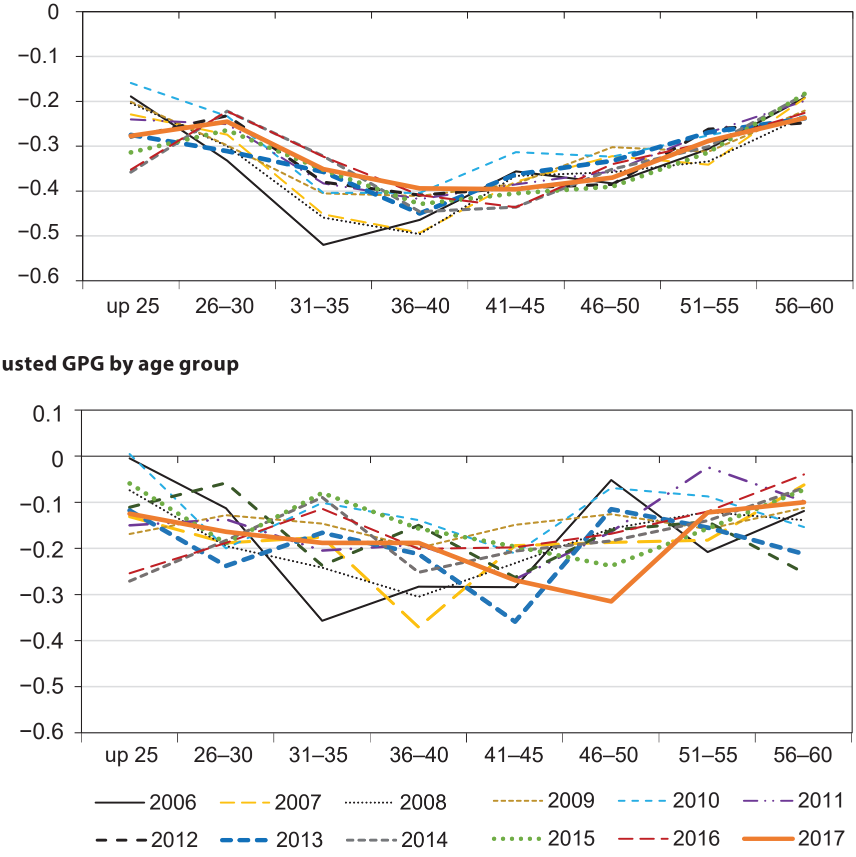
5.2 GPG for various age groups

In order to investigate the distribution of the GPG over the age group, we estimated models (1) and (3) for various age groups for each year. The results are shown in Figure 3.[11]

Figure 3: Gross GPG and adjusted GPG by age groups

Gross GPG by age group

Adjusted GPG by age group



Source: Authors' own computation based on EU-SILC data

The outcome of the regressions provides a set of U-shaped curves, indicating a strong impact of parenthood on the evolution of the GPG over the working life of men

11 As the number of the various regressions is very large, the detailed data can be provided by the authors on request and we only provide the general results here.

and women. The gap widens substantially for the age groups of 30−35 and 36−40 and then narrows gradually as women catch up. The picture is a little blurred for model (3). To obtain a clearer picture of the GPG age dependence, we reran the regressions estimating the gross GPG as well as the selection correction version of the GPG for all the years from 2006 to 2017 together with the dummies for 2007−2017 using the year 2006 as the base. The regression equations used in this part are the following:

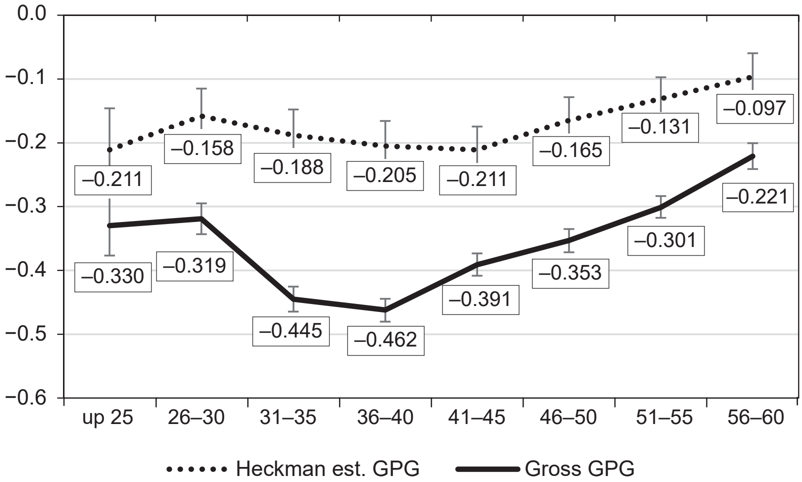
$$(1Y)\ \ln Y_i = \propto + \beta_X SEX_i + \beta_Y YEAR_t + \varepsilon_i \,,$$

$$(3Y)\ \ln Y_i = \propto + \beta_X SEX_i + \beta X_i + \beta_{imr} imr_i + \beta_Y YEAR_t + \varepsilon_i \,,$$

where β_Y is a vector of the regression coefficients which represents the impact of an individual year on income with respect to the base year 2006, and $YEAR_t$ is a vector of the dummy variables representing each respective year. The other variables are defined in the same way as above.

Figure 4 shows the estimates of the gross GPG for various age groups as well as estimates of the adjusted GPGs with their 95% confidence intervals.[12]

Figure 4: Gross GPG and adjusted GPG by age group – aggregate estimate



Source: Authors' own computation based on EU-SILC data

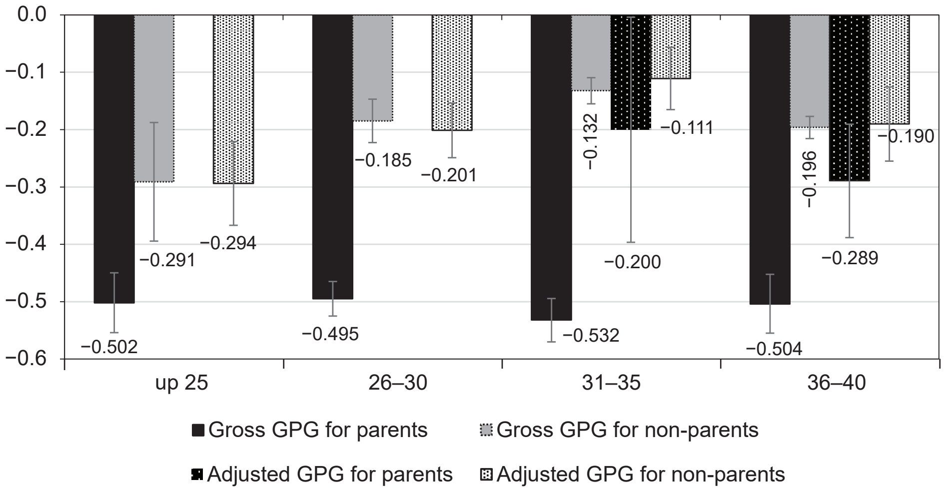
The overall pattern is clear: the GPG constitutes itself in the first two age groups up to 30 years at a level of 32% and then jumps up to levels around 45% for the age

12 Appendix 7 provides detailed regression results for the gross GPG (model 1Y). Detailed regression data for the adjusted GPG (3Y) are available from the authors on request as they are extensive.

from 31 to 40. From the age of 41 on, the GPG decreases gradually to reach the level around 30% after the age of 50 and then continues to close itself to levels just above 20% just before the retirement age. The same shape is expressed by the estimates of the GPG adjusted by control variables as well as for a selection bias, but it is much less manifested. It starts at around 20%. The upward jump is less pronounced than in the case of the gross GPG − it widens to 18.8% for the age group of 31−35, remains the same for the age groups of 36−40 and 41−45 (20.5% and 21.1%, respectively) and then narrows down to approximately 10% for the age group of 56−60. It almost seems that if the retirement age was postponed till 70, the GPG would disappear just before retirement.

In order to analyse the relationship between parenthood and the GPG, we estimated models (1Y) and (3Y) for parents and non-parents separately (*i.e.*, only for the age groups up to 40). In Figure 5, we report the respective regression results[13]. We report only statistically significant results with a 5% level of significance excluding data for the age groups up to 35 for parents in model (3Y).

Figure 5: Gross GPG and adjusted GPG by age group – aggregate estimates for parents and non-parents only (for age cohorts up to 40)



Source: Authors' own computation based on EU-SILC data

The outcomes provide an illuminating picture. For parents, the gross GPG for all the age groups amounts to approximately 50%. The same measure for non-parents is about 30 percentage points lower on average. Moreover, the gross GPG for non-parents does not

13 Detailed regression data are available from the authors on request due to their extensivity.

differ significantly from the adjusted GPG for non-parents. The only statistically significant estimates of the adjusted GPG for parents (age groups of 31−40) also shows a substantial difference from the one for non-parents, although much smaller than in the case of the gross GPG. All the data point to the fact that parenthood drives the largest part of the GPG.

5.3 GPG heterogeneity analysis

Figure 6 shows the resulting estimates of the gross GPG and the adjusted GPG for different demographic groups[14]: models (1edh), (1edm), (1edb), (3edh), (3edm) and (3edb); (1prg), (1nprg), (3prg) and (3nprg); (1public), (1private), (3public) and (3private); (1smallmiddle), (1big), (3smallmiddle) and (3big); (3typf), (3mixed), (3typm). For the heterogeneity analysis, we merged data from individual years to get higher statistical significance as dividing the EU-SILC data according to various characteristic would end up with relatively small samples, affecting the statistical significance[15].

The gross GPG does not vary among the various levels of education − differences between the middle a base levels are not significant. The only significant variation is between high levels and the rest with GPG being higher for high education levels and the difference amounts to approx. 3 percentage points. However, the pattern reverses for adjusted GPG. There, again one cannot distinguish statistically between lower and middle educational levels, but the GPG is substantially lower for higher educational levels and amounts to 7.27% as opposed about 20% for the rest.

The gross GPG in Prague cannot be distinguished statistically from that elsewhere, so there are no significant interregional differences. Thus, it corresponds to the general level of the gross GPG. On the other hand, the situation is different for the adjusted GPG. The adjusted GPG in Prague is substantially lower (13.23%) than that in the other regions (19.04%) and the difference is statistically significant.

There is no significant difference in the GPG levels between the public and the private sectors. However, that may arise from the composition of the EU-SILC statistics for the early years, which did not comprise any substantial groups of public servants as is clear from the Table 3. This omission causes a substantial selection bias within the public sector domain and the data for the public sector for the years 2006−2008.

14 As the number of the various regressions is very large, the detailed data can be provided by the authors on request and we provide only the general results here.

15 By doing so, we lose information on the development over time. The authors conducted the heterogeneity analysis for each year. For those interested, the analysis can be provided by the authors on request, but yields relatively little as opposed to merged numbers, so we do not report it in the paper.

Figure 6: Heterogeneity analysis

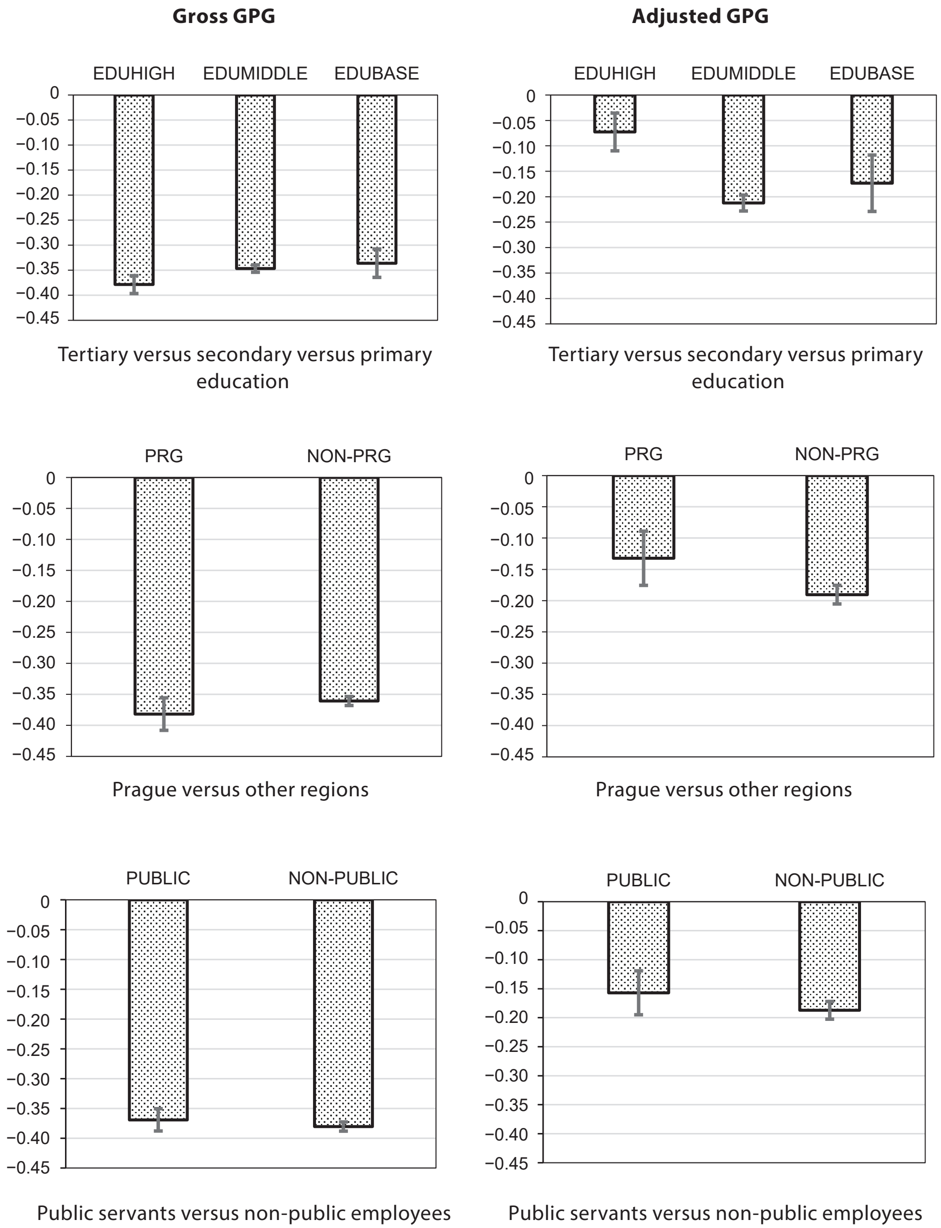
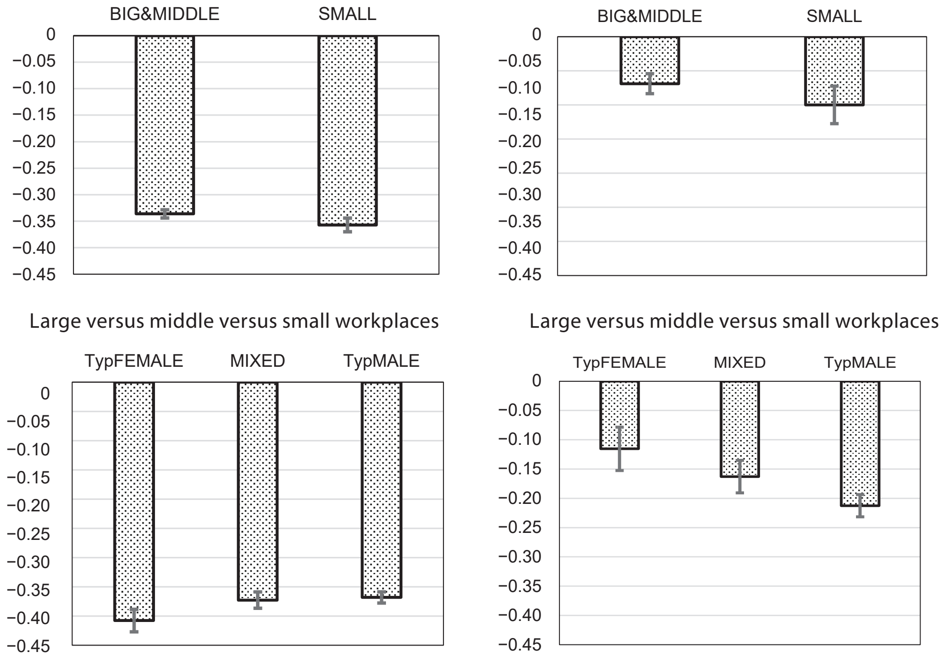


Figure 6: Continuation



Large versus middle versus small workplaces   Large versus middle versus small workplaces

Typically female, male and mixed sectors   Typically female, male and mixed sectors

Source: Authors' own calculation based on EU-SILC data

Table 3: Numbers of public servants in EU-SILC database

	2006	2007	2008	2009	2010	2011
Total	454	604	556	1,160	1,089	1,034
out of which: men	397	519	477	775	697	666
out of which: women	57	85	79	385	392	368

Source: EU-SILC

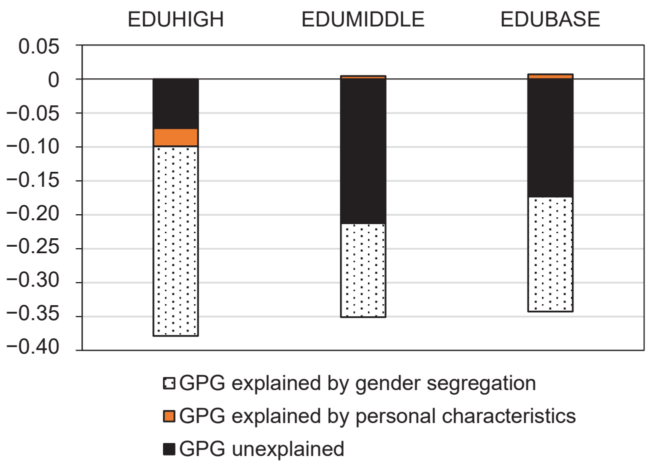
There is a difference between the adjusted GPG in the private (18.7%) and public domains (15.7%); however, the difference is not statistically significant.

The size matters – at least with the GPG. There is a statistically significant difference (approx. 2%) in the size of the gross GPG between large and middle-sized workplaces as opposed to small ones (less than 10 employees). However, the difference for the adjusted GPG looms even larger (more than 3% out of less than 20%).

The most interesting estimates came from those dealing with female as opposed to typically male dominated or mixed jobs. A typically female job was defined as a job in which there is a substantial majority (over 60%) of female workers and vice versa for males and typically male jobs. Mixed jobs are then such employments where one cannot find a clear substantial majority. According to the NACE classification, typical females sectors are those of I, K, P, Q, S and T. Typical male jobs are A through F, H, J and U. The other sectors remain mixed. The female sectors show a larger gross GPG than the other two groups of sectors. However, when adjusted GPG estimates are compared, the typically female sectors end up with the lowest GPG (11.55%) with mixed and male sectors amounting to16.29% and 21.26% respectively, all differences being statistically significant.

To sum up, the high gross GPG is a phenomenon of highly educated workers living in Prague and working in small workplaces in typically female sectors. As for the adjusted GPG, it is almost the opposite – the lowest levels of adjusted GPG are among highly educated people living in Prague and working in typically female sectors. The only factor lowering both kinds of the GPG is workplace size, with large ones having a lower GPG. This somewhat paradoxical outcome suggests a strong role of horizontal as well as vertical segregation of men and women on the labour market, driven mostly by the intra-family division of labour with men being mainly bread-winners as opposed to women being mostly care-takers, either staying out of the labour market or preferring less demanding jobs, lower workloads and not competing all too strongly for a promotion.

Figure 7: Personal characteristics versus labour market segregation



Source: Authors' own calculation based on EU-SILC data

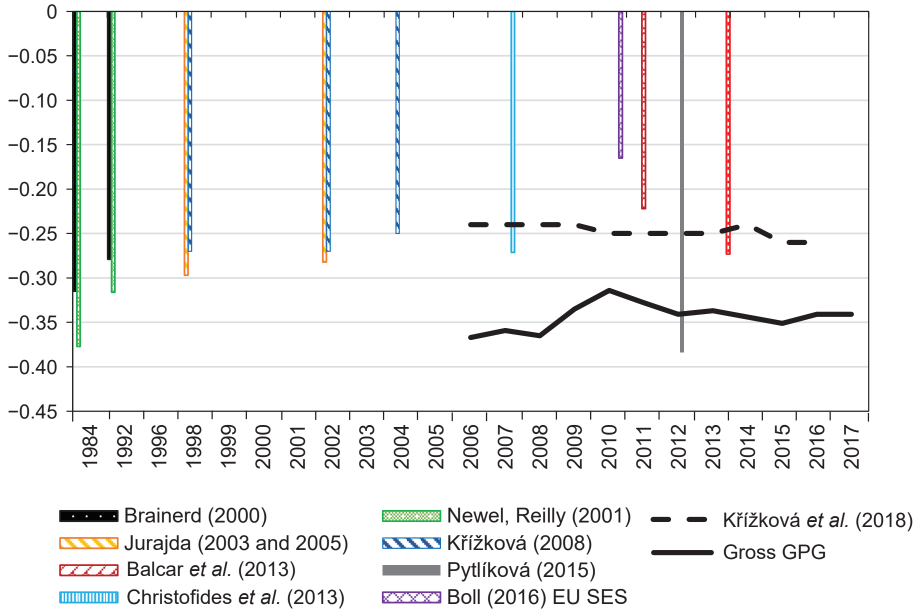
Such a family model seems to be especially numerous within the part of population with high levels of education and less so for other groups (as men and women mostly make couples with their educational peers). To demonstrate such an effect, we divided the control variables into those describing personal characteristics and those describing labour market segregation. Figure 7 shows the results of such a decomposition into education groups[16].

It is very clear that personal characteristics do not explain any substantial part of the GPG. By far the largest share of the explained part is driven by the labour market segregation, which is also the most powerful for the high educational levels in a population. A very similar scenario can be shown for other population subgroups.

5.4 Comparison with other results

As outlined in Section 1, several studies have tried to disentangle the difference in wages between women and men in the Czech Republic (or former Czechoslovakia). Figure 8 provides a graphical overview of estimates produced over the last 25 years, which also encompassed older estimates from pre-revolution times.

Figure 8: Gross GPG comparison with other studies



Source: Estimates taken from studies: Balcar *et al.* (2013); Boll (2016); Brainerd (2000); Christofides *et al.* (2013); Jurajda (2003, 2005); Křížková *et al.* (2018); Newel, Reilly (2001); Pytlíková (2015); and authors' own calculation based on EU-SILC data

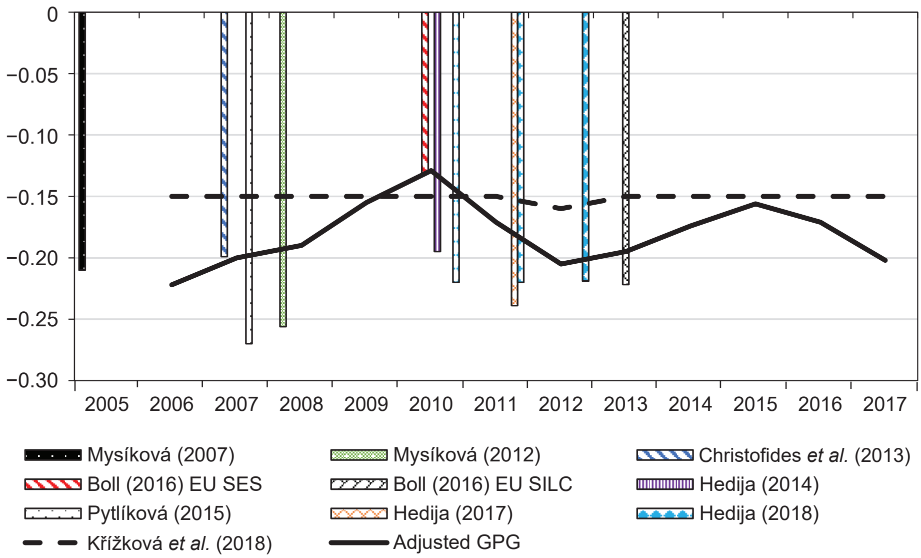
16 The detailed regression data can be provided by the authors on request.

When we compare our results (bold black line labelled Gross GPG) to those of others, we have to distinguish between two groups of estimates: those based on EU-SILC (this paper; Pytliková, 2015; Boll *et al.*, 2016 (SILC); Christofides *et al.*, 2013) and those based on other sources of data (mostly ISPV: Křížková *et al.*, 2018; Křížková, 2008; Jurajda, 2003; Jurajda, 2005; or other data: Boll *et al.*, 2016 (SES);, Balcar, 2013). When comparing these two groups of results, the non-EU-SILC estimates are substantially lower. EU-SILC estimates range between 27.13% (Boll *et al.*, 2016 (SILC)) and 38.4% (Pytliková, 2015) with our estimates lying somewhere in between. The higher levels of estimates in Pytliková (2015) can be expected as she incorporated only the data for employees between 20 and 49 years old that show the largest GPG that is driven by parenthood as was also shown in detail by our own estimates (Figures 4 and 5 above). The studies based on ISPV data provide lower estimates consistently. There might be several explanations for that. Firstly, a large part of the economy are usually excluded from the analysis (companies under 10 employees; Jurajda excluded companies under 100) and such an exclusion can cause a substantial bias downwards as smaller companies return larger GPG as opposed to middle-sized and large enterprises, which was also shown by our heterogeneity analysis above. Secondly, ISPV-based analysis usually measures hourly wages as opposed to monthly wages in analysis based on the EU-SILC. The difference underestimates somewhat the size of the GPG as men usually work longer hours than women. Such an outcome is not entirely exogenous to gender as most mothers adjust their workload according to their family duties as a result of a traditional family structure. This also shows the important role of parenthood in explaining the GPG with ISPV data providing no guidance as they do not include information on children.

As opposed to gross GPG, the adjusted GPG estimates are more consistent among various sources, Křížková *et al.* (2018) being the only outlier not for the size of the estimate but for its almost complete invariance to anything happening in the real economy between 2006 and 2016.

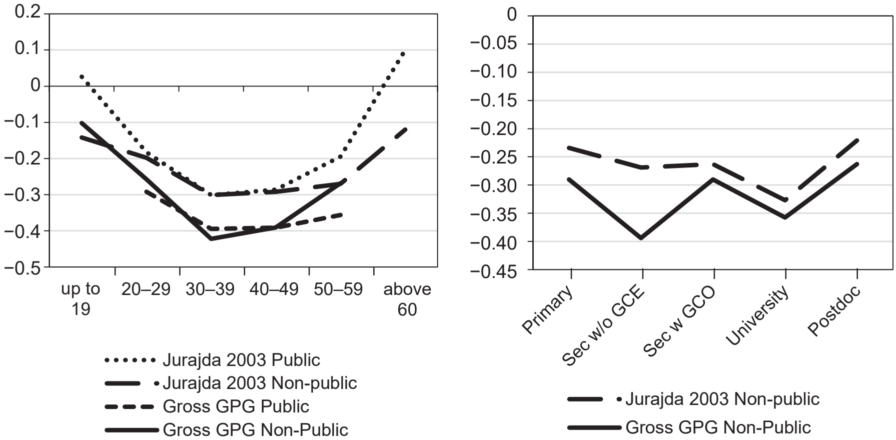
The size of the adjusted GPG varies with the controls included in the regression analysis. Pytliková (2015) is an example of a limited-amount-of-controls analysis, reporting the largest level of adjusted GPG (27%). Nevertheless, most of the estimates show a narrowing of the adjusted GPG in times of economic depression in 2009 and 2010, and our estimate is thus consistent with the overall development as well as size of the adjusted GPG reported by all other studies (regardless of their data sources). The only exception to the rule is Křížková *et al.* (2018). The explained part of the GPG averages to about 50% in our estimates, which is generally a higher share than with the other studies, which can usually explain only about a third of the gross GPG.

Figure 9: Adjusted GPG, comparison with other studies



Source: Estimates taken from studies: Boll (2016) EU SES; Boll (2016) EU SILC; Hedija (2014,2017,2018); Christofides *et al.* (2013); Křížková *et al.* (2018); Mysíková (2007,2012); Pytlíková (2015); and authors' own calculations based on EU-SILC data

Figure 10: GPG comparison for age groups and educational cohorts



Gross GPG for age cohorts

Gross GPG for education categories

Source: Authors' own computation based on EU-SILC data, Jurajda (2003)

Figure 10 also shows a comparison between the gross GPG estimated by Jurajda (2003) for various education subgroups. The results are also strikingly similar.

As our analysis provided evidence of the critical role of parenthood in shaping the GPG, we can compare our results with those of Jurajda (2003), who provided a simplified estimate of the development of the gross GPG by age group. We adjusted the age groups to correspond to those of Jurajda (2003) and excluded small and middle-sized workplaces to obtain the closest dataset to that of Jurajda (2003) and estimated the gross GPG for public and non-public employees. The results are strikingly similar, as shown in Figure 9. Similar comparisons can be provided for the results of Pytliková (2015) with a very similar outlook.

5.5 GPG for different earning groups

In order to investigate how the GPG progresses through the various earning groups, we applied the conditional quantile regression analysis to our data for each year. By doing so, we also explored the possible existence of a glass ceiling or a sticky floor in our sample. According to the literature, a glass ceiling is indicated if the pay gap is substantially larger for the upper quantiles of the income distribution. Similarly, a sticky floor is suspected if there is a substantially larger pay gap for lower quantiles of the income distribution. The outcomes of the analysis are shown in detail in Appendix 8 and an overview in Figure 11. The estimates of all the beta coefficients representing the gross GPG are highly statistically significant.

Figure 11: Gross GPG over various quantiles (2006–2017)

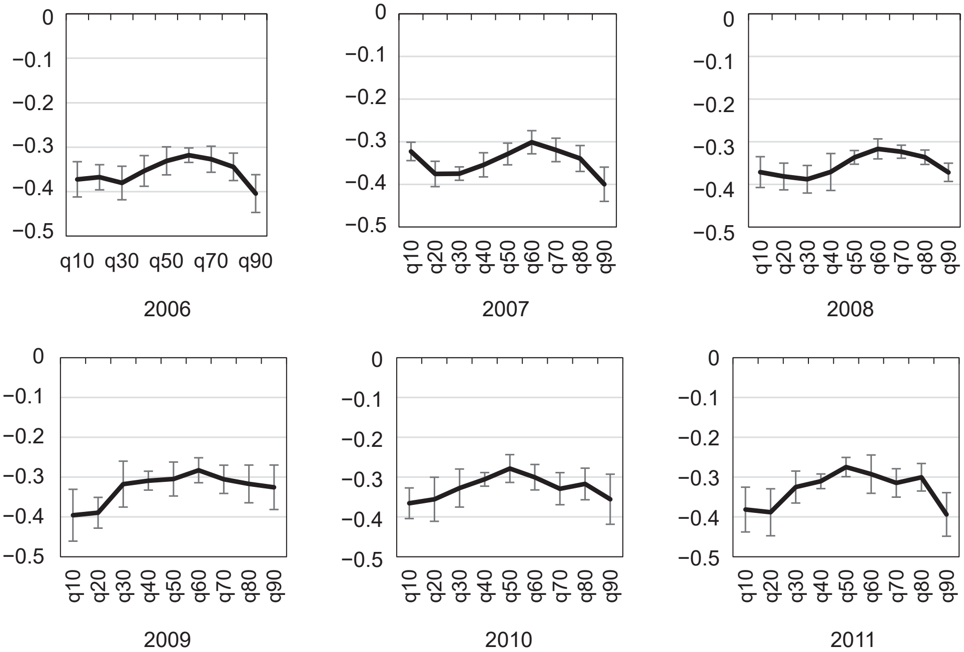
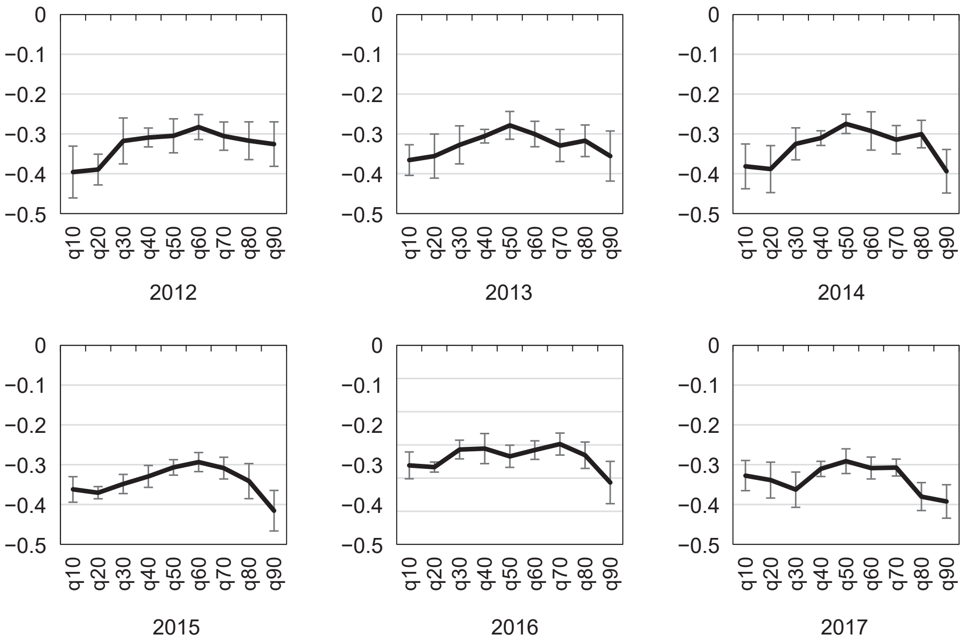


Figure 11: Continuation



Source: Authors' own computation based on EU-SILC data

There is a general pattern in most years for which we had data. That pattern progresses in the form of an inverted U-shaped curve, indicating that the largest GPG manifests itself at the edges of the earning distribution; hence, there is some indication for the existence of both a glass ceiling and a sticky floor. It is not as surprising for the upper part of the income ladder as it is for its lowest parts. The right-hand side of the inverted U-curve remains substantially lower for each year, *i.e.*, the gross GPG is higher for those with higher incomes (esp. the highest decile) than for those around the median of the income distribution. Similarly, for nine years (out of twelve, exceptions are 2009, 2010 and 2017), those with the lowest incomes (the lowest two deciles) experience a larger gross GPG than those with higher incomes.

6. Conclusions

The aim of the study was to provide consistent and comparable estimates of various measures of the GPG in the Czech Republic over a time span of twelve years (2006−2017) in order to be able to evaluate their development. We constructed three measures of the GPG: the gross GPG, the GPG adjusted for control variables and the GPG adjusted for control variables and the selection bias of those not active on the labour market (especially women). The gross GPG remains relatively time-invariant for all the years surveyed (2006–2017) at a level around 35%. There is a slight and statistically weak (10% level of statistical significance) dependence of the size of the GPG on the GDP development over time. Such a size of the GPG is consistent with other findings based on EU-SILC data (Pytliková, 2015; Christofides *et al.*, 2013; Boll *et al.*, 2016) but not in line with others mostly based on ISPV data (Jurajda, 2003; Jurajda, 2005; Křížková, 2008; Křížková, 2018; Balcar, 2013; Boll *et al.*, 2016). The adjusted GPG development captured by our estimates is mostly consistent in size and movements with the general picture provided by other studies with the exception of Křížková *et al.* (2018). The explained part of the GPG varies but about 50% on average can be explained by inclusion of control and selection variables, the rest remaining unexplained. Only about a third of the gross GPG stems from differences in labour market characteristics (endowment effects) if we also include the interaction part. The remaining two thirds result from the coefficient effect.

Our analysis strongly suggests that the driving force behind the GPG formation, size and development is the effect of parenthood on women's labour market decisions, resulting in strong horizontal as well as vertical gender segregation. We find that non-parents face a gross GPG that is about 30 percentage points lower than that for parents. Without the effect of parenthood, the gross GPG and the adjusted GPG cannot be distinguished statistically. The effect of parenthood thus drives all the results. The analysis by the age groups suggests that parenthood casts long shadows. It takes two decades for women to catch up and for the GPG to narrow after the effects of parenthood disappear. For the age cohorts over 50, the GPG equates with the one before parenthood, *i.e.*, approximately 25% for the gross GPG and between 10 and 15% for the adjusted GPG. The data also reveal that parenthood affects the gross GPG much more profoundly than the adjusted GPG, which rises due to the effects of parenthood only by a tenth of its original size. Our conclusions regarding the effect of parenthood are strongly supported by only two other papers that dealt with the same problem (Jurajda, 2003; Pytliková, 2015). As a substantial part of the papers dealing with the GPG use ISPV data, the effect of parenthood is hardly discernible as ISPV data do not include information on employees' children. The conclusion is highly relevant for public policy because if the parenthood drives the GPG rather than other variables (including possible discrimination), the most important tool to lower the existing GPG lies with the policies lowering the costs of parenthood to families and mothers in particular.

Thirdly, we measured the GPG in the Czech Republic across various demographic groups and also across earning quantiles. There are some interesting differences in the size of the GPG for some demographic groups under scrutiny (educational cohorts, Prague citizens versus non-Prague dwellers, public servants versus anyone else, employees of small rather than middle-sized or large companies, employees in female and male dominated sectors). In general, those with higher education, living in Prague and working in typically female sectors in small workplaces face a high gross GPG. At the same time, almost the same group – better-educated, living in Prague, working in typically female sectors – shows the lowest adjusted GPG. The only factor that lowers both kinds of the GPG is workplace size – a lower GPG is attributable to large enterprises.

Last but not least, by employing the quantile regression on our data, we found some indication of the existence of a glass ceiling and a sticky floor as the gross GPG is most pronounced in the highest and the lowest quantiles of the income distribution.

Appendix

Appendix 1: Descriptive statistics

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
TOTAL_INCOME	131,083	213,830	195,605	244,741	247,844	250,762	257,278	261,515	267,865	280,967	291,798	306,802
AGE	51.80	42.50	43.20	43.40	43.70	43.80	44.10	44.30	44.50	44.70	44.80	44.70
PARTNER	0.75	0.81	0.81	0.80	0.80	0.80	0.79	0.79	0.79	0.79	0.79	0.79
CHILDREN	0.62	0.96	0.95	0.92	0.90	0.90	0.89	0.90	0.87	0.85	0.85	0.85
EDUHIGH	0.12	0.15	0.15	0.15	0.17	0.18	0.19	0.19	0.20	0.22	0.23	0.24
EDUMIDDLE	0.75	0.78	0.78	0.77	0.76	0.76	0.75	0.76	0.75	0.74	0.73	0.71
PRAHA	0.09	0.09	0.08	0.09	0.09	0.10	0.10	0.10	0.12	0.13	0.13	0.12
STRCECHY	0.10	0.10	0.10	0.12	0.12	0.12	0.13	0.12	0.12	0.12	0.11	0.11
ODPRAC_LET	26.95	21.42	21.93	22.04	22.15	22.06	22.36	22.54	22.78	22.86	22.91	22.90
CONTRACT	–	0.66	0.67	0.68	0.67	0.67	0.68	0.68	0.67	0.68	0.69	0.71
SUPERVISOR	–	0.15	0.15	0.15	0.16	0.15	0.15	0.15	0.16	0.16	0.16	0.16
Number of observations	11,536	9,695	11,210	9,820	9,114	8,724	8,510	7,980	7,636	7,474	8,091	8,170

Source: Authors' own computation based on EU-SILC data

Appendix 2: Selection equation

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Variables	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**	**lpfW**
lnYOW	−0.0264*** (0.00592)	−0.114*** (0.00841)	−0.128*** (0.00851)	−0.123*** (0.00779)	−0.112*** (0.00729)	−0.105*** (0.00760)	−0.105*** (0.00770)	−0.0777*** (0.00743)	−0.0972*** (0.00812)	−0.113*** (0.00672)	−0.114*** (0.00669)	−0.126*** (0.00679)
PARTNER	−0.422*** (0.0695)	−0.245*** (0.0800)	−0.229*** (0.0799)	−0.222*** (0.0851)	−0.346*** (0.0825)	−0.321*** (0.0845)	−0.409*** (0.0867)	−0.299*** (0.0933)	−0.237** (0.0936)	−0.143 (0.101)	−0.256** (0.0999)	−0.190* (0.102)
PART-NERW	0.426*** (0.0570)	0.211*** (0.0639)	0.144** (0.0633)	0.0579 (0.0676)	0.119* (0.0648)	0.159** (0.0670)	0.228*** (0.0683)	0.102 (0.0709)	0.0718 (0.0745)	0.0271 (0.0796)	0.178** (0.0785)	0.164** (0.0817)
AGE30	0.404*** (0.0767)	−0.305*** (0.0814)	−0.423*** (0.0805)	−0.260*** (0.0880)	−0.255*** (0.0866)	−0.146 (0.0915)	−0.230** (0.0955)	−0.187* (0.0995)	−0.0898 (0.104)	−0.252** (0.110)	−0.302*** (0.108)	−0.138 (0.109)
AGE 31_45	0.878*** (0.0727)	−0.00753 (0.0707)	−0.0387 (0.0693)	0.0232 (0.0757)	−0.133* (0.0712)	0.0305 (0.0736)	0.0366 (0.0758)	−0.00643 (0.0790)	0.0569 (0.0796)	−0.0360 (0.0838)	−0.0841 (0.0826)	−0.0524 (0.0865)
CHILD2W	−1.593*** (0.0795)	−1.435*** (0.0731)	−1.419*** (0.0684)	−1.435*** (0.0740)	−1.339*** (0.0733)	−1.504*** (0.0780)	−1.471*** (0.0804)	−1.505*** (0.0880)	−1.648*** (0.0939)	−1.185*** (0.0954)	−1.340*** (0.0941)	−1.404*** (0.0921)
CHILD 3_5W	−1.079*** (0.0790)	−1.010*** (0.0693)	−0.956*** (0.0653)	−0.905*** (0.0704)	−0.789*** (0.0697)	−0.832*** (0.0720)	−0.852*** (0.0734)	−0.872*** (0.0741)	−0.863*** (0.0771)	−0.694*** (0.0834)	−0.536*** (0.0811)	−0.663*** (0.0846)
CHILD 6_15W	0.00266 (0.0657)	−0.199*** (0.0593)	−0.0926 (0.0580)	−0.160** (0.0630)	−0.0733 (0.0619)	−0.151** (0.0653)	−0.168** (0.0666)	−0.194*** (0.0690)	−0.156** (0.0716)	−0.149** (0.0733)	−0.252*** (0.0731)	−0.153** (0.0747)
EDU-HIGH	0.860*** (0.102)	0.945*** (0.109)	0.897*** (0.104)	0.828*** (0.108)	1.039*** (0.105)	1.169*** (0.110)	1.006*** (0.114)	1.136*** (0.120)	1.183*** (0.126)	1.196*** (0.129)	1.120*** (0.127)	0.976*** (0.131)
EDU-MIDDLE	0.403*** (0.0701)	0.646*** (0.0809)	0.598*** (0.0777)	0.578*** (0.0823)	0.787*** (0.0820)	0.891*** (0.0897)	0.734*** (0.0950)	0.782*** (0.0981)	0.802*** (0.104)	0.851*** (0.108)	0.818*** (0.108)	0.747*** (0.116)
Constant	0.470*** (0.0831)	1.752*** (0.112)	2.037*** (0.113)	1.984*** (0.115)	1.682*** (0.111)	1.460*** (0.115)	1.678*** (0.122)	1.444*** (0.127)	1.473*** (0.134)	1.299*** (0.128)	1.444*** (0.130)	1.525*** (0.137)
Observa-tions	4,024	4,759	5,490	4,790	4,449	4,251	4,156	3,836	3,646	3,550	3,809	3,882

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 3: Gross GPG

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Variables	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**
SEX	−0.367***	−0.359***	−0.365***	−0.335***	−0.314***	−0.328***	−0.341***	−0.337***	−0.344***	−0.351***	−0.341***	−0.341***
	(0.0127)	(0.0113)	(0.00972)	(0.0108)	(0.0115)	(0.0119)	(0.0120)	(0.0127)	(0.0133)	(0.0133)	(0.0125)	(0.0122)
Constant	12.35***	12.41***	12.48***	12.53***	12.55***	12.57***	12.60***	12.60***	12.63***	12.67***	12.71***	12.76***
	(0.00859)	(0.00767)	(0.00680)	(0.00740)	(0.00787)	(0.00817)	(0.00828)	(0.00878)	(0.00919)	(0.00928)	(0.00865)	(0.00847)
Observations	6,252	8,005	8,231	8,136	7,340	7,028	6,899	6,477	6,110	6,019	6,549	6,567
R–squared	0.118	0.112	0.146	0.106	0.092	0.098	0.105	0.098	0.099	0.103	0.102	0.106

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

Appendix 4: GPG with control variables

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Variables	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**
SEX	−0.225*** (0.0257)	−0.219*** (0.0226)	−0.202*** (0.0183)	−0.165*** (0.0208)	−0.133*** (0.0224)	−0.173*** (0.0226)	−0.202*** (0.0221)	−0.201*** (0.0239)	−0.172*** (0.0249)	−0.160*** (0.0249)	−0.173*** (0.0237)	−0.200*** (0.0231)
HUSBAND	−0.0597*** (0.0196)	−0.0291* (0.0175)	−0.0316** (0.0138)	−0.0515*** (0.0160)	−0.0738*** (0.0169)	−0.0706*** (0.0171)	−0.0774*** (0.0167)	−0.0583*** (0.0184)	−0.0647*** (0.0192)	−0.0604*** (0.0192)	−0.0340* (0.0185)	−0.0465*** (0.0178)
WIFE	0.0868*** (0.0205)	0.0959*** (0.0178)	0.0846*** (0.0148)	0.119*** (0.0164)	0.129*** (0.0180)	0.0948*** (0.0182)	0.0676*** (0.0180)	0.0907*** (0.0190)	0.117*** (0.0199)	0.136*** (0.0199)	0.125*** (0.0187)	0.0886*** (0.0186)
PRAHA	0.238*** (0.0270)	0.201*** (0.0239)	0.206*** (0.0203)	0.218*** (0.0229)	0.163*** (0.0240)	0.176*** (0.0238)	0.138*** (0.0235)	0.136*** (0.0255)	0.131*** (0.0254)	0.118*** (0.0254)	0.0878*** (0.0229)	0.101*** (0.0226)
STR CECHY	0.0807*** (0.0191)	0.0853*** (0.0165)	0.0799*** (0.0137)	0.115*** (0.0147)	0.103*** (0.0152)	0.145*** (0.0155)	0.135*** (0.0152)	0.135*** (0.0165)	0.138*** (0.0177)	0.150*** (0.0176)	0.141*** (0.0169)	0.122*** (0.0169)
EDUHIGH	0.695*** (0.0289)	0.568*** (0.0260)	0.564*** (0.0206)	0.520*** (0.0238)	0.494*** (0.0261)	0.576*** (0.0283)	0.537*** (0.0283)	0.499*** (0.0319)	0.462*** (0.0345)	0.504*** (0.0350)	0.502*** (0.0319)	0.532*** (0.0312)
EDUMIDDLE	0.260*** (0.0242)	0.199*** (0.0217)	0.211*** (0.0169)	0.183*** (0.0195)	0.161*** (0.0220)	0.217*** (0.0244)	0.184*** (0.0247)	0.184*** (0.0283)	0.145*** (0.0312)	0.184*** (0.0320)	0.193*** (0.0292)	0.221*** (0.0288)
ODPRAC_LET	0.0174*** (0.00231)	0.0133*** (0.00211)	0.0129*** (0.00173)	0.0188*** (0.00198)	0.0145*** (0.00210)	0.0146*** (0.00217)	0.0146*** (0.00220)	0.0171*** (0.00237)	0.0181*** (0.00249)	0.0173*** (0.00247)	0.0177*** (0.00230)	0.0183*** (0.00225)
ODPRAC_ LET2	−0.000404*** (5.18e−05)	−0.000348*** (4.69e−05)	−0.000318*** (3.82e−05)	−0.000467*** (4.36e−05)	−0.000378*** (4.66e−05)	−0.000364*** (4.81e−05)	−0.000353*** (4.85e−05)	−0.000409*** (5.23e−05)	−0.000427*** (5.50e−05)	−0.000388*** (5.48e−05)	−0.000413*** (5.09e−05)	−0.000413*** (5.04e−05)
SIZEBIG	0.131*** (0.0152)	0.114*** (0.0145)	0.163*** (0.0117)	0.135*** (0.0137)	0.0993*** (0.0145)	0.00489 (0.0141)	0.00459 (0.0138)	−0.00886 (0.0151)	0.115*** (0.0170)	0.142*** (0.0167)	0.142*** (0.0155)	0.131*** (0.0151)
SIZE-MIDDLE	0.0637*** (0.0146)	0.0366*** (0.0140)	0.0929*** (0.0115)	0.0567*** (0.0130)	0.0457*** (0.0138)	−0.0312** (0.0147)	0.00533 (0.0146)	−0.0254 (0.0155)	0.0587*** (0.0164)	0.0610*** (0.0162)	0.0827*** (0.0152)	0.0642*** (0.0152)
MUNIBIG	0.00879 (0.0207)	0.0525*** (0.0185)	0.0286* (0.0152)	0.0187 (0.0176)	0.0622*** (0.0186)	0.0513*** (0.0184)	0.0773*** (0.0185)	0.0908*** (0.0200)	0.0819*** (0.0204)	0.0523** (0.0208)	0.0724*** (0.0186)	0.0702*** (0.0182)

Appendix 4: Continuation

MUNIMIDDLE	0.0411** (0.0178)	0.0261* (0.0156)	0.0405*** (0.0127)	0.0136 (0.0153)	0.0284* (0.0161)	0.0435*** (0.0165)	0.0478*** (0.0168)	0.0609*** (0.0181)	0.0539*** (0.0201)	−0.0108 (0.0203)	−0.0373* (0.0195)	−0.0187 (0.0193)
NACEA	−0.0846*** (0.0313)	−0.128*** (0.0281)	−0.120*** (0.0234)	−0.0238 (0.0266)	−0.0330 (0.0290)	−0.0195 (0.0302)	0.0205 (0.0298)	−0.0133 (0.0323)	−0.0544 (0.0348)	−0.0916*** (0.0347)	0.00532 (0.0325)	−0.0176 (0.0337)
NACEB	−0.627 (0.441)	0.182 (0.253)	0.180 (0.363)	0.164*** (0.0448)	0.193*** (0.0481)	0.252*** (0.0457)	0.283*** (0.0487)	0.269*** (0.0548)	0.200*** (0.0622)	0.0770 (0.0742)	0.0576 (0.0680)	0.0610 (0.0697)
NACEC	−0.0207 (0.0205)	−0.0472** (0.0186)	−0.0482*** (0.0118)	0.0310* (0.0164)	0.0149 (0.0176)	0.0522*** (0.0169)	0.112*** (0.0169)	0.0935*** (0.0181)	0.0344* (0.0207)	0.0242 (0.0205)	0.0454** (0.0190)	0.0488*** (0.0188)
NACED	0.0247 (0.0738)	−0.0364 (0.0680)	0.0503 (0.0520)	0.203*** (0.0419)	0.188*** (0.0430)	0.253*** (0.0428)	0.290*** (0.0431)	0.160*** (0.0505)	0.179*** (0.0555)	0.0972* (0.0581)	0.0624 (0.0549)	0.107** (0.0507)
NACEE	0.0664* (0.0378)	−0.00293 (0.0329)	0.00129 (0.0270)	−0.000891 (0.0492)	−0.0135 (0.0496)	0.0357 (0.0506)	0.0740 (0.0473)	0.0595 (0.0523)	−0.00230 (0.0544)	0.00982 (0.0596)	0.0639 (0.0575)	−0.0149 (0.0565)
NACEF	0.0597 (0.0765)	−0.0911 (0.0693)	−0.0231 (0.0647)	0.0854*** (0.0203)	0.0538** (0.0221)	0.0807*** (0.0226)	0.129*** (0.0223)	0.0797*** (0.0242)	0.0308 (0.0257)	0.0695*** (0.0257)	0.0951*** (0.0248)	0.0802*** (0.0245)
NACEH	−0.0347 (0.0239)	−0.121*** (0.0216)	−0.113*** (0.0163)	0.170*** (0.0223)	0.127*** (0.0238)	0.147*** (0.0242)	0.148*** (0.0241)	0.153*** (0.0261)	0.141*** (0.0287)	0.129*** (0.0283)	0.158*** (0.0264)	0.122*** (0.0264)
NACEI	−0.120*** (0.0374)	−0.174*** (0.0317)	−0.140*** (0.0266)	−0.0283 (0.0287)	−0.0339 (0.0294)	−0.0759** (0.0307)	0.0166 (0.0304)	−0.0320 (0.0335)	−0.0959*** (0.0364)	−0.0825** (0.0381)	−0.0681** (0.0344)	−0.0564 (0.0359)
NACEJ	0.0584** (0.0291)	0.0640** (0.0256)	0.0618*** (0.0199)	0.226*** (0.0337)	0.248*** (0.0359)	0.210*** (0.0363)	0.263*** (0.0351)	0.318*** (0.0370)	0.288*** (0.0396)	0.244*** (0.0394)	0.307*** (0.0368)	0.238*** (0.0354)
NACEK	0.277*** (0.0340)	0.219*** (0.0309)	0.0931*** (0.0256)	0.351*** (0.0335)	0.318*** (0.0336)	0.332*** (0.0342)	0.303*** (0.0326)	0.334*** (0.0355)	0.330*** (0.0408)	0.264*** (0.0378)	0.305*** (0.0338)	0.286*** (0.0336)
NACEM	0.101*** (0.0237)	0.0352* (0.0212)	0.0700*** (0.0155)	0.195*** (0.0284)	0.171*** (0.0306)	0.0936*** (0.0308)	0.177*** (0.0295)	0.181*** (0.0307)	0.143*** (0.0331)	0.215*** (0.0339)	0.172*** (0.0312)	0.190*** (0.0303)
NACEN	−0.0550* (0.0282)	−0.0565** (0.0254)	−0.0478*** (0.0184)	−0.00478 (0.0373)	−0.0498 (0.0357)	−0.109*** (0.0371)	−0.0673* (0.0383)	−0.0809** (0.0388)	−0.0986*** (0.0383)	−0.162*** (0.0401)	−0.220*** (0.0374)	−0.142*** (0.0371)
NACEP	0.0589** (0.0280)	0.0267 (0.0249)	0.000847 (0.0189)	0.0171 (0.0227)	0.0205 (0.0244)	0.00710 (0.0246)	0.0478** (0.0242)	0.0678*** (0.0262)	−0.0114 (0.0279)	−0.0240 (0.0274)	−0.0168 (0.0257)	−0.0147 (0.0259)

Appendix 4: Continuation

NACER	−0.0751**	−0.185***	−0.0675***	−0.00669	−0.0246	−0.0539	−0.0197	0.0480	−0.0233	−0.0355	−0.0581	0.0122
	(0.0319)	(0.0288)	(0.0255)	(0.0385)	(0.0420)	(0.0427)	(0.0438)	(0.0516)	(0.0512)	(0.0495)	(0.0450)	(0.0452)
NACET	−0.261	0.104	–	−0.0660	1.355***	0.323	−0.343	−0.572**	−0.606**	−0.724***	−0.476**	−0.0204
	(0.442)	(0.310)	–	(0.239)	(0.420)	(0.421)	(0.239)	(0.249)	(0.255)	(0.252)	(0.213)	(0.244)
NACEU	−0.373	–	−0.801**	–	0.212	−0.0896	0.122	0.0271	−0.263	−0.491	−0.235	0.0145
	(0.442)	–	(0.363)	–	(0.242)	(0.421)	(0.413)	(0.431)	(0.255)	(0.309)	(0.246)	(0.243)
CONTRACT	–	0.0357***	0.116***	0.0202	0.0628***	0.109***	0.155***	0.175***	0.108***	0.136***	0.0937***	0.0978***
	–	(0.0134)	(0.0125)	(0.0127)	(0.0136)	(0.0131)	(0.0129)	(0.0136)	(0.0156)	(0.0155)	(0.0146)	(0.0147)
SUPERVISOR	–	0.238***	0.252***	0.238***	0.252***	0.272***	0.296***	0.303***	0.288***	0.260***	0.276***	0.257***
	–	(0.0137)	(0.0107)	(0.0129)	(0.0134)	(0.0140)	(0.0139)	(0.0149)	(0.0154)	(0.0155)	(0.0146)	(0.0142)
NACEL	–	–	–	0.185***	0.187***	0.181***	0.232***	0.247***	0.240***	0.114*	0.123**	0.181***
	–	–	–	(0.0635)	(0.0553)	(0.0574)	(0.0595)	(0.0646)	(0.0640)	(0.0606)	(0.0566)	(0.0560)
NACEO	–	–	–	0.169***	0.170***	0.169***	0.219***	0.204***	0.109***	0.107***	0.137***	0.157***
	–	–	–	(0.0231)	(0.0237)	(0.0239)	(0.0240)	(0.0256)	(0.0278)	(0.0278)	(0.0259)	(0.0255)
NACEQ	–	–	–	0.123***	0.128***	0.159***	0.210***	0.196***	0.109***	0.100***	0.0726***	0.0570**
	–	–	–	(0.0223)	(0.0238)	(0.0239)	(0.0237)	(0.0257)	(0.0272)	(0.0266)	(0.0249)	(0.0243)
NACES	–	–	–	−0.0990**	−0.0631	−0.134***	−0.138***	−0.155***	−0.125**	−0.107**	−0.0411	−0.0293
	–	–	–	(0.0417)	(0.0416)	(0.0452)	(0.0469)	(0.0469)	(0.0487)	(0.0434)	(0.0421)	(0.0437)
Constant	11.70***	11.85***	11.77***	11.83***	11.88***	11.86***	11.84***	11.81***	11.86***	11.83***	11.88***	11.93***
	(0.0392)	(0.0354)	(0.0282)	(0.0322)	(0.0352)	(0.0379)	(0.0376)	(0.0422)	(0.0448)	(0.0456)	(0.0414)	(0.0411)
Observations	6,252	8,005	8,231	8,136	7,340	7,028	6,899	6,477	6,110	6,019	6,549	6,567
R–squared	0.317	0.336	0.424	0.359	0.346	0.360	0.383	0.363	0.354	0.367	0.367	0.355

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 5: GPG with control variables accounting for a selection bias

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Variables	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**
SEX	−0.222***	−0.200***	−0.190***	−0.155***	−0.129***	−0.171***	−0.205***	−0.195***	−0.174***	−0.156***	−0.171***	−0.202***
	(0.0275)	(0.0234)	(0.0188)	(0.0212)	(0.0229)	(0.0232)	(0.0226)	(0.0244)	(0.0256)	(0.0255)	(0.0242)	(0.0235)
HUSBAND	−0.0592***	−0.0247	−0.0282**	−0.0476***	−0.0718***	−0.0696***	−0.0786***	−0.0545***	−0.0655***	−0.0584***	−0.0336*	−0.0471***
	(0.0197)	(0.0175)	(0.0139)	(0.0161)	(0.0171)	(0.0173)	(0.0169)	(0.0186)	(0.0194)	(0.0194)	(0.0185)	(0.0178)
WIFE	0.0868***	0.0972***	0.0861***	0.120***	0.129***	0.0950***	0.0674***	0.0913***	0.117***	0.137***	0.125***	0.0885***
	(0.0205)	(0.0177)	(0.0148)	(0.0164)	(0.0180)	(0.0182)	(0.0180)	(0.0190)	(0.0199)	(0.0199)	(0.0187)	(0.0186)
PRAHA	0.238***	0.199***	0.204***	0.216***	0.163***	0.176***	0.138***	0.135***	0.131***	0.118***	0.0876***	0.101***
	(0.0270)	(0.0239)	(0.0203)	(0.0229)	(0.0240)	(0.0238)	(0.0235)	(0.0255)	(0.0255)	(0.0254)	(0.0229)	(0.0226)
STRCECHY	0.0806***	0.0848***	0.0793***	0.116***	0.103***	0.145***	0.135***	0.135***	0.138***	0.150***	0.141***	0.123***
	(0.0191)	(0.0165)	(0.0137)	(0.0147)	(0.0152)	(0.0155)	(0.0152)	(0.0165)	(0.0177)	(0.0176)	(0.0169)	(0.0169)
EDUHIGH	0.693***	0.556***	0.556***	0.514***	0.490***	0.574***	0.539***	0.492***	0.464***	0.498***	0.500***	0.534***
	(0.0295)	(0.0263)	(0.0208)	(0.0240)	(0.0265)	(0.0288)	(0.0287)	(0.0326)	(0.0351)	(0.0356)	(0.0325)	(0.0314)
EDUMIDDLE	0.258***	0.189***	0.205***	0.178***	0.157***	0.215***	0.186***	0.177***	0.147***	0.179***	0.192***	0.223***
	(0.0247)	(0.0219)	(0.0170)	(0.0197)	(0.0224)	(0.0249)	(0.0251)	(0.0289)	(0.0317)	(0.0325)	(0.0297)	(0.0290)
ODPRAC_LET	0.0173***	0.0126***	0.0122***	0.0181***	0.0143***	0.0145***	0.0148***	0.0168***	0.0182***	0.0171***	0.0176***	0.0184***
	(0.00232)	(0.00213)	(0.00175)	(0.00200)	(0.00213)	(0.00219)	(0.00222)	(0.00239)	(0.00250)	(0.00249)	(0.00230)	(0.00226)
ODPRAC_ LET2	−0.000402***	−0.000338***	−0.000308***	−0.000458***	−0.000374***	−0.000362***	−0.000355***	−0.000404***	−0.000428***	−0.000385***	−0.000413***	−0.000414***
	(5.23e−05)	(4.70e−05)	(3.83e−05)	(4.38e−05)	(4.68e−05)	(4.83e−05)	(4.86e−05)	(5.24e−05)	(5.51e−05)	(5.49e−05)	(5.10e−05)	(5.04e−05)
SIZEBIG	0.131***	0.113***	0.162***	0.133***	0.0988***	0.00476	0.00471	−0.00924	0.116***	0.141***	0.142***	0.131***
	(0.0152)	(0.0145)	(0.0117)	(0.0137)	(0.0146)	(0.0141)	(0.0138)	(0.0151)	(0.0170)	(0.0168)	(0.0156)	(0.0152)
SIZE-MIDDLE	0.0635***	0.0354**	0.0909***	0.0555***	0.0454***	−0.0312**	0.00534	−0.0252	0.0588***	0.0603***	0.0825***	0.0647***
	(0.0146)	(0.0139)	(0.0115)	(0.0130)	(0.0138)	(0.0147)	(0.0146)	(0.0155)	(0.0164)	(0.0162)	(0.0153)	(0.0152)
MUNIBIG	0.00879	0.0518***	0.0288*	0.0190	0.0622***	0.0513***	0.0773***	0.0905***	0.0819***	0.0522**	0.0724***	0.0702***
	(0.0207)	(0.0185)	(0.0152)	(0.0176)	(0.0186)	(0.0184)	(0.0185)	(0.0200)	(0.0204)	(0.0208)	(0.0186)	(0.0182)

Appendix 5: Continuation

MUNI-MIDDLE	0.0410** (0.0178)	0.0257* (0.0156)	0.0390*** (0.0127)	0.0131 (0.0153)	0.0281* (0.0161)	0.0434*** (0.0165)	0.0480*** (0.0168)	0.0605*** (0.0181)	0.0542*** (0.0201)	−0.0112 (0.0203)	−0.0375* (0.0195)	−0.0186 (0.0193)
NACEA	−0.0851*** (0.0313)	−0.130*** (0.0281)	−0.121*** (0.0234)	−0.0259 (0.0266)	−0.0336 (0.0290)	−0.0197 (0.0302)	0.0208 (0.0299)	−0.0142 (0.0324)	−0.0543 (0.0348)	−0.0924*** (0.0347)	0.00499 (0.0325)	−0.0170 (0.0337)
NACEB	−0.627 (0.441)	0.187 (0.253)	0.182 (0.363)	0.162*** (0.0448)	0.192*** (0.0481)	0.252*** (0.0457)	0.284*** (0.0487)	0.268*** (0.0548)	0.200*** (0.0622)	0.0762 (0.0742)	0.0574 (0.0680)	0.0611 (0.0697)
NACEC	−0.0210 (0.0206)	−0.0500*** (0.0186)	−0.0491*** (0.0118)	0.0283* (0.0164)	0.0140 (0.0176)	0.0517*** (0.0169)	0.112*** (0.0169)	0.0919*** (0.0181)	0.0347* (0.0207)	0.0230 (0.0206)	0.0450** (0.0191)	0.0493*** (0.0188)
NACED	0.0242 (0.0738)	−0.0400 (0.0680)	0.0474 (0.0520)	0.201*** (0.0419)	0.187*** (0.0430)	0.252*** (0.0429)	0.291*** (0.0431)	0.159*** (0.0506)	0.179*** (0.0555)	0.0968* (0.0581)	0.0622 (0.0549)	0.108** (0.0507)
NACEE	0.0661* (0.0378)	−0.00488 (0.0329)	0.000682 (0.0269)	−0.00254 (0.0492)	−0.0142 (0.0496)	0.0349 (0.0506)	0.0750 (0.0473)	0.0584 (0.0523)	−0.00180 (0.0544)	0.00801 (0.0597)	0.0635 (0.0575)	−0.0142 (0.0565)
NACEF	0.0593 (0.0765)	−0.0958 (0.0693)	−0.0251 (0.0647)	0.0826*** (0.0203)	0.0527** (0.0221)	0.0801*** (0.0226)	0.129*** (0.0223)	0.0781*** (0.0243)	0.0311 (0.0257)	0.0682*** (0.0257)	0.0947*** (0.0249)	0.0810*** (0.0246)
NACEH	−0.0352 (0.0239)	−0.124*** (0.0216)	−0.115*** (0.0163)	0.168*** (0.0223)	0.126*** (0.0238)	0.146*** (0.0243)	0.149*** (0.0241)	0.151*** (0.0261)	0.142*** (0.0287)	0.128*** (0.0283)	0.157*** (0.0264)	0.122*** (0.0264)
NACEI	−0.120*** (0.0374)	−0.178*** (0.0317)	−0.143*** (0.0266)	−0.0321 (0.0287)	−0.0349 (0.0294)	−0.0765** (0.0308)	0.0172 (0.0304)	−0.0340 (0.0336)	−0.0954*** (0.0365)	−0.0838** (0.0382)	−0.0683** (0.0344)	−0.0557 (0.0359)
NACEJ	0.0580** (0.0291)	0.0617** (0.0256)	0.0618*** (0.0199)	0.223*** (0.0337)	0.247*** (0.0359)	0.210*** (0.0363)	0.263*** (0.0351)	0.316*** (0.0370)	0.288*** (0.0396)	0.243*** (0.0394)	0.306*** (0.0368)	0.239*** (0.0355)
NACEK	0.277*** (0.0340)	0.214*** (0.0309)	0.0908*** (0.0256)	0.348*** (0.0335)	0.317*** (0.0336)	0.331*** (0.0342)	0.304*** (0.0326)	0.331*** (0.0356)	0.331*** (0.0408)	0.263*** (0.0378)	0.305*** (0.0338)	0.286*** (0.0336)
NACEM	0.101*** (0.0237)	0.0309 (0.0212)	0.0682*** (0.0155)	0.192*** (0.0284)	0.170*** (0.0307)	0.0927*** (0.0308)	0.178*** (0.0295)	0.178*** (0.0307)	0.144*** (0.0332)	0.214*** (0.0340)	0.171*** (0.0312)	0.191*** (0.0303)
NACEN	−0.0556** (0.0283)	−0.0604** (0.0254)	−0.0505*** (0.0184)	−0.00789 (0.0373)	−0.0509 (0.0358)	−0.109*** (0.0372)	−0.0666* (0.0383)	−0.0823** (0.0388)	−0.0985** (0.0383)	−0.163*** (0.0401)	−0.220*** (0.0374)	−0.141*** (0.0371)
NACEP	0.0586** (0.0281)	0.0243 (0.0249)	−0.000934 (0.0189)	0.0128 (0.0228)	0.0187 (0.0244)	0.00636 (0.0247)	0.0489** (0.0243)	0.0652** (0.0263)	−0.0109 (0.0279)	−0.0257 (0.0275)	−0.0173 (0.0258)	−0.0138 (0.0259)

Appendix 5: Continuation

NACER	−0.0757** (0.0319)	−0.189*** (0.0288)	−0.0698*** (0.0255)	−0.00909 (0.0386)	−0.0250 (0.0420)	−0.0541 (0.0427)	−0.0193 (0.0438)	0.0461 (0.0516)	−0.0233 (0.0512)	−0.0367 (0.0496)	−0.0580 (0.0450)	0.0126 (0.0452)
NACET	−0.261 (0.442)	0.0936 (0.310)	– –	−0.0728 (0.239)	1.373*** (0.420)	0.323 (0.421)	−0.343 (0.239)	−0.573** (0.249)	−0.605** (0.255)	−0.726*** (0.252)	−0.477** (0.213)	−0.0191 (0.244)
NACEU	−0.373 (0.442)	– –	−0.804** (0.363)	– –	0.207 (0.242)	−0.0907 (0.421)	0.123 (0.414)	0.0240 (0.431)	−0.263 (0.255)	−0.492 (0.309)	−0.235 (0.246)	0.0156 (0.244)
lambda	−0.00963 (0.0310)	−0.0914*** (0.0296)	−0.0675*** (0.0241)	−0.0604** (0.0280)	−0.0260 (0.0287)	−0.0127 (0.0308)	0.0160 (0.0303)	−0.0412 (0.0353)	0.0114 (0.0369)	−0.0321 (0.0359)	−0.0107 (0.0344)	0.0167 (0.0329)
CONTRACT	– –	0.0346*** (0.0134)	0.112*** (0.0126)	0.0199 (0.0127)	0.0627*** (0.0136)	0.108*** (0.0131)	0.155*** (0.0129)	0.175*** (0.0136)	0.108*** (0.0156)	0.136*** (0.0155)	0.0937*** (0.0146)	0.0980*** (0.0147)
SUPERVISOR	– –	0.238*** (0.0137)	0.252*** (0.0107)	0.237*** (0.0129)	0.252*** (0.0134)	0.272*** (0.0140)	0.296*** (0.0139)	0.302*** (0.0149)	0.288*** (0.0154)	0.260*** (0.0155)	0.276*** (0.0146)	0.257*** (0.0142)
NACEL	– –	– –	– –	0.181*** (0.0635)	0.186*** (0.0553)	0.181*** (0.0575)	0.232*** (0.0595)	0.245*** (0.0646)	0.240*** (0.0640)	0.112* (0.0606)	0.123** (0.0566)	0.181*** (0.0560)
NACEO	– –	– –	– –	0.166*** (0.0231)	0.169*** (0.0237)	0.168*** (0.0240)	0.220*** (0.0240)	0.202*** (0.0256)	0.110*** (0.0279)	0.106*** (0.0279)	0.137*** (0.0259)	0.158*** (0.0256)
NACEQ	– –	– –	– –	0.120*** (0.0223)	0.127*** (0.0238)	0.158*** (0.0239)	0.210*** (0.0238)	0.194*** (0.0258)	0.109*** (0.0272)	0.0979*** (0.0267)	0.0720*** (0.0250)	0.0579** (0.0244)
NACES	– –	– –	– –	−0.102** (0.0418)	−0.0640 (0.0416)	−0.135*** (0.0453)	−0.137*** (0.0469)	−0.157*** (0.0469)	−0.125** (0.0487)	−0.109** (0.0435)	−0.0417 (0.0421)	−0.0285 (0.0437)
Constant	11.70*** (0.0399)	11.88*** (0.0363)	11.79*** (0.0292)	11.85*** (0.0332)	11.89*** (0.0364)	11.86*** (0.0391)	11.83*** (0.0388)	11.83*** (0.0438)	11.86*** (0.0462)	11.84*** (0.0470)	11.89*** (0.0425)	11.92*** (0.0418)
Observations	6,252	8,005	8,231	8,136	7,340	7,028	6,899	6,477	6,110	6,019	6,549	6,567
R–squared	0.317	0.336	0.424	0.360	0.346	0.360	0.383	0.363	0.354	0.367	0.367	0.355

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 6: Oaxaca–Blinder decompositions

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Variables	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**	**lnY Differential**
Prediction_1	12.35***	12.41***	12.48***	12.53***	12.55***	12.57***	12.60***	12.60***	12.63***	12.67***	12.71***	12.76***
	(0.00870)	(0.00776)	(0.00687)	(0.00770)	(0.00815)	(0.00833)	(0.00833)	(0.00905)	(0.00929)	(0.00951)	(0.00907)	(0.00870)
Prediction_2	11.98***	12.05***	12.12***	12.20***	12.23***	12.24***	12.26***	12.27***	12.28***	12.32***	12.37***	12.42***
	(0.00923)	(0.00827)	(0.00691)	(0.00756)	(0.00809)	(0.00850)	(0.00862)	(0.00895)	(0.00955)	(0.00938)	(0.00856)	(0.00856)
Difference	0.367***	0.359***	0.365***	0.335***	0.314***	0.328***	0.341***	0.337***	0.344***	0.351***	0.341***	0.341***
	(0.0127)	(0.0113)	(0.00974)	(0.0108)	(0.0115)	(0.0119)	(0.0120)	(0.0127)	(0.0133)	(0.0134)	(0.0125)	(0.0122)
Endowments	0.107***	0.0909***	0.104***	0.0751***	0.0883***	0.0920***	0.0943***	0.0883***	0.0885***	0.0981***	0.0699***	0.0699***
	(0.0202)	(0.0169)	(0.0134)	(0.0154)	(0.0165)	(0.0174)	(0.0175)	(0.0187)	(0.0206)	(0.0194)	(0.0178)	(0.0173)
Coefficients	0.276***	0.248***	0.225***	0.196***	0.186***	0.231***	0.265***	0.255***	0.221***	0.210***	0.187***	0.221***
	(0.0227)	(0.0196)	(0.0159)	(0.0186)	(0.0200)	(0.0201)	(0.0195)	(0.0213)	(0.0221)	(0.0222)	(0.0209)	(0.0206)
Interaction	−0.0162	0.0205	0.0355*	0.0642***	0.0404*	0.00504	−0.0189	−0.00617	0.0341	0.0424	0.0835***	0.0500**
	(0.0277)	(0.0233)	(0.0183)	(0.0216)	(0.0233)	(0.0238)	(0.0233)	(0.0253)	(0.0270)	(0.0263)	(0.0245)	(0.0240)
Observations	6,252	8,005	8,231	8,136	7,340	7,028	6,899	6,477	6,110	6,019	6,549	6,567

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 7: Gross GPG per age groups – data merged over the whole time period

	up 25	26–30	31–35	36–40	41–45	46–50	51–55	56–60
Variables	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**	**lnY**
SEX	−0.330*** (0.0237)	−0.319*** (0.0123)	−0.445*** (0.01000)	−0.462*** (0.00914)	−0.391*** (0.00893)	−0.353*** (0.00929)	−0.301*** (0.00870)	−0.221*** (0.0103)
YEAR2007	0.0216 (0.0508)	0.0806*** (0.0258)	0.0478** (0.0230)	0.0606*** (0.0232)	0.0294 (0.0228)	0.0445* (0.0230)	0.0472** (0.0211)	0.0656** (0.0270)
YEAR2008	0.182*** (0.0522)	0.186*** (0.0271)	0.155*** (0.0229)	0.151*** (0.0234)	0.115*** (0.0226)	0.144*** (0.0230)	0.154*** (0.0210)	0.126*** (0.0264)
YEAR2009	0.258*** (0.0533)	0.237*** (0.0278)	0.212*** (0.0230)	0.222*** (0.0232)	0.199*** (0.0227)	0.212*** (0.0231)	0.208*** (0.0209)	0.161*** (0.0263)
YEAR2010	0.304*** (0.0549)	0.292*** (0.0286)	0.232*** (0.0238)	0.229*** (0.0236)	0.231*** (0.0233)	0.237*** (0.0236)	0.222*** (0.0216)	0.209*** (0.0262)
YEAR2011	0.364*** (0.0557)	0.285*** (0.0292)	0.260*** (0.0240)	0.231*** (0.0238)	0.240*** (0.0238)	0.265*** (0.0234)	0.241*** (0.0220)	0.224*** (0.0265)
YEAR2012	0.181*** (0.0554)	0.260*** (0.0294)	0.235*** (0.0243)	0.234*** (0.0231)	0.226*** (0.0234)	0.256*** (0.0236)	0.229*** (0.0221)	0.235*** (0.0263)
YEAR2013	0.119** (0.0571)	0.239*** (0.0297)	0.235*** (0.0251)	0.266*** (0.0232)	0.234*** (0.0238)	0.271*** (0.0238)	0.241*** (0.0227)	0.256*** (0.0267)
YEAR2014	0.138** (0.0581)	0.297*** (0.0304)	0.298*** (0.0257)	0.282*** (0.0237)	0.245*** (0.0237)	0.262*** (0.0241)	0.262*** (0.0230)	0.258*** (0.0270)
YEAR2015	0.108* (0.0581)	0.304*** (0.0312)	0.343*** (0.0262)	0.314*** (0.0237)	0.301*** (0.0233)	0.303*** (0.0242)	0.329*** (0.0231)	0.312*** (0.0275)
YEAR2016	0.232*** (0.0563)	0.366*** (0.0303)	0.371*** (0.0258)	0.346*** (0.0237)	0.368*** (0.0228)	0.340*** (0.0239)	0.365*** (0.0222)	0.324*** (0.0271)
YEAR2017	0.318*** (0.0550)	0.435*** (0.0296)	0.397*** (0.0262)	0.413*** (0.0238)	0.397*** (0.0228)	0.425*** (0.0235)	0.406*** (0.0223)	0.397*** (0.0272)
Constant	12.13*** (0.0381)	12.30*** (0.0199)	12.39*** (0.0178)	12.43*** (0.0180)	12.40*** (0.0179)	12.34*** (0.0178)	12.27*** (0.0166)	12.22*** (0.0205)
Observations	2,619	6,869	11,431	13,604	13,534	13,207	14,208	9,720
R–squared	0.099	0.127	0.180	0.187	0.161	0.132	0.116	0.076

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 8: Quantile regressions: gross GPG

2006									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.372*** (0.0202)	−0.368*** (0.0144)	−0.381*** (0.0193)	−0.353*** (0.0176)	−0.331*** (0.0161)	−0.318*** (0.00831)	−0.327*** (0.0150)	−0.344*** (0.0158)	−0.404*** (0.0217)
Constant	11.79*** (0.00986)	11.96*** (0.00407)	12.10*** (0.00256)	12.21*** (0.00955)	12.31*** (0.0105)	12.42*** (0.00846)	12.55*** (0.0117)	12.68*** (0.0107)	12.94*** (0.0150)
Observations	6,252	6,252	6,252	6,252	6,252	6,252	6,252	6,252	6,252

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

2007									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.323*** (0.0110)	−0.376*** (0.0152)	−0.375*** (0.00801)	−0.354*** (0.0145)	−0.329*** (0.0130)	−0.301*** (0.0139)	−0.319*** (0.0140)	−0.339*** (0.0155)	−0.400*** (0.0205)
Constant	11.84*** (0.0113)	12.06*** (0.00929)	12.17*** (0.00762)	12.28*** (0.00847)	12.39*** (0.00650)	12.47*** (0.0134)	12.60*** (0.00784)	12.75*** (0.0102)	13.01*** (0.0203)
Observations	8,005	8,005	8,005	8,005	8,005	8,005	8,005	8,005	8,005

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

Appendix 8: Continuation

2008									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.371*** (0.0184)	−0.381*** (0.0160)	−0.388*** (0.0165)	−0.371*** (0.0221)	−0.336*** (0.00818)	−0.317*** (0.0120)	−0.323*** (0.00781)	−0.336*** (0.00861)	−0.372*** (0.0109)
Constant	11.98*** (0.0149)	12.16*** (0.00468)	12.27*** (0.00963)	12.37*** (0.0129)	12.44*** (0.00726)	12.54*** (0.0108)	12.66*** (0.00527)	12.79*** (0.00515)	13.02*** (0.0104)
Observations	8,231	8,231	8,231	8,231	8,231	8,231	8,231	8,231	8,231

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

2009									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.284*** (0.0156)	−0.323*** (0.0136)	−0.348*** (0.0157)	−0.310*** (0.0119)	−0.317*** (0.00907)	−0.329*** (0.00837)	−0.337*** (0.0137)	−0.321*** (0.0146)	−0.368*** (0.0190)
Constant	11.98*** (0.0147)	12.17*** (0.00743)	12.31*** (0.0147)	12.39*** (0.0109)	12.50*** (0.00710)	12.61*** (0.00117)	12.73*** (0.00708)	12.87*** (0.00447)	13.11*** (0.0143)
Observations	8,136	8,136	8,136	8,136	8,136	8,136	8,136	8,136	8,136

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses
Source: Authors' own computation based on EU-SILC data

Appendix 8: Continuation

2010									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.277***	−0.321***	−0.351***	−0.319***	−0.302***	−0.284***	−0.293***	−0.302***	−0.360***
	(0.0204)	(0.0207)	(0.0125)	(0.0161)	(0.0130)	(0.0160)	(0.0112)	(0.0134)	(0.0242)
Constant	11.99***	12.19***	12.32***	12.43***	12.53***	12.62***	12.75***	12.89***	13.13***
	(0.0212)	(0.0161)	(0.00169)	(0.0114)	(0.0101)	(0.0143)	(0.00723)	(0.00839)	(0.0125)
Observations	7,340	7,340	7,340	7,340	7,340	7,340	7,340	7,340	7,340

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

2011									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.360***	−0.352***	−0.340***	−0.315***	−0.300***	−0.286***	−0.290***	−0.298***	−0.369***
	(0.0361)	(0.0182)	(0.0215)	(0.0130)	(0.0157)	(0.0144)	(0.0128)	(0.0167)	(0.0270)
Constant	12.01***	12.21***	12.34***	12.45***	12.56***	12.66***	12.77***	12.91***	13.17***
	(0.0242)	(0.0138)	(0.00344)	(0.00625)	(0.00543)	(0.00419)	(0.0132)	(0.0124)	(0.0195)
Observations	7,028	7,028	7,028	7,028	7,028	7,028	7,028	7,028	7,028

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

Appendix 8: Continuation

2012									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.396*** (0.0332)	−0.389*** (0.0197)	−0.318*** (0.0294)	−0.309*** (0.0121)	−0.305*** (0.0218)	−0.283*** (0.0160)	−0.305*** (0.0181)	−0.317*** (0.0242)	−0.325*** (0.0285)
Constant	12.07*** (0.0177)	12.25*** (0.0119)	12.35*** (0.00901)	12.47*** (0.0110)	12.59*** (0.0130)	12.67*** (0.0110)	12.79*** (0.0137)	12.94*** (0.0138)	13.16*** (0.0201)
Observations	6,899	6,899	6,899	6,899	6,899	6,899	6,899	6,899	6,899

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

2013									
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.366*** (0.0197)	−0.356*** (0.0282)	−0.327*** (0.0244)	−0.306*** (0.00872)	−0.278*** (0.0178)	−0.300*** (0.0163)	−0.329*** (0.0205)	−0.317*** (0.0202)	−0.355*** (0.0321)
Constant	12.06*** (0.0158)	12.25*** (0.0100)	12.36*** (0.0141)	12.47*** (0.00753)	12.59*** (0.00933)	12.69*** (0.0131)	12.82*** (0.0155)	12.95*** (0.0100)	13.20*** (0.0277)
Observations	6,477	6,477	6,477	6,477	6,477	6,477	6,477	6,477	6,477

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

Appendix 8: Continuation

				2014					
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.381***	−0.388***	−0.325***	−0.310***	−0.275***	−0.292***	−0.314***	−0.300***	−0.394***
	(0.0287)	(0.0301)	(0.0205)	(0.00939)	(0.0123)	(0.0245)	(0.0182)	(0.0176)	(0.0279)
Constant	12.08***	12.27***	12.39***	12.48***	12.60***	12.70***	12.84***	12.96***	13.25***
	(0.0212)	(0.0188)	(0.0103)	(0.00835)	(0.00768)	(0.0146)	(0.00915)	(0.0163)	(0.0256)
Observations	6,110	6,110	6,110	6,110	6,110	6,110	6,110	6,110	6,110

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

				2015					
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.362***	−0.370***	−0.348***	−0.329***	−0.307***	−0.293***	−0.308***	−0.341***	−0.415***
	(0.0163)	(0.00781)	(0.0123)	(0.0140)	(0.00992)	(0.0122)	(0.0139)	(0.0225)	(0.0260)
Constant	12.10***	12.32***	12.45***	12.56***	12.64***	12.74***	12.87***	13.02***	13.29***
	(0.0158)	(0.00481)	(0.00744)	(0.00829)	(0.00942)	(0.00673)	(0.00630)	(0.0119)	(0.0195)
Observations	6,019	6,019	6,019	6,019	6,019	6,019	6,019	6,019	6,019

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

Appendix 8: Continuation

				2016					
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.362***	−0.367***	−0.314***	−0.311***	−0.334***	−0.316***	−0.298***	−0.331***	−0.414***
	(0.0206)	(0.00775)	(0.0144)	(0.0231)	(0.0170)	(0.0143)	(0.0169)	(0.0203)	(0.0325)
Constant	12.13***	12.34***	12.47***	12.59***	12.69***	12.80***	12.90***	13.06***	13.34***
	(0.0172)	(0.00437)	(0.00516)	(0.00532)	(0.0135)	(0.0136)	(0.0133)	(0.0179)	(0.0186)
Observations	6,549	6,549	6,549	6,549	6,549	6,549	6,549	6,549	6,549

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

				2017					
Variables	***q*10**	***q*20**	***q*30**	***q*40**	***q*50**	***q*60**	***q*70**	***q*80**	***q*90**
SEX	−0.327***	−0.339***	−0.363***	−0.310***	−0.291***	−0.308***	−0.307***	−0.380***	−0.392***
	(0.0193)	(0.0230)	(0.0226)	(0.00971)	(0.0158)	(0.0140)	(0.0108)	(0.0179)	(0.0214)
Constant	12.17***	12.40***	12.53***	12.64***	12.74***	12.84***	12.96***	13.14***	13.35***
	(0.0187)	(0.0195)	(0.0125)	(0.00971)	(0.0110)	(0.00780)	(0.00880)	(0.0131)	(0.0160)
Observations	6,567	6,567	6,567	6,567	6,567	6,567	6,567	6,567	6,567

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Standard errors in parentheses

Source: Authors' own computation based on EU-SILC data

References

Balcar, J., Filipová, L., Machová, Z. (2012). Gender Wage Gap in the Czech Republic: First Descriptive Analysis Based on Survey 2011. *Review of Economic Perspectives – Národohospodářský obzor*, 12(3), 151–167, https://doi.org/10.2478/v10135-012-0007-5

Becker, G. (1981). *A Treatise on the Family*. Cambridge: Harvard University Press.

Becker, G. S. (1965). A Theory of the Allocation of Time. *Economic Journal*, 75(299), 493–517, https://doi.org/10.2307/2228949

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436–455, https://doi.org/10.2307/144855

Boll, C., Leppin, J. S. (2015). Die geschlechtsspezifische Lohnlücke in Deutschland: Umfang, Ursachen und Interpretation. *Wirtschaftsdienst*, 95(4), 249–254, https://doi.org/10.1007/s10273-015-1814-y

Boll, C., Leppin, J. S., Rossen, A., et al. (2016). *Magnitude and Impact Factors of the Gender Pay Gap in EU Countries*. Report Prepared for and Financed by the European Commission – Directorate-General for Justice, European Union (Hrsg.), Hamburg.

Brainerd, E. (2000). Women in Transition: Changes in Gender Wage Differentials in Eastern Europe and the Former Soviet Union. *Industrial and Labor Relations Review*, 54(1), 138–162, https://doi.org/10.2307/2696036

Del Río, C., Gradín, C., Cantó, O. (2011). The Measurement of Gender Wage Discrimination: The Distributional Approach Revisited. *The Journal of Economic Inequality*, 9(1), 57–86, https://doi.org/10.1007/s10888-010-9130-7

Doeringer, P. B., Piore, M. J. (1971). *Internal Labor Markets and Manpower Analysis*. Lexington, MA: Heath.

Dohmen, T., Falk, A. (2010). You Get What You Pay For: Incentives and Selection in the Education System. *The Economic Journal*, 120(546), 256–271, https://doi.org/10.1111/j.1468-0297.2010.02376.x

Donohue, J. J. (2007). Chapter 18 Antidiscrimination Law, in Polinsky, A. M., Shavell, S., eds., *Handbook of Law and Economics*, Volume 2. North Holland, pp. 1387–1472, https://doi.org/10.1016/s1574-0730(07)02018-X

Filer, R. K., Jurajda, Š., Plánovský, J. (1999). Education and Wages in the Czech and Slovak Republics during Transition. *Labour Economics*, 6(4), 581–593, https://doi.org/10.1016/S0927-5371(99)00032-9

Gneezy, U., Niederle, M., Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3), 1049–1074, https://doi.org/10.1162/00335530360698496

Ham, J., Svejnar, J., Terrell, K. (1995). The Czech and Slovak Labor Markets during the Transition, in Commander, S., Coricelli, F., eds., *Unemployment. Restructuring, and the Labour Market in Eastern Europe and Russia*. Washington DC: The World Bank, pp. 91–146.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–163, https://doi.org/10.2307/1912352

Hedija, V. (2014). *Gender Pay Gap in Different Sectors of Czech Economy*. In 32nd International Conference Mathematical Methods in Economics, pp. 275–280.

Hedija, V. (2017). Sector-specific Gender Pay Gap: Evidence from the European Union Countries. *Economic research-Ekonomska istraživanja*, 30(1), 1804–1819, https://doi.org/10.1080/1331677X.2017.1392886

Hedija, V. (2018). Is the Rule of Law Significant for the Explanation of the Differences in the Gender Pay Gap? *Prague Economic Papers*, 27(6), 704–722, https://doi.org/10.18267/j.pep.684

Christofides, L. N., Polycarpou, A., Vrachimis, K. (2013). Gender Wage Gaps, 'Sticky Floors' and 'Glass Ceilings' in Europe. *Labour Economics*, 21, 86–102, https://doi.org/10.1016/j.labeco.2013.01.003

Jurajda, Š. (2003). Gender Wage Gap and Segregation in Enterprises and the Public Sector in Late Transition Countries. *Journal of Comparative Economics*, 31(2), 199–222, https://doi.org/10.1016/S0147-5967(03)00040-4

Jurajda, Š. (2005). Gender Segregation and Wage Gap: An East-West Comparison. *Journal of the European Economic Association*, 3(2–3), 598–607, https://doi.org/10.1162/jeea.2005.3.2-3.598

Křížková, A., Penner, A. M., Petersen, T. (2008). Genderové nerovnosti v odměňování na stejné pracovní pozici: sociální vyloučení žen. *Gender, rovné příležitosti, výzkum*, 9(2), 55–67.

Křížková, A., Pospíšilová, K., Maříková, H., et al. (2018). *Rozdíly v odměňování žen a mužů v ČR. Pracoviště, zaměstnání, stejná práce a rozklad faktorů*. Praha: Ministerstvo práce a sociálních věcí. ISBN 978-80-7421-162-1.

Light, A., Ureta, M. (1995). Early-career Work Experience and Gender Wage Differentials. *Journal of Labor Economics*, 13(1), 121–154, https://doi.org/10.1086/298370

Mysíková, M. (2007). The Gender Wage Gap and its Determinants, in Večerník, J., ed., *The Czech Labour Market: Changing Structures and Work Orientations*. Sociologické studie / Sociological Studies 07:4. Praha: Sociologický ústav AV ČR, v.v.i., pp. 55–68. ISBN 978-80-7330-132-3.

Mysíková, M. (2012). Gender Wage Gap in the Czech Republic and Central European Countries. *Prague Economic Papers*, 21(3), 328–346, https://doi.org/10.18267/j.pep.427

Newell, A., Reilly, B. (2001). The Gender Pay Gap in the Transition from Communism: Some Empirical Evidence. *Economic Systems*, 25(4), 287–304, https://doi.org/10.1016/S0939-3625(01)00028-0

Niederle, M., Vesterlund, L. (2007). Why Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067–1101, https://doi.org/10.1162/qjec.122.3.1067

Oaxaca, R. (1973). Male-female Wage Differentials in Urban Labour Markets. *International Economic Review*, 14(3), 693–709, https://doi.org/10.2307/2525981

Olivetti, C., Petrongolo, B. (2008). Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26(4), 621–654, https://doi.org/10.1086/589458

Pertold-Gebicka, B. (2019). Parental Leave Length and Mothers' Careers: What Can be Inferred from Occupational Allocation? *Applied Economics*, 52(9), 879–904, https://doi.org/10.1080/00036846.2019.1646870

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, 62(4), 659–661.

Pytliková, M. (2015). *Wage Differences Related to Motherhood and Children in the Family*. CERGE-EI IDEA. Prague Study No. 11/2015.

Večerník, J. (1995). Changing Earnings Distribution in the Czech Republic: Survey Evidence from 1988–1994. *Economics of Transition*, 3(3), 355–371, https://doi.org/10.1111/j.1468-0351.1995.tb00147.x

Waldfogel, J. (1997). The Effect of Children on Women's Wages. *American Sociological Review*, 62(2), 209–217, https://doi.org/10.2307/2657300

Zajíčková, D., Zajíček, M. (2020). Mateřská sankce v České republice, její vývoj a zdroje. *Politická ekonomie*, 68(5), 569–604, https://doi.org/10.18267/j.polek.1292

Zajíčková, D., Zajíček, M., Rašticová, M. (2021). Does Anti-Discrimination Legislation Work? The Case of Motherhood Penalty in the Czech Republic. *Employee Responsibilities and Rights Journal*, 33(1), 25–45, https://doi.org/10.1007/s10672-020-09360-0