Universität Hohenheim
Fakultät Wirtschafts- und Sozialwissenschaften

# Recent Developments in Gender Differences in Pay

Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. oec.)

vorgelegt der
Fakultät der Wirtschafts– und Sozialwissenschaften
der Universität Hohenheim
von

**Marina Francesca Katharina Töpfer, M.Sc.**
Institut für Volkswirtschaftslehre
Lehrstuhl für Volkswirtschaftslehre, insbesondere Dienstleistungs- und Arbeitsmarktökonomik (520G)

Stuttgart–Hohenheim 2017

Datum der mündlichen Doktorprüfung: 8. November 2017

Dekan: Prof. Dr. Karsten Hadwich
Prüfungsvorsitz: Prof. Dr. Alfonso Sousa-Poza
Erstgutachter: Prof. Dr. Thomas Beißinger
Zweitgutachter: Prof. Dr. Carolina Castagnetti

*To my family*

# Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Carolina Castagnetti and Professor Thomas Beißinger for their great support and guidance during my doctoral studies. I wish to thank them for the time, advice and assistance they gave me.

Furthermore, I would like to thank my co-authors, Professor Carolina Castagnetti and Professor Luisa Rosti, for their remarkable professionalism and helpfulness during my period(s) as a visiting student/scholar at the University of Pavia as well as for our discussions that were of enormous value for my dissertation project. Their enthusiasm was a continuous source of motivation to me.

I am indebted to Professor Thomas Beißinger for supporting my research stays in Pavia and Chengdu as well as various conferences and workshop participations and providing a productive and stimulating working environment.

I am extremely grateful for financial support from the Ministry of Science, Research and Arts Baden-Württemberg (“Landesgraduiertenförderung”). I am also grateful for financial support for my research stay in China and workshop participation to the thematic network Innovation, Entrepreneurship, and Finance (INEF) and for conference participation to the “Universitätsbund Hohenheim e.V.”. Special thanks to the Italian Institute for the Development of Vocational Training for Workers (ISFOL) and in particular to Emiliano Mandrone for data provision.

I thank all my friends and in particular Giovanni, Bianca, Karthika, Veronica and Vincent as well as my colleagues, Dario, Martyna, Philipp and Sebastian, for their constant encouragement and motivation. I am also grateful to Anita Bott for her support in administrative matters and beyond.

Many thanks to my sisters, Laura and Lisa, my grandparents and Mascha; they have been of great moral support to me. The most special thanks goes to my parents for their invaluable encouragement and constant belief.

Marina Töpfer

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **CATI** | Computer Assisted Telephone Interviewing |
| **CPG** | Counterfactual Pay Gap |
| **CQR** | Conditional Quantile Regression |
| **EU** | European Union |
| **GOEG** | Gender Overall Earnings Gap |
| **GPG** | Gender Pay Gap |
| **IMR** | Inverse Mills Ratio |
| **ISFOL** | Institute for the Development of Vocational Training for Workers |
| **LIE** | Law of Iterated Expectations |
| **MLE** | Maximum Likelihood Estimation |
| **OECD** | Organisation for Economic Co-operation and Development |
| **OLS** | Ordinary Least Squares |
| **OVB** | Omitted Variable Bias |
| **PLUS** | Participation, Labor, Unemployment Survey |
| **PPWG** | Public-Private Sector Wage Gap |
| **RIF-OLS** | Recentered-Influence-Function Ordinary Least Squares |
| **SLS** | Semiparametric Least Squares |
| **SURE** | Seemingly Unrelated Equations |

**UQPE** Unconditional Partial Quantile Effect

**UQR** Unconditional Quantile Regression

**US** United States

# Chapter 1

# Introduction

Gender differences in pay are a global phenomenon and have obtained much attention from both policy makers and economic literature (see for example Blau and Kahn, 1992; 2003; 2006; Goldin, 2014; Fortin, 2015). In order to reduce gender differences in earnings, anti-discrimination laws, female board quotas and family-friendly policies were implemented. A general result in the literature is that the earnings gap between men and women converges over time (e.g. Goldin, 2014). However, in recent years, the rate of convergence slowed down (see Blau and Kahn, 2006; England, 2006). Smaller differences in human capital and labor market presence (Goldin, 2006), occupational segregation (Cotter et al., 2004; England, 2006), technical development (Black and Spitz-Oener, 2010) as well as anti-discrimination legislation and changes in attitudes towards women in the labor market (Goldin, 2006; Fortin, 2015) are identified as major causes for the decline of the Gender Pay Gap (GPG) over time. Despite the implementation of equal-pay legislation and a converging GPG, women continue to earn considerably less than men; about 20.0% in the United States (US) and 16.0% in Europe in 2014, respectively.[1] The pay gap between men and women also varies significantly across sectors, occupations and countries (Melly, 2005; Blau and Kahn, 2006; Lucifora and Meurs, 2006; Arulampalam et al., 2007). In fact, gender differences in pay are a pervasive feature of modern labor markets and exist in various subsamples.

In order to understand the drivers of the GPG, the gap is decomposed in different parts. The standard approach in decomposing mean pay differences between groups is the Oaxaca (1973) and Blinder (1973) method that allows to decompose the wage gap into an explained

[1] The data is taken from the United States Department of Labor (2017) for the US and from Eurostat (2017b) for Europe. See also Blau and Kahn (2003), where the pay gap between men and women in several countries is compared from 1985-1994.

(due to group differences in observable characteristics) and an unexplained part (due to group differences in coefficients). Despite the popularity of the method, it suffers from the index-number problem, i.e. the problem that the choice of the non-discriminatory wage structure influences the estimation results (Reimers, 1983; Cotton, 1988; Neumark, 1988; Oaxaca and Ransom, 1994).[2] The latter is a main pitfall of the Oaxaca-Blinder decomposition model. Various studies have proposed solutions to this problem. The solutions can be divided in two main categories: methods attributing different weights to the various groups (Reimers, 1983; Cotton, 1988) and the intercept-shift approach (Fortin, 2008; Elder et al. 2010). Yet, none of the approaches provides a completely satisfactory solution to the problem (see Lee, 2015, for an overview). Regardless of this problem, the Oaxaca-Blinder approach is still used in academic research (e.g. Mandel and Semyonov, 2014; Juhn and McCue, 2017). One reason for this is its intuitive decomposition (in an explained and an unexplained part; Fortin et al., 2011). The unexplained part, often referred to as discrimination, is identified as the main driver of the wage gap. This result holds across various countries (Blau and Kahn, 1992; 2003) as well as over time (Goldin, 2014). There are also other factors that influence the GPG (in a specific year or time period). Differences in human capital and labor market characteristics, for example, are accounted for by the explained component, while industrial and occupational segregation as well as the persistence of gender stereotypes in the labor market enter in the unexplained part of the pay gap. Similarly, productivity differences are absorbed by this effect (Blau and Kahn, 2006). Consequently, even though the unexplained component of the GPG is generally attributed to gender discrimination, gender differences in unobservables influence this term as well. Hence, it is pivotal to explicitly account for unobservables in the estimation of the GPG and its components. Individual selection in the labor market, in employment or in a specific industry or occupation may be based on individual-specific motivation, ability and productivity. Moreover, the respective selection processes may be different for men and women. For example, women are more likely to work in family-friendly occupations such as jobs with flexible working times or child-care provision (Sorensen, 1989; Waldfogel, 1997; Centra and Cutillo, 2009). It is thus important to account for selection processes into wage work when estimating gender differences in pay. An extensive part of the literature has shown that individual heterogeneity, due to differences in the decision to participate in the labor force or to accept a job offer, has a significant effect on wages (e.g. Heckman, 1979; Sorensen, 1989; Bar et al., 2015). As GPGs differ across subsamples, additional selection procedures matter, such as the choice of a specific job. In order to model both selection decisions, double selectivity approaches are applied (Tunali, 1986; Sorensen, 1989; Dolton and Vignoles, 2000;

[2]Further information about problems of the standard Oaxaca-Blinder model are provided in Chapter 5.

Cutillo and Di Pietro, 2006; Baffoe-Bonnie, 2009). An example of such models are models with 'partial partial observability' by the definition of Meng and Schmidt (1985). These models account simultaneously for sample selection or endogeneity of individual choices (e.g. the participation and job decision). Thereby, unobserved characteristics or unobserved heterogeneity are included in the model (Tunali, 1986; Dolton and Vignoles, 2000). The GPG can then be consistently estimated in the specific subsample (Meng and Schmidt, 1985).

Another body of literature focuses on gender stereotypes in the labor market (for example Becker, 1985; Castagnetti and Rosti, 2013). It is a well-known result of this literature that gender stereotypes contribute to the wage gap among men and women (e.g. Becker, 1985). In particular, significant differences in the assessment of candidates arise from unconscious discriminatory behavior (Lindzey et al., 1998; Schein, 2007). This implies that an identical curriculum is evaluated differently if attributed to a woman compared to a man (Valian, 1998). The impact of stereotypes can be reduced by the use of specific screening devices (such as educational attainment, aptitude tests or letters of recommendation as well as competitive mechanisms ranking the applicants; Castagnetti and Rosti, 2013). Open competition is thereby identified as the most effective screening device for weakening the effects of stereotypes. In Chapter 2 (joint work with Carolina Castagnetti and Luisa Rosti), we use public contests as a special recruitment method of open competition. The analysis focuses on Italy as public contests are particularly regulated by Italian constitutional law and used in both the public and the private sector. Public contests are based on objective criteria and require the employer to provide justification for his or her hiring decision. This allows to eliminate GPGs in the labor market given certain institutional characteristics. Contrary, wages between men and women that are not hired by public contest significantly differ in favor of men. In particular, the GPG is found to be mainly due to the unexplained part and hence gender discrimination may play a role. In the sample of young individuals (18–34 years) recruited by public contest, average earnings are even higher for women than for men. This reversal is driven by the explained part of the gap. Overall, the results suggest that public contests are merit-based and gender-fair instruments for the evaluation of applicants. By using a double selectivity model, we explicitly account for selection into employment and public contests and obtain consistent parameter estimates. The main analysis is conducted on both a panel data set of ten years as well as on the single cross sections. The estimation results are robust over these ten years.

Another phenomenon of modern labor markets is overeducation, i.e. individuals possess higher levels of educational attainment than actually required for their current job (Hartog, 2000). Understanding the reasons of differences in earnings related to overeducation is relevant for policy evaluation as overeducation may implicate a waste of human capital (Flisi et al., 2014). Higher investment in schooling is a loss of resources if both overeducated and properly educated individuals have equal or similar levels of productivity and motivation. However, if there are significant differences between over- and properly educated individuals due to differences in both unobservables and observables, then there is no waste of resources. In Chapter 3 (joint work with Carolina Castagnetti and Luisa Rosti), we estimate the GPG separately for overeducated and properly educated workers using Italian microdata. The study of overeducation and gender differences in pay is particularly interesting for Italy, where, on the one side, the share of individuals with tertiary education is comparably low.[3] On the other side, the problem of overeducation is particularly pronounced in Italy. In Italy, 40.0% of all labor market participants with a university degree are overeducated compared to 21.5% at European Union (EU) level and 26.3% in the US (Cutillo and Di Pietro (2006) and Meliciani and Radicchia (2016) for Italy and Groot and Maassen van den Brink (2000) for the EU and the US). Additionally, the share of female graduates in Italy is relatively high. The latter underlines the importance to consider both the GPG and the effects of overeducation. Our approach presents a novelty in the literature as it integrates insights from two usually separate research fields: gender differences in pay and the effects of overeducation on wages. Up to now, the literature has mainly focused on wage differences between properly and overeducated individuals and detects larger pay gaps between over- compared to properly educated workers (see e.g. Cutillo and Di Pietro, 2006). We find that the GPG is significantly higher among overeducated individuals. In particular, we correct for labor market participation and the overeducation decision, and thus consistently estimate the components of the respective wage gap. We find that the discriminatory part of the gap vanishes in the corrected model. This implies that properly educated individuals would earn more than an overeducated individual with similar characteristics even if the latter was properly matched.

In gender economics, empirical work has recently focused on the estimation beyond the mean (Albrecht et al., 2003; Machado and Mata, 2005; Melly, 2005; Arulampalam et al., 2007; Firpo et al., 2009). The literature identified significant variation in the GPG at different points of the earnings distribution (e.g. Albrecht et al., 2003; Arulampalam et al., 2007). Particularly pronounced wage gaps at the top (glass ceiling) and in fewer cases also at the bottom (sticky floors) of the wage distribution are found. Policy makers are interested in the effects of a policy on different points of the earnings distribution in order

[3]Italy has, with approximately 20.0%, the lowest share of university graduates in Europe (Meliciani and Radicchia, 2016).

to address potential inequality or distributional effects. The latter cannot be detected in the case of the mean analysis (Fortin et al., 2011). The main part of the literature going beyond the mean concentrates on Conditional Quantile Regressions (CQRs). However, Unconditional Quantile Regressions (UQRs) may be preferable over CQRs as they allow for the unconditional mean interpretation and offer a way to conduct path-independent Oaxaca-Blinder type decompositions across the wage distribution (Fortin et al., 2011). The traditional Oaxaca-Blinder decomposition is the benchmark in applied labor economics, when it comes to decomposing mean differences by groups. The approach uses the unconditional mean interpretation and offers an intuitive way to interpret the different components of the wage gap. The detailed Oaxaca-Blinder model allows to attribute to each covariate a specific part of the wage gap. Decompositions based on CQRs are, particularly in the case of detailed decompositions, computationally heavy (Fortin et al., 2011). This is another disadvantage of CQR models. As stated, selection issues may be important in the estimation of the GPG and its components. However, only a comparably small part of the literature beyond the mean accounts for sample selection issues (Buchinsky, 1998; Albrecht et al, 2009). Studies considering wage differences between men and women across the wage distribution and accounting for individual participation or employment decisions find that selection processes vary across quantiles (Albrecht et al., 2009; Chzhen and Mumford, 2011). This result is important for policy makers as not only different policy actions may be required across the wage distribution but also adjusted and unadjusted coefficient estimates may be considerably different. However, the approaches considering selection effects beyond the mean are based on CQRs. In Chapter 4 (single authored), parametric UQRs are used to conduct a detailed decomposition of the GPG at different quantiles (Firpo et al., 2009). The approach is extended such that it can be accounted for quantile-specific employment selection. The selection correction is estimated based on semiparametric methods (Ichimura, 1993; Klein and Spady, 1993). The proposed model offers two methodological advances. First, the more intuitive UQR model or the RIF-OLS model is used and corrected for sample selection. Second, the adjusted UQR model is based on selection processes that do not require any distributional assumptions (Martins, 2001).

So far, sample selection issues at the mean as well as beyond were discussed. Another important body of literature is, as stated, concerned with the decline of the GPG over time. In particular, the unexplained part (attributed to gender discrimination) was found to decrease over time, while the educational level of females was found to increase (e.g. Blau and Kahn, 2006). In Chapter 5 (joint work with Carolina Castagnetti and Luisa Rosti), the convergence of the GPG over the last decade in Italy is examined. An alternative estimation approach is proposed allowing to directly estimate the change of the GPG as well as of its components over time. Up to now, studies examining changes in the wage gap over time and between groups do not directly estimate the difference of the GPG but rather compare ex post the results of the differentials in the corresponding subsamples (Christofides and Michael, 2013; Mandel and Semyonov, 2014). Moreover, statistical inference is often not possible (as no standard errors are provided). Despite the empirical application of the proposed model of the GPG over time, we also estimate the change in the Public-Private Sector Wage Gap (PPWG) between men and women. Sectoral differences in pay are significant and more pronounced for women than for men (Lucifora and Meurs, 2006; Mandel and Semyonov, 2014). The literature identifies a higher unexplained part of the PPWG for women than for men. This effect is more pronounced in the private sector compared to the public sector. Our model confirms that the unexplained part significantly varies between men and women. In line with the results from Chapter 2, institutional differences between the public and the private sector are found to matter as well. In both empirical applications, we estimate the standard procedure as well as our suggested model and compare the results of the two approaches. All in all, the proposed estimation model allows to draw conclusions of changes in mean group differences in wages. In particular, the insights may differ if conclusions are drawn by simply comparing the results for different subsamples. Additionally, we show that our model can be made robust to the index-number problem of the standard Oaxaca-Blinder decomposition (Oaxaca and Ransom, 1994) and the indeterminacy problem of the intercept-shift approach (Lee, 2015).

The empirical part of this thesis focuses on Italy. Italy is particularly interesting for the study of selection into labor force participation or employment as female labor force participation is relatively low in Italy. The literature generally finds that countries with low GPGs have high employment gaps between men and women (Olivetti and Petrongolo, 2008). The GPG in Italy amounted to 5.6% between 2005-2014, while it was equal to 16.8% at EU-28 level.[4] In the same period, 49.8% of all labor market participants in Italy were female, in comparison to 62.1% in the EU. This corresponds to gender employment gaps of 23.3% in Italy and 13.5% at EU-level.[5] The data set used is the Participation, Labor, Unemployment Survey (PLUS) from the Italian Institute for the Development of Vocational Training for Workers (ISFOL). The data was collected in the context of a joint project of ISFOL and the Italian Ministry of Labor and Social Policy that started in 2005. Up to now, the survey was released for the following years with a longitudinal structure: 2005, 2006,

[4]The rates are calculated using data from Eurostat (2017a).

[5]The rates are calculated using data from Eurostat (2017a).

2008, 2010, 2011 and 2014. In this work, the entire release of panel and cross-sectional dimension is used. The survey is particularly designed for the study of wage inequality and includes detailed information on the personal working profiles and the intervieews' motivation to work as well as on the cultural and territorial background of the participants (Mandrone, 2006; Centra and Cutillo, 2009). The survey uses subjective measures only and was conducted by Computer Assisted Telephone Interviewing (CATI). In ISFOL PLUS, the individuals self-select themselves in certain conditions such as whether being active or inactive in the labor market, while in other data sets the individuals' labor market status is often identified via characteristics such as having worked at least one hour during the last week. This characteristic of ISFOL PLUS allows us to identify a homogeneous group of voluntarily unemployed individuals (Chapter 2 and 4) as well as individuals (voluntarily) out of the labor force (Chapter 3). Studies examining selection issues often cannot distinguish between voluntary and involuntary unemployment or voluntary and involuntary absence from the labor market (Heinze et al., 2003). Case-specific sample restrictions are discussed in the corresponding Chapters.

This thesis adds to literature on selection effects on wages and shows that individual participation or employment decisions matter at the mean as well as across the earnings distribution. It is, to the author's best knowledge, the first work that shows the empirical disappearance of the GPG and combines the literature on gender pay differences and overeducation. In particular, it provides a model to adjust the wage equation, the GPG and its components at different quantiles using UQRs. Additionally, this work examines gender differences in pay over time by proposing a novel approach to model changes in the wage gap over time.

# References


Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a Glass Ceiling in Sweden?. *Journal of Labor Economics*, 21(1), 145-177.

Albrecht, J., van Vuuren, A., and Vroman, S. (2009). Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands. *Labour Economics*, 16(4), 383-396.

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2007). Is there a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial and Labor Relations Review*, 60(2), 163-186.

Baffoe-Bonnie, J. (2009). Black–White Wage Differentials in a Multiple Sample Selection Bias Model. *Atlantic Economic Journal*, 37(1), 1-16.

Bar, M., Kim, S., and Leukhina, O. (2015). Gender Wage Gap Accounting: The Role of Selection Bias. *Demography*, 52(5), 1729-1750.

Becker, G. S. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics*, 3(1, Part 2), 33-58.

Black, S. E., and Spitz-Oener, A. (2010). Explaining Women's Success: Technological Change and the Skill Content of Women's Work. *The Review of Economics and Statistics*, 92(1), 187-194.

Blau, F. D., and Kahn, L. M. (1992). The Gender Earnings Gap: Learning from International Comparisons. *The American Economic Review*, 82(2), 533-538.

Blau, F. D., and Kahn, L. M. (2003). Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1), 106-144.

Blau, F. D., and Kahn, L. M. (2006). The US Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4), 436-455.

Buchinsky, M. (1998). The Dynamics of Change in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*, 1-30.

Castagnetti, C., and Rosti, L. (2013). Unfair Tournaments Gender Stereotyping and Wage Discrimination among Italian Graduates. *Gender and Society*, 27(5), 630-658.

Centra, M., and Cutillo, A. (2009). *Differenziale Salariale di Genere e Lavori Tipicamente Femminili*. Studi Isfol, 2.

Christofides, L. N., and Michael, M. (2013). Exploring the Public-Private Sector Wage Gap in European Countries. *IZA Journal of European Labor Studies*, 2(1), 1-53.

Chzhen, Y., and Mumford, K. (2011). Gender Gaps Across the Earnings Distribution for Full-Time Employees in Britain: Allowing for Sample Selection. *Labour Economics*, 18(6), 837-844.

Cotter, D. A., Hermsen, J. M., and Vanneman, R. (2004). G*ender Inequality at Work*. Russell Sage Foundation.

Cotton, J. (1988). On the Decomposition of Wage Differentials. *The Review of Economics and Statistics*, 70(2), 236-243.

Cutillo, A., and Di Pietro, G. (2006). The Effects of Overeducation on Wages in Italy: A Bivariate Selectivity Approach. *International Journal of Manpower*, 27(2), 143-168.

Dolton, P., and Vignoles, A. (2000). The Incidence and Effects of Overeducation in the U.K. Graduate Labour Market. *Economics of Education Review*, 19(2), 179-198.

Elder, T. E., Goddeeris, J. H., and Haider, S. J. (2010). Unexplained Gaps and Oaxaca–Blinder Decompositions. *Labour Economics*. 17(1), 284-290.

England, P. (2006). Toward Gender Equality: Progress and Bottlenecks. In Blau F., Brinton M., and Grusky D. (Eds.), *The Declining Significance of Gender?* (pp. 245-264). Russell Sage Foundation.

Eurostat (2017a). *Employment and Activity by Sex and Age - Annual Data*. http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do (accessed 30-06-2017).

Eurostat (2017b). *Gender Pay Gap in Unadjusted Form.* http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (accessed 27-05-2017).

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3), 953-973.

Flisi, S., Goglio, V., Meroni, E., Rodrigues, M., and Vera-Toscano, E. (2014). Occupational Mismatch in Europe: Understanding Overeducation and Overskilling for Policy Making. *European Commission, JRC Science and Policy Reports, Luxembourg*.

Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources*, 43(4), 884-918.

Fortin, N. M., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. *Handbook of Labor Economics*, 4, 1-102.

Fortin, N. M. (2015). Gender Role Attitudes and Women's Labor Market Participation: Opting-out AIDS and the Persistent Appeal of Housewifery. *Annals of Economics and Statistics*, (117/118), 379-401.

Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. *American Economic Review*, 96(2), 1-21.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091-1119.

Groot, W., and Maassen van den Brink, H. (2000). Overeducation in the Labor Market: A Meta-Analysis. *Economics of Education Review*, 19(2), 149-158.

Hartog, J. (2000). Over-education and Earnings: Where are We, Where Should We Go?. *Economics of Education Review*, 19(2), 131-147.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Heinze, A., Beninger, D., Beblo, M., and Laisney, F. (2003). *Measuring Selectivity-Corrected Gender Wage Gaps in the EU* (No. 03-74). ZEW Discussion Papers.

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single- Index Models. *Journal of Econometrics*, 58(1-2), 71-120.

Juhn, C., and McCue, K. (2017). Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts. *The Journal of Economic Perspectives*, 31(1), 183-204.

Klein, R. W., and Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*, 61(2), 387-421.

Lee, M. J. (2015). Reference Parameters in Blinder-Oaxaca Decomposition: Pooled-Sample versus Intercept-Shift Approaches. *Journal of Economic Inequality*, 13(1), 69-82.

Lindzey, G., Gilbert, D. T., and Fiske, S. T. (1998). *The Handbook of Social Psychology*. Vol. 2. McGraw-Hill.

Lucifora, C., and Meurs, D. (2006). The Public Sector Pay Gap In France, Great Britain and Italy. *Review of Income and Wealth*, 52(1), 43-59.

Machado, J. A., and Mata, J. (2005). Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. *Journal of Applied Econometrics*, 20(4), 445-465.

Mandel, H., and Semyonov, M. (2014). Gender Pay Gap and Employment Sector: Sources of Earnings Disparities in the United States, 1970-2010. *Demography*, 51(5), 1597-1618.

Mandrone, E. (2006). *La Riclassificazione del Lavoro tra Occupazione Standard e Atipica: L'Indagine Isfol Plus 2006*, Studi Isfol Plus.

Martins, M. F. O. (2001). Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics*, 16(1), 23-39.

Meliciani, V., and Radicchia, D. (2016). Informal Networks, Spatial Mobility and Overeducation in the Italian Labour Market. *The Annals of Regional Science*, 56(2), 513-535.

Melly, B. (2005). Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression. *Empirical Economics*, 30(2), 505-520.

Meng, C., Schmidt, P. (1985). On the Cost of Partial Observability in the Bivariate Probit Model. *International Economic Review*, 26(1), 71-85.

Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *Journal of Human Resources*, 23(3), 279-295.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

Oaxaca R. and Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61(1), 5-21.

Olivetti, C. and Petrongolo, B. (2008). Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26(4), 621-654.

Reimers, C. W. (1983). Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics*, 65(4), 570-579.

Schein, V. E. (2007). Women in Management: Reflections and Projections. *Women in Management Review*, 22(1), 6-18.

Sorensen, E. (1989). Measuring the Pay Disparity between typically Female Occupations and other Jobs: A Bivariate Selectivity Approach. *Industrial and Labor Relations Review*, 42(4), 624-639.

Tunali, I. (1986). A General Structure for Models of Double-Selection and an Application to a joint Migration/Earnings Process with Remigration. *Research in Labor Economics*, 8 (Part B), 235-282.

United States Department of Labor (2017). *Annual Earnings by Sex, 1960-2014*. https://www.dol.gov/wb/stats/earnings_2014.htm (accessed 06-07-2017).

Valian, V. (1998). Sex, Schemas, and Success: What's Keeping Women Back?. *Academe*, 84(5), 50-55.

Waldfogel, J. (1997). The Effect of Children on Women's Wages. *American Sociological Review*, 62(2), 209-217.

# Chapter 2

# Discriminate Me – if You Can! The Disappearance of the Gender Pay Gap among Public-Contest Selected Employees

## 2.1 Introduction

There is a huge literature on the Gender Pay Gap (GPG) and on its narrowing in recent years (e.g. Blau and Kahn, 2003; 2006; 2007; 2016; Goldin, 2014; Kahn, 2015). Nonetheless, women still earn considerably less than men; about 20.0% in the United States of America and 15.0% in Europe.[1] Despite the empirical finding that the difference in pay between men and women has decreased in the last decades, a consistent part of the GPG remains unexplained and this part has not declined over time but has been roughly stable over the past 30 years (Blau and Kahn, 2016). The unexplained GPG, i.e. the portion of the GPG not due to gender differences in observed characteristics, is generally taken to be an estimate of gender discrimination[2] and may be influenced by cognitive processes such as stereotyping. The literature on gender stereotypes finds that systematic errors in screening and performance assessment of applicants arise from unconscious discriminatory behavior (Lindzey et al., 1998; Schein, 2007). This literature is relevant for the analysis of the GPG as it highlights

[1] See Kahn (2015), where the GPG in several countries is compared from 2010 to 2012.

[2] However, as stressed by Blau and Kahn (2006), the unexplained portion of the GPG may include effects of unobserved productivity differentials.

how, because of stereotypes, an identical curriculum is evaluated in a substantially different way if attributed to a woman instead of to a man (Valian, 1998). Based on findings from social psychologists that discriminatory attitudes and stereotyping may be unconscious and therefore difficult to detect and erase, Blau and Kahn (2016) argue that as gender discrimination has become less socially acceptable, it has become less overt and more subtle as well as unconscious. Castagnetti and Rosti (2013) identify specific environments in which the use of stereotypes is expected to be more likely to exert an influence on screening devices[3] and show that the unexplained component of the GPG increases in line with the expected influence of stereotypes. Open competition is thereby identified as a screening device that weakens or even deletes the impact of stereotypes on performance appraisal in the labor market. Public contests are a special recruitment method of open competition. They increase the accuracy of assessment as they require the use of objective criteria and justification of the candidate choice (Dobbs and Crano, 2001). This allows to increase the probability of fair assessment for both men and women compared to other recruitment methods.[4] The aim of this paper is to show that the selection procedure of public contests may counteract the discrimination mechanisms in the hiring process. The results suggest that public contests are merit-based and gender-fair instruments for the evaluation of applicants. The paper focuses on Italian public contests as their implementation is strictly regulated by Italian constitutional law. Italian public contests are based on clear rules that are difficult to distort. The analysis of the Italian case is of particular interest as public contests are the major way of recruitment in the public sector. In particular, the Italian Constitution states that in Italy every public employee shall be recruited following open competition procedures. Recruitment not by public contest in the public sector is only possible if explicitly regulated by law.[5] Due to this regulation, there are also employees in the public sector not selected by public contest. Legislative acts regulating recruitment in public employment by mechanisms different from public contests were adopted by a wave of reforms starting in the 1980's. These reforms aimed to reorganize public-sector employment in Italy (Carinci et al., 2002).[6] Even

[3]Screening devices include for example educational attainment, aptitude tests or letters of recommendation as well as competitive mechanisms ranking the applicants.

[4]Dobbs and Crano (2001) argue that individuals who have to justify their decisions have a stronger incentive to bypass their stereotyped impressions than those that do not have to provide justifications. As a consequence, when decision makers are required to justify their choices and describe the criteria they use to evaluate candidates, as in open competition, they are less likely to discriminate against a specific group.

[5]In general, in other European countries, only higher public servants are recruited through open competition, yet with different legal constraints compared to the Italian public contest. Other public-sector recruitment takes place via private methods of recruitment such as candidate submittal or agency recruitment. The latter is the main procedure used in the private sector. See Cardona (2006) for a detailed description of the recruitment in civil service systems in Europe.

[6]Examples are 'ope legis' promotions and hiring of 'pro tempore' workers.

though public contests are the main recruitment method in the public sector, public-contest recruitment takes also place in the private sector. However, public contests in the public or private sector may evolve differently, as public-sector employees enjoy different institutional norms compared to private-sector employees. In public employment, for example pay levels are largely immutable and jobs are highly stable in order to guarantee the efficient exercise of public functions (Carinci et al., 2002). Therefore, as a robustness check, we analyze the GPG for public-contest and not public-contest selected employees separately for the public and private sector.

The underlying study uses the survey ISFOL PLUS of the Italian labor market over the period 2005–2014. The survey was conducted by Isfol. The empirical strategy relies on the estimation and comparison of the GPG between two groups of employees which differ by hiring method (i.e. recruitment by public contest or not). The GPG is estimated using the Oaxaca (1973) and Blinder (1973) decomposition. Additionally, we account for double selection into the sample (considering selection in employment as well as selection in public-contest recruitment). The double selection model is applied in order to detect selection differences by recruitment group and gender as well as to account for otherwise unobservable characteristics of the individuals in the sample.

The main findings of the paper can be summarized as follows. First, we find that the GPG vanishes among public-contest selected employees and even reverses in favor of women in the sample of young employees. Moreover, the reversal is entirely due to the explained component of the decomposition. To the best of our knowledge, this is the first work that shows the empirical disappearance and reversal of the GPG. In addition, the result is robust over time as it holds for the period of ten years considered in this paper. As a second result, we show that recruitment through public contest has a sizeable positive effect on wages (stronger for women than for men). This finding is in line with empirical evidence of a public-sector wage premium in Italy given that about 90.0% of the individuals in our sample hired by public contest work in the public sector. Both selection bias channels (i.e. the employment and recruitment decision) are indeed relevant for explaining the GPG. Given more equal and less discriminatory assessment of candidates, women may self-select themselves in public-contest recruitment (endogeneity bias). The work decision depends on individual heterogeneity and unobservable characteristics (sample selection bias). Therefore, it is crucial to account for self-selection deriving from both biases. Thanks to the detailed questionnaire underlying the data we use, we are able to identify instruments for the employment selection as well as for the recruitment selection. The analysis accounting for double selection suggests that public contests are merit-based selection methods. Individuals (both men and women) recruited by public contest possess more favorable unobserved characteristics (and wages) than individuals not recruited by public contest. Moreover, the Counterfactual Pay Gap (CPG) adjusted for double selection shows that in the case of recruitment without public contest the unexplained component is (still) highly statistically significant and positive. The results show that public contests are gender-fair and merit-based mechanisms for applicants' evaluation. Public contests are merit-based because public-contest selected employees hold better productive characteristics than unselected ones. They are gender-fair because among public-contest selected employees women's characteristics are rewarded as men's. We show that, even though public contests are not entirely driven by the institutional framework of the public sector, the institutional environment plays an important role in making public contests effective mechanisms for gender-fair and merit-based recruitment.

The paper is organized as follows. Section 2.2 describes the data. Section 2.3 shows the effect of public contests on log hourly wages. Section 2.4 provides evidence on public contests as gender-fair selection methods. Section 2.5 extends the analysis to a double sample selection model, thereby accounting for sample selection and endogeneity problems. In Section 2.6, the counterfactual GPG adjusted for selectivity bias is computed. As a robustness check, we compute in Section 2.7 the GPG by public-contest recruitment separately for the public and private sector. Section 2.8 concludes.

## 2.2 Data and Descriptive Statistics

The empirical analysis is based on microdata collected by the Italian Institute for the Development of Vocational Training for Workers (ISFOL) in the Participation, Labor, Unemployment Survey (PLUS). ISFOL released up to now the following data waves with a longitudinal structure; 2005, 2006, 2008, 2010, 2011 and 2014. The empirical analysis is conducted by exploiting both the cross-section and panel dimension of the data set. In particular, the panel data set includes all individuals that have been interviewed for at least two periods. However, given that the focus is mainly on the impact of individual variables constant across time as being hired by public contest[7] and that part of the analysis relies on the Oaxaca-Blinder decomposition model of the GPG or an extension of it, we base our estimates on a pooled Ordinary Least Squares (OLS) regression model. In the panel estimation, we include wave or year dummies as explanatory variables.

In our analysis, we focus on full-time employees aged between 18 and 64 years. Part-time workers are excluded from the sample as they have a larger dispersion in pay than their full-

[7]The number of transitions in and out of being hired by public contest is very low (about 1.0%).

time colleagues that may raise the probability of earning less than the average hourly wage. Moreover, the incidence of part-time work differs significantly between men and women in favor of women (e.g. Chzhen and Mumford, 2011). Similarly, self-employed workers are not considered in the study, as the focus in this paper is employees' selection mechanisms, but self-employed are unselected or, if selection takes place it serves as an entrance examination. An example are notaries, where the main aim pursued is not to fill job vacancies but to ensure the citizens on the quality of the services provided. The analysis is also constrained to earnings from the main job only, i.e. from the job that yields the highest income. As only 1.0-2.0% of the individuals in the sample have more than one job, the impact of this restriction should be negligible. Similarly, we exclude all individuals with disabilities (less than 2.0% of all observations). The sample is further restricted by excluding students and pensioners as well as individuals not disposable to work or involuntarily unemployed. This restriction is justified by the aim to form a homogeneous sample of employed individuals and individuals that are voluntarily out of work (Heinze et al., 2003). In the sample, individuals not in wage work are individuals indicating not to accept every job offer but only job offers in line with their characteristics (such as the level of educational attainment or labor market experience). Thus, in the sample all individuals out of employment are voluntarily out of work or in search unemployment. Consequently, the employment decision considered in this paper consists in the decision between voluntary or search unemployment and employment. We do not take into account the selectivity bias stemming from labor market participation but the bias deriving from search unemployment. We are aware of the fact that the selectivity bias from the labor force participation may be important for Italy given a comparably low female labor force participation rate in Italy (Olivetti and Petrongolo, 2008; Centra and Cutillo, 2009). However, as this participation bias is well known for the Italian case, we focus here on search unemployment that – similar to labor market participation – may be significantly different for men and women and particularly relevant for Italy.[8]

We delete observations with missing values on other variables used in the analysis. Then we are left with a sample size of 72,928 for the panel and 109,172 for the cross sections. In the panel, 39,345 are female (54.0%) and 33,538 are male employees (46.0%) (in the cross sections; 59,406 women and 49,766 men). Out of this sample 41,845 (58,151) individuals are employed: 19,398 are female (26,956) and 22,447 (31,195) are male employees in the panel data (and the cross sections, respectively). In the panel data, 6,798 male (45.6%) and 8,116 female employees (54.4%) entered via public contest in their current job. In the

[8]The observation of a positive wage may depend either on the decision of the employee to accept a job offer or not, or on the firm decision to hire the candidate or not (Nicaise, 2001; Baffoe-Bonnie, 2009). We assume that the selection into employment depends only on the individual decision and not on the firm decision.

cross-section dimension 9,255 men (45.2%) and 11,230 women (54.8%) were recruited by means of public contest. We use log hourly wages as dependent variable. The basic hourly wage rate is defined by the net monthly wage perceived and divided by the number of actual working hours per month.[9] A complete list of variables included in the analysis along with the corresponding definition and coding is provided in Appendix 2.A.

Table 2.1 for the panel and Table 2.B.1 for the cross sections report means and standard deviations for some of the variables considered in the analysis. On average, more women are recruited via public contests than men (*Female*). Generally, employees hired by public contest have higher educational attainment (*Educ*) and hence graduated more often from university (*University_Degree*).[10] Moreover, they have more often obtained the maximum grade when holding a university degree (*Max_D_Mark*). On average, public-contest selected employees have more experience (*Exper*) as well as job tenure (*Tenure*). Individuals hired by public contest are on average more often married (*Married*) and have more often children (*Kids*) as well as young children (*Kids_10*). Public-contest selected employees are on average more than ten years older than employees not selected by public contest. Our data show that the selection by public contest is not a prerogative of the public sector; about 9.0% of the recruitment in the private sector takes place by public contest. Similarly, approximately 16.0% of the observed individuals employed in the public sector are not hired by public contest. More than 90.0% of the employees hired by public contest have an unlimited contract, while on average only 80.0% of employees not hired by public contest have an unlimited contract (*Contract_Type*). Finally, public-contest selected employees are more often occupied in highly specialized and intellectual occupations (*Manager*), while individuals not hired by public contest are slightly more often engaged in intermediary positions (*Intermed_Prof*).

---

[9] The survey includes monthly gross earnings, which however contain for more than 98.0% of all individuals in the data missing values. Alternatively, using gross annual earnings would be possible. Yet, when dividing gross annual earnings by the number of months in a calendar year plus one (in order to account for a potential 13$^{\text{th}}$ income), the difference between the artificially created monthly gross income and the reported monthly gross income amounts on average to more than 800 Euros per month. Therefore, we prefer to use the reported monthly net income. Individuals with children obtain tax credits in Italy – given an annual gross income below 95,000 Euro (see for example Worldwide Tax Summaries, 2017, for further details). The tax credit is rated on a monthly basis but depends on the annual gross income and is granted yearly. Consequently, the children-related tax credit does not impact on the monthly perceived net income and does not directly affect monthly net wages in Italy.

[10] The dummy variable *University_Degree* is presented only for illustration of the variation of university graduation in the data and not included in the regressions as the effect of having a university degree on earnings is already captured by the years of schooling completed *Educ*.

Table 2.1 Descriptive Statistics – Panel

| Year | Panel | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Public-Contest Selected Employees | | Not Public-Contest Selected Employees | |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. |
| Female | 0.544 | 0.498 | 0.419 | 0.493 |
| Educ | 14.043 | 2.218 | 12.573 | 2.812 |
| University_Degree | 0.464 | 0.499 | 0.231 | 0.422 |
| Max_D_Mark | 0.094 | 0.292 | 0.035 | 0.182 |
| Exper | 25.500 | 10.839 | 16.627 | 12.978 |
| Tenure | 20.684 | 11.248 | 10.860 | 11.047 |
| Married | 0.743 | 0.437 | 0.466 | 0.499 |
| Kids | 0.758 | 0.429 | 0.480 | 0.500 |
| Kids_10 | 0.300 | 0.458 | 0.261 | 0.439 |
| Age | 48.713 | 10.092 | 37.317 | 12.484 |
| Public_Sector | 0.912 | 0.284 | 0.163 | 0.370 |
| Contract_Type | 0.927 | 0.261 | 0.778 | 0.416 |
| Manager | 0.387 | 0.487 | 0.147 | 0.355 |
| Intermed_Prof | 0.496 | 0.500 | 0.455 | 0.498 |
| Observations | 14,914 | | 26,931 | |

## 2.3 The Effect of Public-Contest Selection on Earnings

The unadjusted GPG[11] is a key indicator used within the European employment strategy to monitor imbalances in wages between men and women. The Eurostat data show that in the period considered, 2005-2014, the GPG is estimated to be on average 16.8% in the EU-27[12] as a whole and 5.6% in Italy.[13] In our data the gender gap in hourly wages among full-time

[11]"The unadjusted gender pay gap provides an overall picture of gender inequality in hourly pay. This gap represents the difference between the average [...] hourly earnings of men and women expressed as a percentage of average [...] hourly earnings of men. It is called unadjusted as it does not take into account all of the factors that influence the gender pay gap, such as differences in education, labor market experience or type of job" (Eurostat, 2016).

[12]EU-27 include: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden and the United Kingdom.

[13]According to Eurostat (2017), the GPG in Italy varies between 4.4% and 7.3% in the period considered. The GPG indicator is calculated using the Structure of Earnings Survey - NACE Rev. 2. The population consists of all paid employees in enterprises with 10 employees or more.

employees is 6.6% for the panel data set and varies between 10.1% and 3.1% for the cross sections (see Table 2.2).

Table 2.2 GPG of Net Hourly Wages

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| GPG in % | 6.60 | 10.10 | 8.18 | 7.10 | 5.66 | 3.06 | 3.85 |
| Observations | 41,845 | 9,520 | 9,305 | 8,601 | 9,420 | 8,977 | 12,328 |

*Source:* Own elaboration on ISFOL PLUS.

A small GPG in hourly wage does not imply a thin overall income inequality between men and women within the economy. When considering the annual income instead of the hourly wage, the differential increases significantly due to a lower number of hours worked by female employees. Moreover, besides the GPG and the gender gap in paid hours, it is important to consider gender gaps in employment, as they also contribute substantially to increases of the difference in average earnings between men and women. In particular, in countries with low female employment rates, women choosing to work may decide to do so due to relatively higher job profiles and earnings expectations. In order to give a complete picture of the GPG, Eurostat has developed a synthetic indicator called Gender Overall Earnings Gap (GOEG). This indicator measures the impact of three combined factors (hourly earnings, hours paid and employment rate) on average earnings of all men of working age compared to all women of working age. Eurostat (2016) estimates the 2010 GOEG to amount to 44.3% in Italy, and to 41.1% for the EU-27. At EU-level, the GOEG was mostly driven by the GPG (contribution of 37.0%), the gender employment gap (contribution of 35.0%) and the gender gap in paid hours (28.0%). In Italy the gender gap in employment rates was the main contributor to the total earnings gap (contribution of 65.0%), followed by the gender gap in paid hours (26.0%) and by the GPG (contribution of 9.0%, see Eurostat, 2016). This result suggests that considering sample selection is particularly relevant for Italy. Even though the GPG in hourly wages is only a part of the overall income inequality by gender in Italy, it is the analysis of that (relatively small) gap which brings out discrimination from the data and drives the gender gap in both paid hours and employment rates.

The paper analyzes the GPG separately for employees recruited by public contest and employees recruited by different methods. The basis for the following analysis is the estimation of a standard Mincer-type wage equation separately for gender and recruitment group augmented by factors such as human capital and job as well as personal and family background characteristics. In order to identify public-contest recruitment, we use the indicator variable *Public_Contest*, which is equal to one if the individual was hired by public contest and zero otherwise. The estimation results show that recruitment carried out by public contest has a positive effect on wages. Entering in employment by public contest has a relevant positive impact on log hourly wages in all data waves (see Table 2.3).[14] The negative and significant coefficient estimate of the variable *Female* confirms the usual result of the literature: being a woman significantly reduces wages. The coefficient estimate of the interaction term *Contfem* being positive and significant shows that female employees receive from public-contest selection a wage premium. In particular, women hired by public contest perceive a wage premium – except for 2005 – at par or even higher than the gender penalty. Hence, the bonus received by female employees for public-contest recruitment outweighs the negative wage effect of being a woman significantly in the majority of the data waves. Table 2.4 shows the effect of public contests on earnings in a Mincer-type wage model. Indeed, recruitment through public contest has a sizeable positive effect on earnings and the dummy *Public_Contest* emerges as one of the most important among the considered variables to predict earnings. In the full sample of individuals aged 18-64, the wage premium for public-contest selection lies between 5.8% and 10.6%. This finding is in line with empirical evidence of a public-sector wage premium in Italy (Cappellari, 2002; Dickson et al., 2014) as the correlation between being hired by public contest in Italy and working in the public sector is very high. About 90.0% of the individuals in our sample hired by public contest are employed in the public sector. Again, being a woman has a significant negative impact on earnings. In our sample, earnings are reduced between 8.3% and 14.1%, all else equal. Positive and significant coefficient estimates of the interaction term *Contfem* show, again, that female employees receive from public-contest employment a wage premium.[15]

Both theoretical literature and empirical evidence on the GPG suggest that small differences in the early career greatly expand with age and give rise to large lifelong observed gender disparities in earnings (Lazear and Rosen, 1990; Blau and Kahn, 2000). This is driven by gender differences in promotion, bargaining and particularly women's absence from the labor market due to childbearing and -care (Blackaby et al., 2005; Niederle and Vesterlund, 2007; Fortin, 2008; Heilman and Okimoto, 2008; Bertrand et al., 2011). As the positive effect of public-contest selection impacts to a greater extent on early wages, we expect to find

[14] Table 2.3 shows the regression output of log hourly wages on the indicator variables *Public_Contest* and *Female* as well as the interaction term *Contfem*. The variable *Contfem* is given by the interaction between the indicator variables *Female* and *Public_Contest*.

[15] The other explanatory variables included in the regression impact on wages as expected. For a full list of the regression output, see Table 2.B.2 for the results of the panel data and Tables 2.B.3–2.B.8 for the cross-sectional data in Appendix 2.B.

a stronger effect of public-contest recruitment among young people by taking the early age as a proxy for the early career. The results presented in Table 2.4 (columns (2), (4) and (6), respectively) confirm that the positive effect on wages of recruitment carried out by public contest is stronger in the early career. Moreover, the positive effect of recruitment through public contest is less volatile and generally higher among young employees: their earnings increase between 7.3% and 14.4% if individuals are selected by public contest (compared to the non selected). The coefficient estimate of the variable *Female* impacts as in the regression on the full sample negatively and statistically significantly on log hourly earnings. The coefficient estimate of the variable *Contfem* is again statistically significant and positive. Hence, being a woman and entering in employment by public contest, all else equal, raises log hourly wages on average also for the young sample. As public contests are assumed to be less discriminatory or discretionary than other private methods of recruitment, they may be preferred by women (all else equal). In line with this, Table 2.B.9 and Tables 2.B.10-2.B.15 show that the positive effect on wages of recruitment carried out by public contest is stronger for women than for men.[16]

Table 2.3 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest*, *Female* and Interactive Effect *Contfem* Only – Full Sample

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Variables | | | | Full Sample | | | |
| Female | -0.099*** | -0.140*** | -0.126*** | -0.106*** | -0.084*** | -0.049*** | -0.065*** |
| | (0.006) | (0.011) | (0.011) | (0.011) | (0.013) | (0.013) | (0.011) |
| Public_Contest | 0.362*** | 0.347*** | 0.356*** | 0.367*** | 0.366*** | 0.367*** | 0.285*** |
| | (0.006) | (0.013) | (0.012) | (0.013) | (0.014) | (0.014) | (0.012) |
| Contfem | 0.092*** | 0.101*** | 0.122*** | 0.100*** | 0.085*** | 0.058*** | 0.074*** |
| | (0.009) | (0.018) | (0.018) | (0.019) | (0.021) | (0.021) | (0.017) |
| Constant | 2.065*** | 1.942*** | 1.958*** | 1.999*** | 2.015*** | 2.016*** | 2.077*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.008) | (0.008) | (0.007) |
| Year Dummies | Yes | No | No | No | No | No | No |
| Observations | 41,845 | 9,520 | 9,305 | 8,601 | 9,420 | 8,977 | 12,328 |
| R-squared | 0.171 | 0.179 | 0.188 | 0.187 | 0.134 | 0.136 | 0.099 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

[16] In all years, except for 2005.

Table 2.4 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| *Panel* | | | | | | |
| Public_Contest | 0.068*** | 0.128*** | | | | |
| | (0.006) | (0.016) | | | | |
| Female | -0.104*** | -0.069*** | -0.064*** | -0.008 | -0.104*** | -0.069*** |
| | (0.005) | (0.008) | (0.006) | (0.021) | (0.005) | (0.008) |
| Contfem | 0.040*** | 0.050*** | | | | |
| | (0.008) | (0.019) | | | | |
| *2005* | | | | | | |
| Public_Contest | 0.062*** | 0.070** | | | | |
| | (0.014) | (0.035) | | | | |
| Female | -0.132*** | -0.138*** | -0.105*** | -0.050 | -0.137*** | -0.139*** |
| | (0.010) | (0.014) | (0.013) | (0.045) | (0.010) | (0.014) |
| Contfem | 0.025 | 0.062 | | | | |
| | (0.015) | (0.040) | | | | |
| *2006* | | | | | | |
| Public_Contest | 0.060*** | 0.102*** | | | | |
| | (0.014) | (0.033) | | | | |
| Female | -0.096*** | -0.067*** | -0.043*** | 0.033 | -0.101*** | -0.070*** |
| | (0.010) | (0.015) | (0.013) | (0.042) | (0.011) | (0.015) |
| Contfem | 0.047*** | 0.059 | | | | |
| | (0.015) | (0.037) | | | | |
| *2008* | | | | | | |
| Public_Contest | 0.056*** | 0.134*** | | | | |
| | (0.013) | (0.032) | | | | |
| Female | -0.111*** | -0.089*** | -0.066*** | -0.014 | -0.115*** | -0.094*** |
| | (0.010) | (0.014) | (0.014) | (0.043) | (0.010) | (0.014) |
| Contfem | 0.037** | 0.027 | | | | |
| | (0.016) | (0.039) | | | | |
| *2010* | | | | | | |
| Public_Contest | 0.076*** | 0.112*** | | | | |
| | (0.015) | (0.040) | | | | |
| Female | -0.093*** | -0.043** | -0.068*** | -0.030 | -0.089*** | -0.040** |
| | (0.012) | (0.018) | (0.016) | (0.053) | (0.013) | (0.018) |
| Contfem | 0.031* | 0.038 | | | | |
| | (0.018) | (0.045) | | | | |
| *2011* | | | | | | |
| Public_Contest | 0.101*** | 0.111*** | | | | |
| | (0.015) | (0.031) | | | | |
| Female | -0.079*** | -0.024 | -0.058*** | 0.108** | -0.080*** | -0.023 |
| | (0.013) | (0.019) | (0.015) | (0.051) | (0.013) | (0.019) |
| Contfem | 0.023 | 0.105*** | | | | |
| | (0.019) | (0.039) | | | | |
| *2014* | | | | | | |
| Public_Contest | 0.072*** | 0.119*** | | | | |
| | (0.013) | (0.030) | | | | |
| Female | -0.087*** | -0.048*** | -0.073*** | -0.027 | -0.085*** | -0.050*** |
| | (0.011) | (0.017) | (0.013) | (0.034) | (0.011) | (0.017) |
| Contfem | 0.014 | 0.019 | | | | |
| | (0.016) | (0.037) | | | | |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

*Notes:* The regressions include the full set of control variables. The full regression output can be found in Table 2.B.2 –2.B.8 in Appendix 2.B. The regression on each data set, panel or cross sections, contains sectoral as well as year or wave dummies.

## 2.4 The GPG by Public-Contest Selection

In the previous Section, we have found evidence that hiring carried out by public contest has a positive effect on earnings that is more pronounced for female and young employees. In this Section we use the standard Oaxaca (1973) and Blinder (1973) methodology to decompose the GPG. We analyze the GPG all else equal as well as the (so-called) discriminatory part of the wage gap for both public-contest recruited employees and not public-contest recruited employees. We assume that public contests, contrary to other private, methods of recruitment are merit-based and gender-fair. Indeed, private recruitment methods are more discretionary and unregulated and hence may create conditions for gender discrimination to flourish. Therefore, we expect that both the GPG as well as the discriminatory part are lower among public-contest selected employees. By using the implicit assumptions in Oaxaca (1973) and Blinder (1973), we decompose the wage differential in two parts; endowments and coefficients:

$$\begin{aligned}\overline{ln(W_M)} - \overline{ln(W_F)} &= \bar{X}_M'\hat{\beta}_M - \bar{X}_F'\hat{\beta}_F \\ &= (\bar{X}_M' - \bar{X}_F')\hat{\beta}_M + \bar{X}_F'(\hat{\beta}_M - \hat{\beta}_F) \qquad (2.1)\end{aligned}$$

where $\overline{ln(W_M)}$ and $\overline{ln(W_F)}$ are the log hourly wages for the male and female sample evaluated at the mean, respectively, with $\bar{X}_G$ and $\hat{\beta}_G$ being $K \times 1$ vectors of average characteristics and estimated coefficients for $G = (F, M)$, where $G = F$ stands for female and $G = M$ stands for male. The first term is the endowments effect that evaluates the GPG in terms of characteristics at the rate of return of men.[17] As different endowments should have different effects on earnings, the difference in endowments represents the explained component of the Oaxaca-Blinder decomposition. The second term is the coefficients effect evaluating the GPG in terms of different returns for female characteristics. As the same endowments should have the same effect on earnings for both men and women, coefficients should not differ by gender, which is why this term represents the unexplained part of the GPG. If the GPG depends mainly on the difference in returns on characteristics, this may indicate the presence of gender discrimination.[18]

[17] Thus, we follow the standard approach using male coefficients as non-discriminatory wage structure.

[18] As stated before, the unexplained part of the GPG is often taken to be an estimate for gender discrimination. However, the unexplained portion of the GPG may include the effects of unmeasured productivity and some of the explanatory variables, such as the regressors accounting for gender differences in industries or occupations, may be affected by discrimination (Blau and Kahn, 2006).

In the case of public-contest recruitment, the GPG vanishes from 2005 onwards (see Table 2.5).[19] In contrast, Table 2.6 shows that if individuals do not enter by public contests in employment, there is a significant GPG in all years. In particular, the component generally referred to as discrimination is the main driver of the pay gap in all years. in fact, the endowments effect is mostly not statistically significant. Table 2.5 shows that the GPG among public-contest selected employees vanishes in the full sample of individuals aged 18-64 and even reverses in the young sample aged 18-34 years (Table 2.7). This is because the career path erodes the head start that young women receive by public-contest recruitment. Recruitment carried out by public contest significantly reverses the GPG among young employees in all years, except for 2014, where the reversal is not statistically significant. Moreover, the reversed wage gap is mainly explained by endowments, i.e. by the fact that women have better observable characteristics than men. The unexplained component is never statistically significant: given the same set of observable labor market characteristics for men and women, the difference in coefficients by gender is negligible (i.e. not statistically significant). In contrast, in the sample of young individuals not hired by public contest (Table 2.8), the GPG is either statistically significant and positive or zero, apart for 2011, where, however, the reversed GPG is substantially smaller compared to the reversal found for young public-contest selected employees. In the case of young individuals not hired by public contest, the coefficients component is significant (when a positive and significant GPG is found) suggesting that discrimination may already play a role in the early career, when individuals are not recruited by public contest. The different magnitude of the GPGs shown in Tables 2.6 and 2.8 may be due to the fact that even small differences at the start of the career expand greatly in the career path and give rise to large lifelong wage gaps.

[19] We do not decompose the zero-GPG arguing that in the absence of a pay disparity, there is no need to decompose the wage gap.

Table 2.5 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap, Public-Contest Selected Employees – Full Sample 18-64

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Differential* | | | | | | | |
| Male Wages (Log Hourly Wages) | 2.356*** | 2.289*** | 2.313*** | 2.366*** | 2.381*** | 2.383*** | 2.361*** |
| | (0.005) | (0.011) | (0.010) | (0.011) | (0.012) | (0.012) | (0.010) |
| Female Wages (Log Hourly Wages) | 2.352*** | 2.250*** | 2.310*** | 2.360*** | 2.382*** | 2.392*** | 2.370*** |
| | (0.005) | (0.010) | (0.010) | (0.010) | (0.011) | (0.011) | (0.009) |
| Difference | 0.004 | 0.039*** | 0.004 | 0.005 | -0.001 | -0.009 | -0.009 |
| | (0.007) | (0.014) | (0.014) | (0.015) | (0.016) | (0.016) | (0.014) |
| *Decomposition* | | | | | | | |
| Explained | | -0.050*** | | | | | |
| | | (0.013) | | | | | |
| Unexplained | | 0.089*** | | | | | |
| | | (0.015) | | | | | |
| Observations | 14,914 | 3,679 | 3,482 | 2,978 | 3,037 | 2,905 | 4,404 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.6 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap, Not Public-Contest Selected Employees – Full Sample 18-64

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Differential* | | | | | | | |
| Male Wages (Log Hourly Wages) | 2.001*** | 1.942*** | 1.958*** | 1.999*** | 2.015*** | 2.016*** | 2.077*** |
| | (0.004) | (0.007) | (0.007) | (0.007) | (0.008) | (0.008) | (0.007) |
| Female Wages (Log Hourly Wages) | 1.901*** | 1.802*** | 1.832*** | 1.893*** | 1.932*** | 1.967*** | 2.012*** |
| | (0.004) | (0.009) | (0.008) | (0.009) | (0.010) | (0.010) | (0.008) |
| Difference | 0.100*** | 0.140*** | 0.126*** | 0.106*** | 0.084*** | 0.049*** | 0.065*** |
| | (0.006) | (0.011) | (0.011) | (0.011) | (0.013) | (0.013) | (0.011) |
| *Decomposition* | | | | | | | |
| Explained | 0.005 | 0.004 | 0.035*** | -0.002 | 0.010 | -0.031*** | -0.012* |
| | (0.004) | (0.008) | (0.009) | (0.009) | (0.009) | (0.008) | (0.007) |
| Unexplained | 0.095*** | 0.136*** | 0.091*** | 0.108*** | 0.074*** | 0.081*** | 0.077*** |
| | (0.006) | (0.011) | (0.012) | (0.012) | (0.014) | (0.014) | (0.012) |
| Observations | 26,931 | 5,841 | 5,823 | 5,623 | 6,383 | 6,072 | 7,924 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.7 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap, Public-Contest Selected Employees – Young Sample 18-34

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Differential* | | | | | | | |
| Male Wages (Log Hourly Wages) | 2.054*** | 1.988*** | 2.005*** | 2.066*** | 2.052*** | 2.035*** | 2.172*** |
| | (0.015) | (0.033) | (0.027) | (0.030) | (0.039) | (0.030) | (0.028) |
| Female Wages (Log Hourly Wages) | 2.185*** | 2.083*** | 2.142*** | 2.217*** | 2.184*** | 2.241*** | 2.229*** |
| | (0.010) | (0.021) | (0.023) | (0.025) | (0.022) | (0.024) | (0.022) |
| Difference | -0.131*** | -0.094** | -0.137*** | -0.151*** | -0.132*** | -0.206*** | -0.057 |
| | (0.018) | (0.039) | (0.036) | (0.039) | (0.045) | (0.038) | (0.035) |
| *Decomposition* | | | | | | | |
| Explained | -0.149*** | -0.164** | -0.098** | -0.128** | -0.157*** | -0.160*** | |
| | (0.025) | (0.064) | (0.050) | (0.060) | (0.054) | (0.049) | |
| Unexplained | 0.017 | 0.070 | -0.039 | -0.024 | 0.024 | -0.047 | |
| | (0.031) | (0.077) | (0.062) | (0.072) | (0.069) | (0.058) | |
| Observations | 2,088 | 576 | 484 | 394 | 517 | 444 | 851 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.8 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap, Not Public-Contest Selected Employees – Young Sample 18-34

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Differential* | | | | | | | |
| Male Wages (Log Hourly Wages) | 1.835*** | 1.775*** | 1.778*** | 1.832*** | 1.852*** | 1.869*** | 1.934*** |
| | (0.005) | (0.009) | (0.009) | (0.010) | (0.010) | (0.010) | (0.010) |
| Female Wages (Log Hourly Wages) | 1.819*** | 1.711*** | 1.756*** | 1.811*** | 1.860*** | 1.903*** | 1.933*** |
| | (0.006) | (0.010) | (0.010) | (0.010) | (0.014) | (0.014) | (0.013) |
| Difference | 0.016** | 0.064*** | 0.022 | 0.021 | -0.007 | -0.033* | 0.002 |
| | (0.007) | (0.013) | (0.013) | (0.014) | (0.017) | (0.017) | (0.016) |
| *Decomposition* | | | | | | | |
| Explained | -0.049*** | -0.073*** | | | | -0.059*** | |
| | (0.006) | (0.011) | | | | (0.014) | |
| Unexplained | 0.065*** | 0.138*** | | | | 0.026 | |
| | (0.009) | (0.016) | | | | (0.022) | |
| Observations | 14,368 | 3,061 | 3,154 | 3,015 | 3,563 | 3,207 | 3,703 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

## 2.5 Accounting for Double Sample Selection: Model and Results

The GPG disappears when employees are recruited by public contests and even reverses in favor of women among public-contest selected employees aged 34 or younger. This may be due to the fairness of the public-contest selection mechanism rewarding women's characteristics as men's. However, the selection process into public-contest or non public-contest recruitment may be non-random and different for men and women (Heckman, 1979). The selection rule depends on two individual decisions; the work decision and the entry choice (recruitment by public contest or not). Our setup refers to the case of a censored probit, i.e. partial partial observability in the sense of Meng and Schmidt (1985). The output of the first decision is always observed, but the output of the second decision is observed if and only if the individual is in employment. In the model, the individual's work and entry decision are estimated simultaneously. The selection into employment may depend on positive factors such as individual ability, motivation or educational quality that raise both the probability of being employed and the level of wages but are omitted in the estimation of the earnings equation as these factors are unobservable. Additionally, we need to correct for any possible endogeneity bias that may result when the individual decision for public-contest recruitment also depends on the individual work decision. The selection rules are described by the following relations:

$$\text{Employment Selection:} \qquad Y_{iW}^{*} = Z_i^{'}\gamma + u_{iW} \tag{2.2}$$

$$\text{Public-Contest Selection:} \qquad Y_{iR}^{*} = Q_i^{'}\alpha + u_{iR} \tag{2.3}$$

where $Y_{iW}^{*}$ represents the unobservable index function underlying individual $i$'s decision whether to work or not and $Y_{iR}^{*}$ represents the unobservable index function underlying individual $i$'s decision to use the channel of public contest or not; with $Z_i$ and $Q_i$ being $K_z \times 1$ and $K_Q \times 1$ vectors of explanatory variables, respectively, and the error terms $u_i$ are assumed to be $N(0,1)$ with $Cov(u_W, u_R) = \rho$. The corresponding coefficients are the $K_Z \times 1$ vector $\gamma$ and the $K_Q \times 1$ vector $\alpha$, respectively.[20] The model is completed with wage equations for paid-employees. We estimate the model separately for the recruitment method chosen as well as for the female and male sample. The model can also be consistently estimated by Maximum Likelihood Estimation (MLE). Yet, the number of parameters to

[20]Further details on the methodology can be found in Appendix 2.C.

be estimated is rather large and by using FMLE we run into many convergence failures of the optimization methods. Therefore, we follow Tunali (1986) and Sorensen (1989) in extending the Heckman (1976; 1979) and Lee (1979; 1983) procedure by including selectivity coefficients as explanatory variables in the wage regression. The method proposed by Tunali (1986) is a two-step procedure that in the first step estimates equations (2.2) and (2.3) via MLE in order to obtain consistent estimates of the correction or selectivity terms, $\bar{\lambda}_W$ and $\bar{\lambda}_R$. This procedure allows wages to be generated through multiple selection rules explicitly recognizing the roles of both the work and the recruitment decision for the determination of the individual's employment status.

Table 2.9 presents the estimation results of the bivariate probit regression for men and women from the panel data. In Appendix 2.B, Tables 2.B.16–2.B.21, the corresponding results of the cross sections are shown. The parameter $\rho$ measuring the correlation of the residuals from the two models shows that the unobservable parts of the two equations are strongly and positively correlated for both men and women. Hence, it is important to model the two decisions jointly. The estimated values for $\rho$ suggest that there are positive and significant selection (or truncation) effects and that those who select into public-contest employment receive higher wages than a randomly chosen individual not selected into public-contest recruitment with a similar set of characteristics would receive. For identification of the two selection processes, each selection equation must contain at least one variable that is correlated with the respective decision but uncorrelated with the earnings equation or the other selection equation. In the case of the work decision, we use the indicator variables *Kids* and *Kids_10* as instruments. The intuition behind is that women with children and in particular with young children spend a significant amount of time with child-rearing and -caring and hence have a lower probability of accepting wage offers (Martins, 2001; Mulligan and Rubinstein, 2008; Lee, 2009; Chang, 2011). In contrast, men with children or young children have higher employment probability. This derives from the persistence of the male-breadwinner and mother-caretaker model in particular in Southern European countries like Italy, Greece or Spain (Mínguez, 2004). Similarly, we add the dummy *Partner_Works*[21] only to the employment equation following the literature that finds a strong relationship between the decision of women to work and spousal income (Devreux, 2004; Bar et al., 2015). The dummy variable *Age5064* controls for the effect of being on the last career stage on the employment probability. These regressors are assumed to affect individual reservation earnings but not the level of wages or individual preferences for a particular

[21] The dummy variable *Partner_Works* is equal to one if the partner of the individual is employed and zero otherwise.

recruitment method directly. We include also an indicator variable controlling for whether the individual has the Italian citizenship or not in the employment equation as there may be earnings as well as employment discrimination based on different cultural backgrounds of the individuals (Neuman and Oaxaca, 2003; Piazzalunga, 2015).[22] Additionally, controls for the individual's geographic position are included in the employment equation (as well as in the wage equation) as the probability of finding a job may increase significantly from the South to the Centre and the North of Italy. We do not include the dummies *North* and *Centre* in the recruitment equation as public contests do not differ across regions but are organized centrally. Similarly, whether an individual lives in Northern, Central or Southern Italy should not affect the probability of public-contest admission.

Variables measuring the level of work satisfaction (including the level of satisfaction with the working climate, the job stability, the working time and the task at the current job) are included in the recruitment-choice equation (as well as in the wage equation) as they are assumed to affect, besides the level of wages, the individual's propensity of public-contest recruitment. Indeed, jobs with public-contest entry may offer particular job characteristics. We use the indicator variables *Reloc* and *Risp* in order to identify the public-contest decision. *Reloc* indicates whether the individual relocated for his or her current job. In the public as well as in the private sector, public contests are generally organized at a central level and refer to positions in different regions. Consequently, the decision to participate in a public contest implies a willingness to relocate. The indicator variable *Risp* accounts for whether the individual answered questions on the quality of public services provided (such as on infrastructure, regularity of public bus services in the individual's city of residence or on the quality of health services provided by the government). Individuals responding to these questions may be more caring for others or interested in changing the current level of public services. Thus, they may be more prone to public-contest selection as, firstly, there is a strong correlation between public-sector employment and public-contest recruitment.[23] This implies that individuals wanting to improve or contribute to the quality of public services offered are more likely to choose public contests as channel of recruitment. Secondly, it may imply an inner consciousness for fairness as well as an aversion against discriminatory behavior of any kind and consequently higher probability to choose public-contest recruitment. Both

[22]The indicator variable *Italian* is not included in the recruitment equation as the general requirement for participation in a public contest (in particular in the public sector) is to hold the Italian citizenship and hence the indicator variable *Italian* does often not vary for public-contest selected individuals.

[23]We assume that public-sector employment is particularly interesting for individuals concerned about the quality of public services provided.

instrumental variables are excluded from the earnings as well as the employment equations as they should not affect marginal productivity or reservation earnings.

In the second step, the (double) selection-corrected wage equations are estimated for the respective subsamples. Adding the selection terms $\bar{\lambda}_W$ and $\bar{\lambda}_R$ to the earnings equations allows us to consistently estimate the earnings for public-contest and not public-contest selected individuals, respectively (Lee, 1983; Tunali, 1986):

$$\overline{ln(W_G^m)} = \bar{X}_G^{m'} \hat{\beta}_G^m + \hat{\delta}_{W,G}^m \bar{\lambda}_{W,G}^m + \hat{\delta}_{R,G}^m \bar{\lambda}_{R,G}^m \tag{2.4}$$

where $m = (PC, NPC)$, $m = PC$ controls for individuals selected by public contest and $m = NPC$ for individuals not selected by public contest, and $G = (F, M)$, where $G = F$ accounts for female and $G = M$ for male. Following Heinze et al. (2003), when considering sample selection, the decomposition in equation (2.1) becomes:

$$\begin{aligned}\overline{ln(W_M^m)} - \overline{ln(W_F^m)} = & (\bar{X}_M^{m'} - \bar{X}_F^{m'})\hat{\beta}_M^m + \bar{X}_F^{m'}(\hat{\beta}_M^m - \hat{\beta}_F^m) \\ & + (\hat{\delta}_{W,M}^m \bar{\lambda}_{W,M}^m - \hat{\delta}_{W,F}^m \bar{\lambda}_{W,F}^m) + (\hat{\delta}_{R,M}^m \bar{\lambda}_{R,M}^m - \hat{\delta}_{R,F}^m \bar{\lambda}_{R,F}^m)\end{aligned} \tag{2.5}$$

The double selection mechanism may reveal benefits from selection through public contest for men and women. If the selection effect of both the employment decision and the recruitment decision is significant and positive, women and men selected by public contest would have higher unobserved characteristics and wages than women and men with the same observed characteristics not selected by public contest. Table 2.10 defines the four selection variables considered in this study and presents the coefficient estimates of the selection terms for both men and women in the full sample.[24] The signs of the estimated coefficients of the respective $\lambda$'s[25] for the employment decision, $\lambda_W^{PC}$ and $\lambda_W^{NPC}$, are positive and statistically significant. Thus, individuals in employment have on average higher unobservable characteristics compared to otherwise observationally identical unemployed individuals. This means that those unobserved characteristics raising the probability of being employed also increase wages. If not positive, the coefficient estimate of the selectivity variable $\lambda_W$ is generally not statistically significant.[26] In this case, employees not selected (or selected) by public contest have almost the same unobserved characteristics and wage offers than unemployed individuals. In the sample of individuals recruited by public contest,

[24]The complete wage regressions with selection variables for both the panel and cross sections are provided in Appendix 2.B, Table 2.B.22 and Table 2.B.23–2.B.28.

[25]In the following, for simplicity; $\bar{\lambda} = \lambda$.

[26]Except for women not selected by public contest in 2006 and men not selected by public contest in 2011.

the positive sign of the estimated coefficient of $\lambda_R^{PC}$ indicates that those unobserved positive characteristics raising the probability of winning a contest also increase wages. Hence, individuals that are recruited by public contest have more favorable unobserved characteristics and wages than individuals not recruited by public contest would have obtained if they were recruited by public contest. In contrast, as expected, the selectivity variable $\lambda_R^{NPC}$ has negative and (generally) statistically significant coefficient estimates. Employees recruited without public contest have lower levels of unobserved characteristics impacting negatively on the wage level than individuals actually selected by public contest.

To sum up, we find evidence that individuals recruited by public contest have more favorable unobserved characteristics and earnings than other employees with similar observed characteristics and actually unemployed individuals would have obtained if they were recruited by public contest.

The results shown in Table 2.10 strengthen the results found in Section 2.4 that public contests are merit-based selection methods. The coefficients of $\lambda_R^{PC}$, positive and significant, confirm that women selected by public contest have better unobserved characteristics than women not selected by public contest. The male coefficients of $\lambda_R^{PC}$ are statistically insignificant, and thus men do neither receive a wage premium nor a wage penalty from public-contest recruitment. All in all, the positive effect from public-contest selection is more pronounced for women than for men.

Table 2.9 Bivariate Probit Estimation by Gender – Panel

| Year | Panel | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.066*** | 0.025*** | 0.044*** | 0.005*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Educ | 0.804*** | 0.493*** | 0.528*** | 0.217*** |
| | (0.015) | (0.010) | (0.013) | (0.011) |
| Married | 0.138*** | -0.060** | 0.300*** | 0.528*** |
| | (0.023) | (0.025) | (0.024) | (0.030) |
| Homeowner | 0.164*** | 0.088*** | 0.088*** | 0.193*** |
| | (0.027) | (0.018) | (0.025) | (0.021) |
| Age5064 | | 0.789*** | | 0.345*** |
| | | (0.030) | | (0.030) |
| Italian | | 0.259*** | | 0.246*** |
| | | (0.059) | | (0.093) |
| North | | 0.797*** | | 0.776*** |
| | | (0.016) | | (0.017) |
| Centre | | 0.499*** | | 0.504*** |
| | | (0.019) | | (0.021) |
| Partner_Works | | 0.036* | | 0.128*** |
| | | (0.021) | | (0.026) |
| Kids | | -0.172*** | | 0.153*** |
| | | (0.024) | | (0.026) |
| Kids_10 | | -0.152*** | | -0.023 |
| | | (0.024) | | (0.034) |
| Work_Climate | -0.026** | | -0.063*** | |
| | (0.013) | | (0.012) | |
| Work_Stab | 0.166*** | | 0.233*** | |
| | (0.010) | | (0.011) | |
| Work_Time | 0.082*** | | 0.057*** | |
| | (0.013) | | (0.012) | |
| Work_Task | -0.032** | | 0.000 | |
| | (0.014) | | (0.013) | |
| Reloc | 0.417*** | | 0.442*** | |
| | (0.039) | | (0.029) | |
| Risp | 0.051** | | 0.123*** | |
| | (0.023) | | (0.023) | |
| Constant | -7.000*** | -3.123*** | -5.638*** | -1.506*** |
| | (0.100) | (0.081) | (0.085) | (0.108) |
| $\rho$ | 0.618*** | | 1.222*** | |
| | (0.055) | | (0.111) | |
| Year Dummies | Yes | Yes | Yes | Yes |
| Observations | 39,345 | | 33,538 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.10 Selection Variables – Definition and Values

| Year | Panel | | 2005 | | 2006 | | 2008 | | 2010 | | 2011 | | 2014 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| | Women | Men | Women | Men | Women | Men | Women | Men | Women | Men | Women | Men | Women | Men |
| $\lambda_W^{PC}$ | | | | | | | | | | | | | | |
| measures the selection bias | 0.126*** | 0.209* | 0.019 | 0.375* | 0.047 | 0.469** | 0.033 | 0.272 | 0.077 | -0.168 | -0.018 | 0.451* | -0.055 | 0.233 |
| from the *work decision* | (0.035) | (0.111) | (0.045) | (0.210) | (0.080) | (0.219) | (0.084) | (0.252) | (0.104) | (0.422) | (0.110) | (0.261) | (0.104) | (0.291) |
| for those selected | | | | | | | | | | | | | | |
| by public contest | | | | | | | | | | | | | | |
| $\lambda_R^{PC}$ | | | | | | | | | | | | | | |
| measures the selection bias | 0.179*** | -0.009 | 0.118** | -0.085 | 0.178** | 0.088 | 0.206** | -0.099 | 0.116 | 0.041 | 0.165 | 0.094 | 0.181** | -0.027 |
| from the *recruitment decision* | (0.035) | (0.039) | (0.053) | (0.070) | (0.088) | (0.077) | (0.082) | (0.098) | (0.089) | (0.096) | (0.102) | (0.097) | (0.088) | (0.084) |
| for those selected | | | | | | | | | | | | | | |
| by public contest | | | | | | | | | | | | | | |
| Observations | 8,116 | 6,798 | 1,987 | 1,692 | 1,715 | 1,767 | 1,586 | 1,392 | 1,621 | 1,416 | 1,656 | 1,249 | 2,665 | 1,739 |
| $\lambda_W^{NPC}$ | | | | | | | | | | | | | | |
| measures the selection bias | -0.041 | -0.013 | 0.009 | -0.050 | -0.109** | 0.006 | -0.065 | -0.006 | 0.023 | -0.041 | 0.162 | -0.297*** | 0.010 | 0.074 |
| from the *work decision* | (0.032) | (0.043) | (0.037) | (0.072) | (0.053) | (0.081) | (0.054) | (0.074) | (0.087) | (0.089) | (0.101) | (0.080) | (0.092) | (0.115) |
| for those NOT selected | | | | | | | | | | | | | | |
| by public contest | | | | | | | | | | | | | | |
| $\lambda_R^{NPC}$ | | | | | | | | | | | | | | |
| measures the selection bias | -0.305** | -0.169*** | -0.115** | -0.097* | -0.220*** | -0.123** | -0.338*** | -0.168*** | -0.313*** | -0.251*** | -0.506*** | -0.363*** | -0.346*** | -0.116 |
| from the *recruitment decision* | (0.034) | (0.032) | (0.058) | (0.055) | (0.059) | (0.061) | (0.060) | (0.061) | (0.089) | (0.077) | (0.090) | (0.070) | 0.088) | (0.086) |
| for those NOT selected | | | | | | | | | | | | | | |
| by public contest | | | | | | | | | | | | | | |
| Observations | 11,282 | 15,649 | 2,526 | 3,315 | 2,368 | 3,455 | 2,370 | 3,253 | 2,588 | 3,795 | 2,512 | 3,560 | 3,428 | 4,496 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

## 2.6 Counterfactual GPG Adjusted for Double Selection

In Section 2.4 we have shown that the GPG in the case of public-contest recruitment is reduced to zero (see Table 2.5). In the case of non public-contest recruitment, the coefficients effect was identified to be the main driver of a positive and significant GPG. In order to verify the robustness of these results in the presence of selection bias, we compute the CPG adjusted for (double) selectivity bias for individuals not selected by public contests.[27] In particular, the CPG adjusted for double selection is given by:

$$CPG^m = \bar{X}_F^{m'}(\hat{\beta}_M^m - \hat{\beta}_F^m) + (\hat{\delta}_{W,M}^m - \hat{\delta}_{W,F}^m)\bar{\lambda}_{W,F}^m + (\hat{\delta}_{R,M}^m - \hat{\delta}_{R,F}^m)\bar{\lambda}_{R,F}^m \tag{2.6}$$

where $m = (PC, NPC)$.

Table 2.11 shows the estimated adjusted difference in pay between men and women not hired by public contest after having corrected for (double) selectivity bias. For individuals not selected by public contest, the counterfactual analysis corrected for selectivity bias confirms the results obtained before. In both cases, with and without taking into account the correction for selection bias, the unexplained part turns out to be the most important driver of the GPG in the case of non public-contest selection. The estimation results predict that women in non public-contest jobs earn between 3.3% (in 2011)[28] and 12.5% (in 2005) less than they would earn if they were remunerated as men. Hence, in the case of non public-contest recruitment, a non-trivial pay disparity between women and men still exists even after adjusting for selectivity and productivity-related characteristics.

Table 2.11 CPG Adjusted for Double Selection, Not Public-Contest Selected Employees – Full Sample 18-64

| Year | Panel | 2005 | 2006 | 2008 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Difference | 0.100*** | 0.140*** | 0.126*** | 0.106*** | 0.084*** | 0.049*** | 0.065*** |
| | (0.006) | (0.011) | (0.011) | (0.011) | (0.013) | (0.013) | (0.011) |
| Counterfactual | 0.109*** | 0.125*** | 0.109*** | 0.124*** | 0.086*** | 0.033 | 0.105*** |
| | (0.013) | (0.028) | (0.032) | (0.030) | (0.025) | (0.025) | (0.021) |
| Observations | 26,931 | 5,841 | 5,823 | 5,623 | 6,383 | 6,072 | 7,924 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

[27] As in the case of public contest recruitment, the GPG – except for 2005 – is zero and thus no additional insights can be gained by looking at the CPG for public-contest recruited individuals.

[28] In 2001, the CPG is statistically insignificant.

## 2.7 Public- versus Private-Sector Recruitment

The institutional rules and practices that typically apply to public-sector employment to some extent insulate public-sector jobs from the uncertainties of labor market forces. Public-sector jobs are generally more stable over time and more tightly linked to experience and education than private-sector jobs. Overall, this higher degree of job security for civil servants comes along with higher barriers to entry into public employment. Differences in entry requirements, wage-setting practices, contract types and career paths between the public and the private sector affect the public-private sector gap in both pay and lifetime values. In the 1980's, permanent hiring without public contests of 'pro-tempore' workers[29] and 'ope legis' promotions has increased recruitment in general and, in particular, recruitment without public contests in the public sector (Craveri, 2016). Pay increases for public-sector employees were automatic until the reform of 1993.[30] Since then, remuneration is determined by employee collective agreements. Automatic wage increases and career promotions linked to seniority were substituted with more discretionary, selective and performance-related mechanisms. After changes in public employment in the 2000's[31] for the purpose of optimizing labor productivity and in order to make the organization of public-sector employment more flexible, hiring on open-ended contracts and flexible forms of employment are now allowed in the public administration. Moreover, collective agreements regulating fixed-term contracts, training contracts, and the supply of temporary labor are now possible. Hence, institutional differences between the public and private sector (e.g. in the legislative regulation, employment relationships or personnel policies) persist, but are decreasing nowadays (Postel-Vinay, 2015).

We calculate the GPG by entry channel only for the public sector, in order to show that the disappearance of the wage gap is due to the mechanism of public contests and not due to public-sector employment. The decomposition results for the public sector are shown in Table 2.12. The results confirm the results of the previous analysis. In the case of public-contest recruitment, the GPG is insignificant and thus disappears on average, while in the case of non public-contest recruitment, even in the public sector, there is a positive and statistically significant GPG. In the case of non-public contest recruitment, the unexplained or coefficients part is again the main driver of the wage gap. This suggests that the disappearance of the

[29] 'Pro-tempore' workers are individuals in temporary employment.

[30] The main changes were introduced by legislative decree 29/1993, and subsequently by legislative decrees 396/1997, 80/1998 and 387/1998, with consequences on the status of public-sector employees, their employment relationships and personnel policies.

[31] In particular: law 133/2008 and the legislative decree 150/2009.

GPG among public-contest selected employees is not entirely driven by the institutional environment of the public sector (as without the mechanism of public-contest selection, a significant and positive GPG for public-sector employees is found).

In the following, we test whether the impact of public contests on the GPG depends on the institutional environment of the public sector at all. Therefore, we analyze the GPG by recruitment method only for the private sector. Table 2.13 presents the decomposition result by public-contest recruitment for the private sector only. We find a positive and significant GPG, regardless of whether individuals are selected by public contest or not. Moreover, the difference in pay between men and women is even higher for public-contest selected employees in the private sector compared to not public-contest selected individuals. In both subsamples, the unexplained as well as the explained part are positive and statistically significant. Thus, the mechanism of public contests as gender-fair and merit-based screening devices requires specific institutional environments. In the public sector, these institutional requirements are given. Even though the institutional background is not the only factor contributing to the success of merit-based and gender-fair screening via public contests (there is a positive and significant GPG among public servants not selected by public contest), it is a crucial one.

Table 2.12 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap – Public Sector

| Year | Panel | |
|---|---|---|
| | (1) | (2) |
| | Public-Contest Selected Employees | Not Public-Contest Selected Employees |
| *Differential* | | |
| Male Wages (Log Hourly Wages) | 2.366*** | 2.180*** |
| | (0.006) | (0.011) |
| Female Wages (Log Hourly Wages) | 2.367*** | 2.129*** |
| | (0.005) | (0.011) |
| Difference | -0.001 | 0.052*** |
| | (0.007) | (0.015) |
| *Decomposition* | | |
| Explained | | -0.024** |
| | | (0.012) |
| Unexplained | | 0.075*** |
| | | (0.017) |
| Observations | 13,595 | 4,394 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table 2.13 Log Hourly Wages and Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap – Private Sector

| Year | Panel | |
|---|---|---|
| | (1) | (2) |
| | Public-Contest Selected Employees | Not Public-Contest Selected Employees |
| *Differential* | | |
| Male Wages (Log Hourly Wages) | 2.289*** | 1.974*** |
| | (0.014) | (0.004) |
| Female Wages (Log Hourly Wages) | 2.122*** | 1.843*** |
| | (0.014) | (0.005) |
| Difference | 0.167*** | 0.131*** |
| | (0.020) | (0.006) |
| *Decomposition* | | |
| Explained | 0.029* | 0.032*** |
| | (0.017) | (0.004) |
| Unexplained | 0.138*** | 0.099*** |
| | (0.023) | (0.006) |
| | | |
| Observations | 1,319 | 22,537 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

## 2.8 Conclusion

This paper analyzes the effect of hiring methods on earnings in Italy. Therefore, the GPG is decomposed in an explained and an unexplained component using the Oaxaca-Blinder decomposition approach. The estimates are then corrected for double selection using the partial partial observability approach by Meng and Schmidt (1985). The decision to enter in employment as well as the individual's entry decision are thereby modeled simultaneously. Employment selection may be particularly relevant for Italy given especially pronounced gender differences in employment and labor force participation. Similarly, public-contest selection may be non-random and different for men and women. In fact, public-contest recruitment may be preferred by women given less potential for discriminatory behavior in the hiring process. Consequently, failure to account for (double) sample selection leads to inconsistent estimates of the gender-specific wage equations as well as of the components of the GPG.

Our results suggest that public contests reduce the conditions for gender discrimination to flourish and are merit-based and gender-fair mechanisms for performance appraisal. They are merit-based because employees hired by public contest hold better observable and unobservable characteristics than unselected employees. They are gender-fair because among public-contest selected employees, there is no gender-related penalty on wages. We show that recruitment carried out by public contests erases the GPG in the full sample of individuals aged 18-64, and even reverses the gap in favor of women among young employees. The relatively strong wage gap in favor of young women is only explained by endowments, i.e. by the fact that women have better observable characteristics than men. The reversal of the GPG observed among public-contest selected young employees vanishes in the full sample, even in the case of public-contest recruitment. This is because the career path erodes the head start that young women receive by public-contest recruitment. This result is in line with the literature finding that women are less often promoted and have generally more breaks in their careers due to childbearing and -care (Blackaby et al., 2005; Niederle and Vesterlund, 2007; Fortin, 2008; Heilman and Okimoto, 2008; Bertrand et al., 2011). In the case of employees not hired by public contest, the component accounting for discrimination is the main driver of the disparity in pay. Even after adjusting the unexplained component of the Oaxaca-Blinder decomposition for double selection (i.e. estimating the CPG), we still find a substantial GPG for not public-contest selected individuals. The robust CPG underpins the result that public-contest recruitment significantly impacts on gender differences in pay. In particular, it provides an option for gender-fair screening. Analyzing the GPG by recruitment in different institutional environments, i.e. in the public and the private sector, shows that the institutional background does indeed matter for gender-fair and merit-based screening. A positive and significant GPG persists for both public-contest and not public-contest selected employees in private-sector employment. However, this does not necessarily imply that the disappearance of the GPG among public-contest selected employees is entirely driven by the public sector. In fact, a statistically significant and positive GPG among individuals not selected by public contest in the public sector is found. Moreover, the decomposition reveals that the wage gap is entirely due to the unexplained component. Thus, even though an institutional environment similar to that in the public sector was found to be important for the mechanism of public-contest selection to work, the results suggest that public contests are gender-neutral and merit-based procedures picking out the most deserving participants as they are less discretionary and more regulated by law than other screening devices. Further research on the environments necessary for public contests as gender-fair and merit-based screening devices is certainly needed. To the best of our knowledge, so far no other research establishes a relationship between recruitment procedures and the GPG and shows empirically the disappearance of the wage gap – given certain institutional characteristics.

# Appendices

## Appendix 2.A Definition of Variables

Table 2.A.1 Definition of Variables

| Variable Name | Definition |
|---|---|
| | Dependent Variables |
| Net_ Hourly_Wage | Hourly wages in Euros and net of taxes and social security contributions |
| Log_Hourly_Wage | The natural log of net hourly earnings; wages are in Euros and net of taxes and social security contributions |
| Employment | One if the respective individual decided to accept a wage offer, i.e. to enter in employment, zero if (voluntarily) unemployed |
| Public_Contest | One if individual entered via public contest in the current job, zero otherwise *Public_Contest* is also used as independent variable |
| | Independent Variables |
| Female | One if the respective individual is a woman, zero otherwise |
| Contfem | Interactive effect of the dummy variables *Public_Contest* and *Female*, i.e. one if the respective employee entered via public contest in his or her current job and is female, zero otherwise |
| Exper | Number of years of work experience |
| Exper2 | *Exper* squared |
| Tenure | Number of years worked for current employer |
| Educ | Number of years of schooling completed |
| University_Degree | One if the respective individual has graduated from university, zero otherwise |
| Max_D_Mark | One if the maximum degree mark was attained, i.e. *110 e lode*, in the case of graduation from university, zero otherwise |
| North | One if the respective individual lives and works in the North of Italy, zero otherwise |
| Centre | One if the respective individual lives and works in the Centre of Italy, zero otherwise |
| Age | Age of the respective individual (in years) $\in (18, 64)$ |
| Age5064 | One if the age of the respective individual is between 50 and 64 years, zero otherwise |
| Married | One if the respective individual is married, zero otherwise |
| Italian | One if the respective individual holds the Italian citizenship, zero otherwise |

| | |
|---|---|
| Hometime | Years the respective individual spent out of the labor force |
| Educ_Moth_Uni | One if the mother's education is equal to *University_Degree*, i.e. the mother holds a university degree, zero otherwise |
| Educ_Fath_Uni | One if the father's education is equal to *University_Degree*, i.e. the father holds a university degree, zero otherwise |
| Kids | One if the respective individual has at least one child, zero otherwise |
| Kids_10 | One if the age of the youngest child is below 10 years, zero otherwise |
| Homeowner | One if the respective individual owns a house, zero otherwise This includes bank loan-financed houses |
| Partner_Works | One if the partner of the respective individual is employed, zero otherwise |
| Risp | One if the respective individual responds to questions on the quality of public services, zero otherwise |
| Reloc | One if the respective individual relocated in order to take the current job, zero otherwise |
| Work_Climate | Level of satisfaction with working climate at current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Stab | Level of satisfaction with stability of current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Time | Level of satisfaction with working time at current job $\in (0,4)$ where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Task | Level of satisfaction with tasks at current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Contract_Type | One if the respective individual holds an unlimited contract, zero otherwise |
| Manager | One if the respective individual is occupied in an intellectual profession; scientific or highly specialized occupations, zero otherwise |
| Intermediate_Prof | One if the respective individual is occupied in an intermediary position in the commercial, technical or administrative sector, in health services or is a technician, zero otherwise |
| Sec_02 - Sec_15 | Sectoral dummies for employment in manufacturing, energy, construction, tourism, commerce, transport, communication, financial activities, service industry, public administration, education, health, sciences and family services, respectively |
| Public_Sector | Dummy variable for public-sector employment; one if the respective individual is employed in the public sector, zero otherwise |
| Year_1-Year_5 | Year dummies, one if year = 2005, 2006, 2008, 2010, 2011, respectively, and zero otherwise |

Selection Correction Terms

| | |
|---|---|
| $\lambda_W^{PC}$ | Measures the selection bias from the work decision for those selected by public contest |

| | |
|---|---|
| $\lambda_R^{PC}$ | Measures the selection bias from the recruitment decision for those selected by public contest. |
| $\lambda_W^{NPC}$ | Measures the selection bias from the work decision for those not selected by public contest |
| $\lambda_R^{NPC}$ | Measures the selection bias from the recruitment decision for those not selected by public contest |

# Appendix 2.B Descriptive Statistics Cross Sections and Further Estimation Results

Table 2.B.1 Descriptive Statistics – Cross Sections

| Year | 2005 | | | | 2006 | | | | 2008 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | Public-Contest Selected Employees | | Not Public-Contest Selected Employees | | Public-Contest Selected Employees | | Not Public-Contest Selected Employees | | Public-Contest Selected Employees | | Not Public-Contest Selected Employees | |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. |
| Female | 0.540 | 0.499 | 0.433 | 0.496 | 0.493 | 0.500 | 0.407 | 0.491 | 0.533 | 0.4990 | 0.4215 | 0.4938 |
| Educ | 13.869 | 2.342 | 12.198 | 2.873 | 13.922 | 2.303 | 12.325 | 2.849 | 14.076 | 2.218 | 12.539 | 2.860 |
| University_Degree[a] | 0.435 | 0.496 | 0.173 | 0.378 | 0.443 | 0.497 | 0.193 | 0.395 | 0.474 | 0.4994 | 0.2356 | 0.4244 |
| Max_D_Mark | 0.095 | 0.293 | 0.029 | 0.167 | 0.095 | 0.293 | 0.029 | 0.166 | 0.076 | 0.2643 | 0.0212 | 0.1439 |
| Exper | 24.053 | 10.459 | 16.815 | 12.568 | 24.639 | 10.484 | 16.200 | 12.597 | 25.509 | 10.5136 | 16.2659 | 13.0295 |
| Tenure | 19.045 | 10.705 | 10.469 | 10.487 | 19.628 | 10.765 | 9.957 | 10.521 | 20.693 | 11.0043 | 10.9641 | 11.1586 |
| Married | 0.739 | 0.439 | 0.512 | 0.500 | 0.742 | 0.437 | 0.474 | 0.499 | 0.744 | 0.4366 | 0.4635 | 0.4987 |
| Kids | 0.788 | 0.409 | 0.810 | 0.392 | 0.744 | 0.437 | 0.434 | 0.496 | 0.750 | 0.4330 | 0.4261 | 0.4946 |
| Kids_10 | 0.070 | 0.255 | 0.080 | 0.271 | 0.190 | 0.392 | 0.168 | 0.374 | 0.145 | 0.3526 | 0.1428 | 0.3499 |
| Age | 47.064 | 9.880 | 37.035 | 12.192 | 47.779 | 9.900 | 36.612 | 12.237 | 49.161 | 9.8433 | 37.2170 | 12.5640 |
| Public_Sector | 0.921 | 0.270 | 0.161 | 0.368 | 0.916 | 0.277 | 0.144 | 0.351 | 0.904 | 0.2952 | 0.1590 | 0.3657 |
| Contract_Type | 0.923 | 0.267 | 0.829 | 0.376 | 0.929 | 0.256 | 0.788 | 0.409 | 0.932 | 0.2521 | 0.7736 | 0.4185 |
| Manager | 0.442 | 0.497 | 0.111 | 0.314 | 0.352 | 0.478 | 0.105 | 0.307 | 0.244 | 0.4296 | 0.1149 | 0.3189 |
| Intermed_Prof | 0.453 | 0.498 | 0.490 | 0.500 | 0.527 | 0.499 | 0.497 | 0.500 | 0.636 | 0.4811 | 0.4964 | 0.5000 |
| Observations | 3,679 | | 5,841 | | 3,482 | | 5,823 | | 2,978 | | 5,623 | |

| Year | 2010 | | | | 2011 | | | | 2014 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) | (22) | (23) | (24) |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. |
| Female | 0.534 | 0.499 | 0.406 | 0.491 | 0.570 | 0.495 | 0.414 | 0.493 | 0.605 | 0.489 | 0.433 | 0.496 |
| Educ | 14.036 | 2.237 | 12.578 | 2.802 | 14.071 | 2.186 | 12.532 | 2.804 | 14.183 | 2.061 | 12.854 | 2.749 |
| University_Degree[a] | 0.467 | 0.499 | 0.231 | 0.422 | 0.468 | 0.499 | 0.224 | 0.417 | 0.481 | 0.500 | 0.277 | 0.447 |
| Max_D_Mark | 0.111 | 0.314 | 0.041 | 0.198 | 0.081 | 0.273 | 0.034 | 0.182 | 0.056 | 0.230 | 0.026 | 0.160 |
| Exper | 25.273 | 11.522 | 16.220 | 13.062 | 25.506 | 11.342 | 16.397 | 13.001 | 24.734 | 11.843 | 18.136 | 12.879 |
| Tenure | 20.717 | 11.693 | 10.818 | 11.057 | 21.076 | 11.596 | 11.038 | 11.120 | 20.782 | 11.882 | 12.614 | 11.384 |
| Married | 0.708 | 0.455 | 0.450 | 0.498 | 0.724 | 0.447 | 0.446 | 0.497 | 0.760 | 0.427 | 0.546 | 0.498 |
| Kids | 0.717 | 0.451 | 0.409 | 0.492 | 0.734 | 0.442 | 0.424 | 0.494 | 0.723 | 0.448 | 0.491 | 0.500 |
| Kids_10 | 0.149 | 0.356 | 0.137 | 0.344 | 0.145 | 0.352 | 0.144 | 0.351 | 0.138 | 0.345 | 0.160 | 0.367 |
| Age | 48.380 | 10.730 | 36.863 | 12.682 | 48.769 | 10.620 | 37.503 | 12.553 | 47.864 | 11.590 | 39.352 | 12.325 |
| Public_Sector | 0.900 | 0.301 | 0.161 | 0.367 | 0.899 | 0.302 | 0.172 | 0.378 | 0.915 | 0.279 | 0.193 | 0.395 |
| Contract_Type | 0.924 | 0.266 | 0.779 | 0.415 | 0.925 | 0.264 | 0.779 | 0.415 | 0.945 | 0.228 | 0.832 | 0.374 |
| Manager | 0.391 | 0.488 | 0.168 | 0.374 | 0.418 | 0.493 | 0.187 | 0.390 | 0.527 | 0.499 | 0.192 | 0.394 |
| Intermed_Prof | 0.466 | 0.499 | 0.416 | 0.493 | 0.450 | 0.498 | 0.377 | 0.485 | 0.386 | 0.487 | 0.454 | 0.498 |
| Observations | 3,037 | | 6,383 | | 2,905 | | 6,072 | | 4,404 | | 7,924 | |

[a] The dummy variable *University_Degree* is not used in the regression analysis in both the panel and the cross sections. It is only presented for illustration of the variation of university graduation in the data.

Table 2.B.2 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – Panel

| Year | Panel | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.068*** | 0.128*** | | | | |
| | (0.006) | (0.016) | | | | |
| Female | -0.104*** | -0.069*** | -0.064*** | -0.008 | -0.104*** | -0.069*** |
| | (0.005) | (0.008) | (0.006) | (0.021) | (0.005) | (0.008) |
| Contfem | 0.040*** | 0.050*** | | | | |
| | (0.008) | (0.019) | | | | |
| Exper | 0.018*** | 0.020*** | 0.014*** | 0.016** | 0.018*** | 0.020*** |
| | (0.001) | (0.003) | (0.001) | (0.008) | (0.001) | (0.003) |
| Exper2 | -0.000*** | -0.000*** | -0.000*** | -0.001 | -0.000*** | -0.000** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.006*** | 0.002*** | 0.001 | 0.004*** | 0.007*** |
| | (0.000) | (0.001) | (0.000) | (0.003) | (0.000) | (0.001) |
| Educ | 0.148*** | 0.098*** | 0.196*** | 0.100*** | 0.126*** | 0.100*** |
| | (0.004) | (0.007) | (0.006) | (0.019) | (0.005) | (0.008) |
| Max_D_Mark | 0.047*** | 0.051*** | 0.029** | 0.049* | 0.043*** | 0.044** |
| | (0.009) | (0.016) | (0.012) | (0.027) | (0.015) | (0.020) |
| Contract_Type | 0.040*** | 0.034*** | 0.004 | -0.018 | 0.042*** | 0.038*** |
| | (0.007) | (0.008) | (0.015) | (0.025) | (0.007) | (0.009) |
| Work_Climate | 0.000 | -0.004 | 0.002 | 0.003 | -0.000 | -0.005 |
| | (0.003) | (0.005) | (0.004) | (0.012) | (0.004) | (0.005) |
| Work_Stab | 0.005* | 0.005 | 0.008* | 0.026** | 0.005 | 0.004 |
| | (0.003) | (0.004) | (0.004) | (0.012) | (0.003) | (0.005) |
| Work_Time | 0.017*** | 0.028*** | 0.005 | 0.017 | 0.022*** | 0.029*** |
| | (0.003) | (0.004) | (0.004) | (0.011) | (0.003) | (0.005) |
| Work_Task | 0.014*** | 0.009* | 0.018*** | 0.002 | 0.012*** | 0.010* |
| | (0.003) | (0.005) | (0.004) | (0.012) | (0.004) | (0.005) |
| Intermed_Prof | 0.037*** | 0.023*** | 0.008 | 0.012 | 0.055*** | 0.026*** |
| | (0.005) | (0.008) | (0.009) | (0.025) | (0.006) | (0.008) |
| Manager | 0.158*** | 0.060*** | 0.155*** | 0.077** | 0.139*** | 0.054*** |
| | (0.007) | (0.013) | (0.011) | (0.030) | (0.010) | (0.015) |
| North | 0.028*** | 0.066*** | -0.022*** | -0.033* | 0.057*** | 0.082*** |
| | (0.005) | (0.008) | (0.006) | (0.019) | (0.006) | (0.009) |
| Centre | 0.011** | 0.037*** | -0.006 | -0.029 | 0.026*** | 0.049*** |
| | (0.005) | (0.010) | (0.008) | (0.023) | (0.007) | (0.010) |
| Home_Time | 0.006*** | 0.013*** | 0.004*** | 0.010*** | 0.007*** | 0.013*** |
| | (0.000) | (0.001) | (0.001) | (0.003) | (0.001) | (0.001) |
| Married | 0.051*** | 0.048*** | 0.036*** | 0.085*** | 0.056*** | 0.045*** |
| | (0.005) | (0.008) | (0.007) | (0.019) | (0.006) | (0.009) |
| Italian | 0.061** | 0.029 | 0.074 | 0.117 | 0.065** | 0.027 |
| | (0.026) | (0.032) | (0.067) | (0.185) | (0.028) | (0.033) |
| Homeowner | 0.027*** | 0.011 | 0.021** | -0.014 | 0.031*** | 0.015 |
| | (0.005) | (0.009) | (0.009) | (0.022) | (0.006) | (0.010) |
| Educ_Fath_Uni | 0.012 | -0.026* | 0.006 | -0.033 | 0.008 | -0.026 |
| | (0.009) | (0.014) | (0.012) | (0.027) | (0.013) | (0.017) |
| Educ_Moth_Uni | 0.008 | 0.024 | 0.019 | 0.006 | 0.003 | 0.025 |
| | (0.012) | (0.016) | (0.018) | (0.036) | (0.016) | (0.018) |
| Constant | 0.942*** | 1.066*** | 1.038*** | 1.285*** | 0.957*** | 1.040*** |
| | (0.034) | (0.053) | (0.077) | (0.220) | (0.039) | (0.056) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 41,845 | 16,456 | 14,914 | 2,088 | 26,931 | 14,368 |
| R-squared | 0.380 | 0.170 | 0.342 | 0.210 | 0.243 | 0.115 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.3 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2005

| Year | 2005 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.062*** | 0.070** | | | | |
| | (0.014) | (0.035) | | | | |
| Female | -0.132*** | -0.138*** | -0.105*** | -0.050 | -0.137*** | -0.139*** |
| | (0.010) | (0.014) | (0.013) | (0.045) | (0.010) | (0.014) |
| Contfem | 0.025 | 0.062 | | | | |
| | (0.015) | (0.040) | | | | |
| Exper | 0.019*** | 0.020*** | 0.014*** | 0.032* | 0.020*** | 0.018*** |
| | (0.002) | (0.005) | (0.003) | (0.017) | (0.002) | (0.006) |
| Exper2 | -0.000*** | -0.000 | -0.000** | -0.001 | -0.000*** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Tenure | 0.002*** | 0.001 | 0.002** | -0.010* | 0.002*** | 0.004* |
| | (0.001) | (0.002) | (0.001) | (0.005) | (0.001) | (0.002) |
| Educ | 0.167*** | 0.154*** | 0.200*** | 0.199*** | 0.145*** | 0.149*** |
| | (0.008) | (0.014) | (0.012) | (0.038) | (0.010) | (0.015) |
| Max_D_Mark | 0.048** | 0.020 | 0.027 | 0.005 | 0.048 | 0.002 |
| | (0.020) | (0.037) | (0.025) | (0.068) | (0.033) | (0.043) |
| Contract_Type | 0.035** | 0.035** | 0.010 | 0.042 | 0.035** | 0.031* |
| | (0.014) | (0.017) | (0.031) | (0.061) | (0.015) | (0.018) |
| Work_Climate | 0.006 | -0.007 | 0.003 | -0.027 | 0.007 | -0.002 |
| | (0.006) | (0.009) | (0.009) | (0.024) | (0.007) | (0.010) |
| Work_Stab | 0.007 | -0.003 | 0.011 | -0.004 | 0.006 | -0.001 |
| | (0.005) | (0.007) | (0.008) | (0.026) | (0.006) | (0.007) |
| Work_Time | 0.012** | 0.028*** | -0.003 | 0.046* | 0.022*** | 0.024*** |
| | (0.005) | (0.008) | (0.009) | (0.023) | (0.007) | (0.009) |
| Work_Task | 0.020*** | 0.028*** | 0.023*** | 0.048** | 0.016** | 0.023** |
| | (0.006) | (0.009) | (0.008) | (0.024) | (0.008) | (0.010) |
| Intermed_Prof | 0.030*** | 0.017 | -0.003 | -0.053 | 0.051*** | 0.028* |
| | (0.010) | (0.014) | (0.019) | (0.046) | (0.012) | (0.015) |
| Manager | 0.251*** | 0.097*** | 0.268*** | 0.077 | 0.200*** | 0.088** |
| | (0.017) | (0.031) | (0.025) | (0.063) | (0.024) | (0.037) |
| North | 0.047*** | 0.086*** | -0.012 | -0.049 | 0.087*** | 0.111*** |
| | (0.009) | (0.015) | (0.013) | (0.037) | (0.011) | (0.016) |
| Centre | 0.042*** | 0.060*** | 0.023 | -0.022 | 0.065*** | 0.079*** |
| | (0.011) | (0.018) | (0.016) | (0.047) | (0.015) | (0.020) |
| Home_Time | 0.007*** | 0.014*** | 0.005*** | 0.008 | 0.008*** | 0.015*** |
| | (0.001) | (0.003) | (0.002) | (0.007) | (0.001) | (0.003) |
| Married | 0.036*** | 0.086*** | 0.016 | 0.147*** | 0.045*** | 0.070*** |
| | (0.010) | (0.019) | (0.014) | (0.044) | (0.013) | (0.021) |
| Italian | 0.019 | -0.004 | 0.290 | | 0.009 | -0.002 |
| | (0.059) | (0.086) | (0.232) | | (0.061) | (0.086) |
| Homeowner | 0.017* | -0.035** | 0.013 | -0.050 | 0.017 | -0.030 |
| | (0.010) | (0.017) | (0.016) | (0.042) | (0.013) | (0.018) |
| Educ_Fath_Uni | 0.017 | -0.012 | -0.005 | 0.010 | 0.034 | -0.010 |
| | (0.019) | (0.033) | (0.026) | (0.063) | (0.029) | (0.040) |
| Educ_Moth_Uni | -0.004 | 0.013 | -0.030 | -0.094 | 0.020 | 0.035 |
| | (0.026) | (0.038) | (0.039) | (0.081) | (0.035) | (0.042) |
| Constant | 0.788*** | 0.791*** | 0.677*** | 0.700*** | 0.788*** | 0.797*** |
| | (0.075) | (0.120) | (0.246) | (0.201) | (0.085) | (0.123) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9,520 | 3,637 | 3,679 | 576 | 5,841 | 3,061 |
| R-squared | 0.442 | 0.238 | 0.385 | 0.250 | 0.314 | 0.171 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.4 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2006

| Year | 2006 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.060*** | 0.102*** | | | | |
| | (0.014) | (0.033) | | | | |
| Female | -0.096*** | -0.067*** | -0.043*** | 0.033 | -0.101*** | -0.070*** |
| | (0.010) | (0.015) | (0.013) | (0.042) | (0.011) | (0.015) |
| Contfem | 0.047*** | 0.059 | | | | |
| | (0.015) | (0.037) | | | | |
| Exper | 0.018*** | 0.021*** | 0.016*** | 0.039** | 0.018*** | 0.019*** |
| | (0.002) | (0.005) | (0.003) | (0.016) | (0.002) | (0.005) |
| Exper2 | -0.000*** | -0.000 | -0.000*** | -0.002** | -0.000*** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.004* | 0.002* | -0.000 | 0.004*** | 0.005** |
| | (0.001) | (0.002) | (0.001) | (0.006) | (0.001) | (0.002) |
| Educ | 0.137*** | 0.088*** | 0.185*** | 0.101*** | 0.112*** | 0.089*** |
| | (0.008) | (0.013) | (0.012) | (0.037) | (0.010) | (0.014) |
| Max_D_Mark | 0.034 | 0.056 | -0.022 | 0.005 | 0.093** | 0.066 |
| | (0.021) | (0.040) | (0.025) | (0.065) | (0.038) | (0.050) |
| Contract_Type | 0.041*** | 0.045*** | -0.011 | -0.024 | 0.048*** | 0.050*** |
| | (0.014) | (0.016) | (0.032) | (0.051) | (0.015) | (0.017) |
| Work_Climate | -0.001 | -0.010 | -0.003 | -0.009 | 0.001 | -0.008 |
| | (0.006) | (0.010) | (0.009) | (0.028) | (0.007) | (0.010) |
| Work_Stab | 0.006 | 0.007 | 0.017* | 0.034 | 0.003 | 0.005 |
| | (0.005) | (0.008) | (0.010) | (0.025) | (0.006) | (0.008) |
| Work_Time | 0.017*** | 0.026*** | 0.009 | 0.017 | 0.019*** | 0.025*** |
| | (0.006) | (0.008) | (0.010) | (0.027) | (0.007) | (0.008) |
| Work_Task | 0.020*** | 0.025** | 0.020** | 0.034 | 0.020*** | 0.022** |
| | (0.006) | (0.010) | (0.009) | (0.025) | (0.007) | (0.010) |
| Intermed_Prof | 0.045*** | 0.049*** | 0.002 | 0.008 | 0.071*** | 0.056*** |
| | (0.010) | (0.015) | (0.018) | (0.043) | (0.012) | (0.016) |
| Manager | 0.198*** | 0.082*** | 0.177*** | 0.127** | 0.177*** | 0.068** |
| | (0.016) | (0.029) | (0.024) | (0.057) | (0.022) | (0.035) |
| North | 0.017** | 0.053*** | -0.038*** | -0.039 | 0.054*** | 0.069*** |
| | (0.009) | (0.015) | (0.013) | (0.036) | (0.012) | (0.016) |
| Centre | -0.005 | 0.032* | -0.038** | -0.090* | 0.025* | 0.052*** |
| | (0.011) | (0.016) | (0.017) | (0.051) | (0.013) | (0.017) |
| Home_Time | 0.007*** | 0.014*** | 0.005*** | 0.001 | 0.008*** | 0.016*** |
| | (0.001) | (0.003) | (0.001) | (0.007) | (0.001) | (0.003) |
| Married | 0.043*** | 0.026 | 0.034** | 0.070 | 0.044*** | 0.021 |
| | (0.010) | (0.018) | (0.014) | (0.045) | (0.014) | (0.020) |
| Italian | 0.078 | 0.032 | -0.247*** | | 0.097* | 0.040 |
| | (0.050) | (0.056) | (0.087) | | (0.050) | (0.056) |
| Homeowner | 0.026*** | 0.019 | 0.013 | -0.020 | 0.030** | 0.023 |
| | (0.010) | (0.018) | (0.016) | (0.048) | (0.012) | (0.019) |
| Educ_Fath_Uni | 0.026 | -0.007 | 0.027 | -0.029 | 0.027 | 0.004 |
| | (0.018) | (0.029) | (0.024) | (0.057) | (0.026) | (0.034) |
| Educ_Moth_Uni | 0.017 | 0.018 | 0.019 | -0.025 | 0.003 | 0.022 |
| | (0.022) | (0.029) | (0.035) | (0.074) | (0.029) | (0.031) |
| Constant | 0.858*** | 0.942*** | 1.257*** | 1.085*** | 0.859*** | 0.917*** |
| | (0.066) | (0.096) | (0.114) | (0.228) | (0.073) | (0.100) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9,305 | 3,638 | 3,482 | 484 | 5,823 | 3,154 |
| R-squared | 0.427 | 0.204 | 0.351 | 0.246 | 0.292 | 0.133 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.5 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2008

| Year | 2008 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.056*** | 0.134*** | | | | |
| | (0.013) | (0.032) | | | | |
| Female | -0.111*** | -0.089*** | -0.066*** | -0.014 | -0.115*** | -0.094*** |
| | (0.010) | (0.014) | (0.014) | (0.043) | (0.010) | (0.014) |
| Contfem | 0.037** | 0.027 | | | | |
| | (0.016) | (0.039) | | | | |
| Exper | 0.017*** | 0.010 | 0.012*** | -0.013 | 0.017*** | 0.012 |
| | (0.002) | (0.007) | (0.003) | (0.015) | (0.002) | (0.008) |
| Exper2 | -0.000*** | -0.000 | -0.000 | 0.001 | -0.000*** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.007** | 0.001 | 0.002 | 0.004*** | 0.008** |
| | (0.001) | (0.003) | (0.001) | (0.006) | (0.001) | (0.003) |
| Educ | 0.146*** | 0.117*** | 0.209*** | 0.075 | 0.119*** | 0.123*** |
| | (0.008) | (0.013) | (0.013) | (0.046) | (0.009) | (0.014) |
| Max_D_Mark | 0.056** | 0.085*** | 0.008 | 0.058 | 0.088*** | 0.091** |
| | (0.022) | (0.033) | (0.029) | (0.065) | (0.032) | (0.036) |
| Contract_Type | 0.029** | 0.036** | 0.034 | -0.018 | 0.025* | 0.043*** |
| | (0.013) | (0.015) | (0.034) | (0.050) | (0.015) | (0.016) |
| Work_Climate | -0.011** | -0.008 | -0.016* | -0.016 | -0.008 | -0.007 |
| | (0.005) | (0.009) | (0.009) | (0.027) | (0.006) | (0.009) |
| Work_Stab | 0.019*** | 0.016* | 0.008 | 0.029 | 0.023*** | 0.014 |
| | (0.005) | (0.008) | (0.009) | (0.030) | (0.006) | (0.009) |
| Work_Time | 0.023*** | 0.030*** | 0.018* | 0.041** | 0.024*** | 0.029*** |
| | (0.005) | (0.008) | (0.009) | (0.021) | (0.007) | (0.009) |
| Work_Task | 0.011* | 0.013 | 0.008 | -0.033 | 0.012* | 0.018* |
| | (0.005) | (0.009) | (0.009) | (0.026) | (0.007) | (0.009) |
| Intermed_Prof | 0.050*** | 0.023 | 0.027 | 0.039 | 0.065*** | 0.025 |
| | (0.010) | (0.017) | (0.020) | (0.062) | (0.012) | (0.018) |
| Manager | 0.138*** | 0.006 | 0.134*** | 0.087 | 0.126*** | -0.008 |
| | (0.017) | (0.032) | (0.025) | (0.082) | (0.025) | (0.037) |
| North | 0.023** | 0.072*** | -0.029** | -0.073 | 0.053*** | 0.090*** |
| | (0.009) | (0.016) | (0.014) | (0.049) | (0.012) | (0.017) |
| Centre | 0.013 | 0.064*** | -0.027 | -0.017 | 0.040*** | 0.074*** |
| | (0.011) | (0.020) | (0.017) | (0.052) | (0.015) | (0.022) |
| Home_Time | 0.005*** | 0.013*** | 0.002 | 0.003 | 0.007*** | 0.014*** |
| | (0.001) | (0.003) | (0.002) | (0.008) | (0.001) | (0.003) |
| Married | 0.049*** | 0.051*** | 0.011 | 0.048 | 0.066*** | 0.059*** |
| | (0.009) | (0.014) | (0.015) | (0.038) | (0.012) | (0.016) |
| Italian | 0.041 | 0.021 | 0.047 | -0.059 | 0.047 | 0.024 |
| | (0.055) | (0.073) | (0.110) | (0.264) | (0.059) | (0.078) |
| Homeowner | 0.035*** | 0.038** | 0.011 | -0.000 | 0.045*** | 0.041** |
| | (0.012) | (0.019) | (0.024) | (0.061) | (0.014) | (0.020) |
| Educ_Fath_Uni | 0.038** | 0.009 | 0.027 | -0.120 | 0.041 | 0.031 |
| | (0.019) | (0.031) | (0.025) | (0.074) | (0.027) | (0.035) |
| Educ_Moth_Uni | -0.007 | 0.009 | 0.051 | 0.086 | -0.036 | -0.009 |
| | (0.026) | (0.033) | (0.044) | (0.100) | (0.031) | (0.034) |
| Constant | 0.972*** | 0.978*** | 1.152*** | 1.886*** | 0.959*** | 0.906*** |
| | (0.068) | (0.103) | (0.143) | (0.440) | (0.075) | (0.107) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | | |
| Observations | 8,601 | 3,409 | 2,978 | 394 | 5,623 | 3,015 |
| R-squared | 0.426 | 0.225 | 0.335 | 0.282 | 0.302 | 0.163 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.6 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2010

| Year | 2010 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.076*** | 0.112*** | | | | |
| | (0.015) | (0.040) | | | | |
| Female | -0.093*** | -0.043** | -0.068*** | -0.030 | -0.089*** | -0.040** |
| | (0.012) | (0.018) | (0.016) | (0.053) | (0.013) | (0.018) |
| Contfem | 0.031* | 0.038 | | | | |
| | (0.018) | (0.045) | | | | |
| Exper | 0.020*** | 0.018** | 0.019*** | -0.005 | 0.019*** | 0.020** |
| | (0.002) | (0.008) | (0.003) | (0.015) | (0.002) | (0.009) |
| Exper2 | -0.000*** | -0.000 | -0.000*** | -0.000 | -0.000*** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.001) |
| Tenure | 0.003*** | 0.012*** | 0.003*** | 0.016*** | 0.004*** | 0.012*** |
| | (0.001) | (0.003) | (0.001) | (0.006) | (0.001) | (0.003) |
| Educ | 0.140*** | 0.090*** | 0.184*** | 0.079* | 0.122*** | 0.093*** |
| | (0.010) | (0.018) | (0.015) | (0.043) | (0.013) | (0.020) |
| Max_D_Mark | 0.070*** | 0.087*** | 0.056** | 0.089* | 0.066** | 0.086** |
| | (0.020) | (0.032) | (0.025) | (0.046) | (0.032) | (0.040) |
| Contract_Type | 0.035** | 0.031* | 0.011 | 0.002 | 0.035** | 0.033* |
| | (0.016) | (0.018) | (0.038) | (0.057) | (0.017) | (0.020) |
| Work_Climate | -0.004 | -0.006 | 0.003 | -0.002 | -0.008 | -0.006 |
| | (0.006) | (0.012) | (0.009) | (0.022) | (0.009) | (0.013) |
| Work_Stab | 0.002 | 0.002 | -0.008 | 0.013 | 0.006 | 0.001 |
| | (0.006) | (0.010) | (0.010) | (0.026) | (0.007) | (0.011) |
| Work_Time | 0.020*** | 0.028** | 0.005 | 0.013 | 0.026*** | 0.030** |
| | (0.007) | (0.012) | (0.010) | (0.023) | (0.009) | (0.013) |
| Work_Task | 0.012* | -0.002 | 0.011 | -0.055* | 0.012 | 0.005 |
| | (0.007) | (0.013) | (0.010) | (0.030) | (0.009) | (0.014) |
| Intermed_Prof | 0.029** | 0.000 | -0.001 | 0.068 | 0.046*** | -0.002 |
| | (0.012) | (0.019) | (0.021) | (0.048) | (0.014) | (0.020) |
| Manager | 0.122*** | 0.038 | 0.108*** | 0.088 | 0.113*** | 0.032 |
| | (0.016) | (0.029) | (0.025) | (0.057) | (0.022) | (0.034) |
| North | 0.035*** | 0.067*** | -0.002 | 0.009 | 0.054*** | 0.075*** |
| | (0.011) | (0.020) | (0.015) | (0.043) | (0.015) | (0.022) |
| Centre | 0.007 | 0.018 | 0.000 | -0.010 | 0.016 | 0.020 |
| | (0.012) | (0.022) | (0.018) | (0.054) | (0.017) | (0.024) |
| Home_Time | 0.006*** | 0.013*** | 0.004** | 0.012* | 0.007*** | 0.012*** |
| | (0.001) | (0.003) | (0.002) | (0.007) | (0.001) | (0.003) |
| Married | 0.057*** | 0.040** | 0.049*** | 0.086** | 0.057*** | 0.033 |
| | (0.010) | (0.018) | (0.015) | (0.040) | (0.014) | (0.020) |
| Italian | 0.007 | 0.062 | 0.156 | 0.131 | -0.001 | 0.059 |
| | (0.037) | (0.044) | (0.198) | (0.359) | (0.038) | (0.044) |
| Homeowner | 0.045*** | 0.036 | 0.074** | 0.060 | 0.037** | 0.032 |
| | (0.015) | (0.024) | (0.031) | (0.076) | (0.018) | (0.025) |
| Educ_Fath_Uni | -0.020 | -0.066** | -0.006 | -0.077 | -0.040 | -0.063* |
| | (0.020) | (0.032) | (0.026) | (0.064) | (0.029) | (0.037) |
| Educ_Moth_Uni | 0.013 | 0.026 | 0.028 | 0.104 | 0.009 | 0.006 |
| | (0.028) | (0.037) | (0.039) | (0.077) | (0.036) | (0.042) |
| Constant | 1.017*** | 1.018*** | 0.980*** | 1.350*** | 1.039*** | 0.979*** |
| | (0.058) | (0.103) | (0.217) | (0.451) | (0.069) | (0.110) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9,420 | 4,080 | 3,037 | 517 | 6,383 | 3,563 |
| R-squared | 0.315 | 0.117 | 0.316 | 0.242 | 0.183 | 0.075 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.7 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2011

| Year | 2011 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.101*** | 0.111*** | | | | |
| | (0.015) | (0.031) | | | | |
| Female | -0.079*** | -0.024 | -0.058*** | 0.108** | -0.080*** | -0.023 |
| | (0.013) | (0.019) | (0.015) | (0.051) | (0.013) | (0.019) |
| Contfem | 0.023 | 0.105*** | | | | |
| | (0.019) | (0.039) | | | | |
| Exper | 0.012*** | 0.014* | 0.009*** | 0.017 | 0.013*** | 0.013 |
| | (0.002) | (0.008) | (0.003) | (0.015) | (0.002) | (0.008) |
| Exper2 | -0.000*** | -0.001 | -0.000 | -0.001 | -0.000*** | -0.001 |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Tenure | 0.004*** | 0.008*** | 0.003*** | 0.009 | 0.004*** | 0.007** |
| | (0.001) | (0.003) | (0.001) | (0.007) | (0.001) | (0.003) |
| Educ | 0.162*** | 0.074*** | 0.216*** | 0.116** | 0.139*** | 0.072*** |
| | (0.010) | (0.019) | (0.016) | (0.047) | (0.012) | (0.020) |
| Max_D_Mark | 0.025 | 0.022 | 0.047* | 0.080 | -0.015 | 0.004 |
| | (0.021) | (0.035) | (0.025) | (0.055) | (0.035) | (0.042) |
| Contract_Type | 0.043*** | 0.031 | 0.015 | -0.005 | 0.044** | 0.037* |
| | (0.016) | (0.019) | (0.033) | (0.061) | (0.018) | (0.020) |
| Work_Climate | 0.006 | 0.004 | 0.010 | 0.057** | 0.005 | -0.004 |
| | (0.007) | (0.013) | (0.009) | (0.027) | (0.009) | (0.014) |
| Work_Stab | -0.002 | -0.001 | 0.005 | 0.011 | -0.004 | -0.000 |
| | (0.006) | (0.010) | (0.010) | (0.028) | (0.007) | (0.010) |
| Work_Time | 0.007 | 0.019 | 0.006 | -0.012 | 0.008 | 0.023* |
| | (0.007) | (0.012) | (0.010) | (0.026) | (0.009) | (0.013) |
| Work_Task | 0.003 | 0.000 | 0.011 | -0.038 | -0.001 | 0.004 |
| | (0.006) | (0.011) | (0.009) | (0.026) | (0.008) | (0.012) |
| Intermed_Prof | 0.022* | 0.033* | 0.008 | -0.048 | 0.036** | 0.039* |
| | (0.012) | (0.019) | (0.022) | (0.050) | (0.014) | (0.020) |
| Manager | 0.106*** | 0.081*** | 0.121*** | -0.023 | 0.085*** | 0.088*** |
| | (0.016) | (0.029) | (0.025) | (0.058) | (0.021) | (0.032) |
| North | 0.034*** | 0.078*** | -0.008 | -0.021 | 0.056*** | 0.094*** |
| | (0.011) | (0.020) | (0.016) | (0.046) | (0.015) | (0.022) |
| Centre | 0.023* | 0.033 | 0.026 | -0.041 | 0.027 | 0.045* |
| | (0.013) | (0.023) | (0.018) | (0.052) | (0.017) | (0.025) |
| Home_Time | 0.008*** | 0.012*** | 0.007*** | 0.009 | 0.009*** | 0.012*** |
| | (0.001) | (0.003) | (0.002) | (0.007) | (0.001) | (0.003) |
| Married | 0.054*** | 0.067*** | 0.058*** | 0.114*** | 0.051*** | 0.064*** |
| | (0.011) | (0.021) | (0.016) | (0.040) | (0.016) | (0.025) |
| Italian | 0.158*** | 0.063 | -0.016 | -0.441*** | 0.174*** | 0.071 |
| | (0.058) | (0.084) | (0.194) | (0.143) | (0.060) | (0.086) |
| Homeowner | 0.027* | -0.006 | 0.030 | -0.062 | 0.025 | 0.001 |
| | (0.015) | (0.025) | (0.022) | (0.060) | (0.019) | (0.027) |
| Educ_Fath_Uni | 0.012 | -0.028 | 0.030 | 0.030 | -0.007 | -0.047 |
| | (0.021) | (0.034) | (0.031) | (0.060) | (0.029) | (0.040) |
| Educ_Moth_Uni | 0.006 | 0.051 | -0.035 | -0.053 | 0.036 | 0.070* |
| | (0.028) | (0.036) | (0.052) | (0.077) | (0.034) | (0.040) |
| Constant | 0.905*** | 1.198*** | 1.074*** | 1.986*** | 0.946*** | 1.161*** |
| | (0.078) | (0.124) | (0.210) | (0.265) | (0.087) | (0.131) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 8,977 | 3,651 | 2,905 | 444 | 6,072 | 3,207 |
| R-squared | 0.302 | 0.111 | 0.314 | 0.228 | 0.158 | 0.069 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

## Table 2.B.8 OLS Estimates of Log Hourly Wages with Indicator Variable *Public_Contest* and Interactive Effect *Contfem* – 2014

| Year | 2014 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variables | Full Sample including Individuals Aged 18-64 | Individuals Aged 18-34 | Individuals Hired by Public Contest and Aged 18-64 | Individuals Hired by Public Contest and Aged 18-34 | Individuals not Hired by Public Contest and Aged 18-64 | Individuals not Hired by Public Contest and Aged 18-34 |
| Public_Contest | 0.072*** | 0.119*** | | | | |
| | (0.013) | (0.030) | | | | |
| Female | -0.087*** | -0.048*** | -0.073*** | -0.027 | -0.085*** | -0.050*** |
| | (0.011) | (0.017) | (0.013) | (0.034) | (0.011) | (0.017) |
| Contfem | 0.014 | 0.019 | | | | |
| | (0.016) | (0.037) | | | | |
| Exper | 0.014*** | 0.020*** | 0.006** | 0.015 | 0.016*** | 0.023*** |
| | (0.002) | (0.007) | (0.003) | (0.014) | (0.002) | (0.008) |
| Exper2 | -0.000*** | -0.001** | -0.000 | -0.000 | -0.000*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.004 | 0.003*** | -0.009 | 0.004*** | 0.007** |
| | (0.001) | (0.003) | (0.001) | (0.006) | (0.001) | (0.003) |
| Educ | 0.122*** | 0.059*** | 0.151*** | 0.141*** | 0.110*** | 0.050*** |
| | (0.008) | (0.015) | (0.012) | (0.033) | (0.010) | (0.017) |
| Max_D_Mark | 0.006 | 0.003 | 0.032 | -0.044 | -0.025 | 0.017 |
| | (0.023) | (0.034) | (0.025) | (0.076) | (0.042) | (0.038) |
| Contract_Type | 0.053*** | 0.048** | 0.138*** | 0.202*** | 0.022 | 0.020 |
| | (0.015) | (0.021) | (0.031) | (0.052) | (0.017) | (0.023) |
| Work_Climate | -0.006 | -0.005 | -0.005 | -0.010 | -0.007 | -0.003 |
| | (0.005) | (0.011) | (0.008) | (0.024) | (0.007) | (0.012) |
| Work_Stab | 0.004 | 0.004 | -0.010 | -0.041** | 0.009 | 0.013 |
| | (0.005) | (0.010) | (0.008) | (0.020) | (0.007) | (0.011) |
| Work_Time | 0.016*** | 0.024** | 0.019** | 0.030 | 0.016** | 0.022** |
| | (0.005) | (0.010) | (0.008) | (0.023) | (0.007) | (0.011) |
| Work_Task | 0.011* | -0.005 | 0.017** | 0.016 | 0.007 | -0.010 |
| | (0.006) | (0.011) | (0.009) | (0.025) | (0.008) | (0.013) |
| Intermed_Prof | 0.039*** | 0.033* | -0.030 | 0.035 | 0.063*** | 0.037* |
| | (0.011) | (0.019) | (0.020) | (0.062) | (0.012) | (0.020) |
| Manager | 0.220*** | 0.162*** | 0.218*** | 0.203*** | 0.193*** | 0.153*** |
| | (0.015) | (0.028) | (0.024) | (0.072) | (0.020) | (0.031) |
| North | 0.021** | 0.044** | -0.019 | 0.031 | 0.045*** | 0.047** |
| | (0.009) | (0.018) | (0.014) | (0.039) | (0.013) | (0.020) |
| Centre | -0.011 | -0.019 | -0.015 | -0.006 | -0.004 | -0.022 |
| | (0.011) | (0.021) | (0.015) | (0.049) | (0.014) | (0.023) |
| Home_Time | 0.004*** | 0.001 | 0.003** | -0.002 | 0.004*** | 0.002 |
| | (0.001) | (0.003) | (0.001) | (0.006) | (0.001) | (0.003) |
| Married | 0.043*** | 0.045*** | -0.001 | -0.008 | 0.060*** | 0.060*** |
| | (0.009) | (0.016) | (0.014) | (0.033) | (0.012) | (0.018) |
| Italian | 0.040 | 0.020 | -0.099 | -0.081 | 0.060 | 0.026 |
| | (0.040) | (0.075) | (0.075) | (0.113) | (0.043) | (0.077) |
| Homeowner | 0.012 | 0.002 | 0.039* | -0.001 | 0.001 | 0.006 |
| | (0.013) | (0.019) | (0.023) | (0.041) | (0.015) | (0.020) |
| Educ_Fath_Uni | 0.022 | 0.028 | 0.006 | 0.001 | 0.030 | 0.038 |
| | (0.017) | (0.027) | (0.022) | (0.049) | (0.024) | (0.032) |
| Educ_Moth_Uni | -0.007 | 0.003 | -0.024 | -0.067 | 0.006 | 0.016 |
| | (0.022) | (0.034) | (0.032) | (0.060) | (0.030) | (0.039) |
| Constant | 1.193*** | 1.457*** | 1.422*** | 1.492*** | 1.184*** | 1.444*** |
| | (0.059) | (0.112) | (0.105) | (0.209) | (0.071) | (0.121) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 12,328 | 4,554 | 4,404 | 851 | 7,924 | 3,703 |
| R-squared | 0.262 | 0.119 | 0.264 | 0.126 | 0.157 | 0.077 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.9 OLS Estimates of Log Hourly Wages by Age and Gender – Panel

| Year | Panel | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.111*** | 0.066*** | 0.183*** | 0.115*** |
| | (0.008) | (0.007) | (0.014) | (0.017) |
| Exper | 0.015*** | 0.020*** | 0.014*** | 0.024*** |
| | (0.001) | (0.001) | (0.005) | (0.004) |
| Exper2 | -0.000*** | -0.000*** | -0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.003*** | 0.007*** | 0.006*** |
| | (0.000) | (0.000) | (0.002) | (0.002) |
| Educ | 0.155*** | 0.138*** | 0.101*** | 0.094*** |
| | (0.006) | (0.005) | (0.011) | (0.010) |
| Max_D_Mark | 0.037*** | 0.048*** | 0.050** | 0.034 |
| | (0.012) | (0.016) | (0.020) | (0.028) |
| Contract_Type | 0.014 | 0.072*** | 0.006 | 0.061*** |
| | (0.010) | (0.009) | (0.012) | (0.011) |
| Work_Climate | -0.001 | 0.002 | 0.002 | -0.009 |
| | (0.004) | (0.004) | (0.007) | (0.007) |
| Work_Stab | 0.002 | 0.010*** | -0.001 | 0.012** |
| | (0.004) | (0.004) | (0.006) | (0.006) |
| Work_Time | 0.016*** | 0.018*** | 0.031*** | 0.026*** |
| | (0.004) | (0.004) | (0.007) | (0.006) |
| Work_Task | 0.014*** | 0.013*** | 0.010 | 0.006 |
| | (0.004) | (0.004) | (0.007) | (0.007) |
| Intermed_Prof | 0.072*** | 0.017*** | 0.054*** | 0.012 |
| | (0.009) | (0.006) | (0.013) | (0.010) |
| Manager | 0.180*** | 0.155*** | 0.086*** | 0.051*** |
| | (0.012) | (0.010) | (0.021) | (0.018) |
| North | 0.008 | 0.045*** | 0.073*** | 0.061*** |
| | (0.007) | (0.006) | (0.013) | (0.010) |
| Centre | 0.005 | 0.016** | 0.056*** | 0.023* |
| | (0.008) | (0.007) | (0.014) | (0.013) |
| Home_Time | 0.006*** | 0.007*** | 0.011*** | 0.015*** |
| | (0.001) | (0.001) | (0.002) | (0.002) |
| Married | 0.043*** | 0.052*** | 0.048*** | 0.068*** |
| | (0.006) | (0.007) | (0.011) | (0.013) |
| Italian | 0.040 | 0.075* | 0.023 | 0.028 |
| | (0.034) | (0.039) | (0.049) | (0.033) |
| Homeowner | 0.026*** | 0.028*** | 0.013 | 0.015 |
| | (0.008) | (0.007) | (0.013) | (0.012) |
| Educ_Fath_Uni | 0.015 | 0.009 | -0.039* | -0.014 |
| | (0.012) | (0.013) | (0.021) | (0.020) |
| Educ_Moth_Uni | 0.048*** | -0.036** | 0.075*** | -0.021 |
| | (0.017) | (0.017) | (0.025) | (0.021) |
| Constant | 0.895*** | 0.891*** | 1.004*** | 1.046*** |
| | (0.048) | (0.047) | (0.081) | (0.062) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes | Yes |
| Observations | 19,398 | 22,447 | 8,090 | 8,366 |
| R-squared | 0.405 | 0.364 | 0.209 | 0.130 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.10 OLS Estimates of Log Hourly Wages by Age and Gender – 2005

| Year | 2005 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.072*** | 0.075*** | 0.143*** | 0.057 |
| | (0.015) | (0.016) | (0.029) | (0.038) |
| Exper | 0.017*** | 0.020*** | 0.014 | 0.024*** |
| | (0.002) | (0.002) | (0.009) | (0.006) |
| Exper2 | -0.000*** | -0.000*** | -0.000 | -0.001* |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.002* | 0.002** | 0.001 | 0.001 |
| | (0.001) | (0.001) | (0.003) | (0.003) |
| Educ | 0.182*** | 0.152*** | 0.174*** | 0.134*** |
| | (0.012) | (0.011) | (0.021) | (0.018) |
| Max_D_Mark | 0.054** | 0.028 | 0.063 | -0.103 |
| | (0.024) | (0.035) | (0.044) | (0.067) |
| Contract_Type | 0.022 | 0.050*** | 0.018 | 0.046** |
| | (0.020) | (0.019) | (0.026) | (0.023) |
| Work_Climate | 0.000 | 0.012 | -0.001 | -0.012 |
| | (0.008) | (0.008) | (0.013) | (0.014) |
| Work_Stab | 0.003 | 0.010* | -0.007 | 0.005 |
| | (0.007) | (0.006) | (0.010) | (0.010) |
| Work_Time | 0.005 | 0.019** | 0.009 | 0.048*** |
| | (0.008) | (0.007) | (0.012) | (0.011) |
| Work_Task | 0.022*** | 0.016** | 0.033** | 0.018 |
| | (0.008) | (0.008) | (0.014) | (0.012) |
| Intermed_Prof | 0.063*** | 0.015 | 0.050** | 0.004 |
| | (0.017) | (0.012) | (0.021) | (0.019) |
| Manager | 0.254*** | 0.268*** | 0.110** | 0.099** |
| | (0.026) | (0.022) | (0.044) | (0.039) |
| Sec_2 | 0.067** | 0.091*** | 0.080** | 0.043 |
| North | 0.051*** | 0.045*** | 0.150*** | 0.035* |
| | (0.014) | (0.011) | (0.024) | (0.018) |
| Centre | 0.042*** | 0.044*** | 0.119*** | 0.010 |
| | (0.016) | (0.014) | (0.027) | (0.025) |
| Home_Time | 0.008*** | 0.006*** | 0.014*** | 0.015*** |
| | (0.001) | (0.001) | (0.004) | (0.003) |
| Married | 0.030** | 0.040*** | 0.071*** | 0.116*** |
| | (0.013) | (0.014) | (0.026) | (0.028) |
| Italian | 0.017 | 0.024 | -0.001 | -0.024 |
| | (0.089) | (0.049) | (0.141) | (0.041) |
| Homeowner | 0.007 | 0.028** | -0.040 | -0.021 |
| | (0.014) | (0.014) | (0.025) | (0.023) |
| Educ_Fath_Uni | 0.025 | 0.006 | -0.007 | -0.014 |
| | (0.025) | (0.029) | (0.045) | (0.050) |
| Educ_Moth_Uni | 0.015 | -0.020 | 0.045 | 0.009 |
| | (0.037) | (0.037) | (0.057) | (0.050) |
| Constant | 0.667*** | 0.772*** | 0.593*** | 0.865*** |
| | (0.117) | (0.073) | (0.196) | (0.101) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 4,513 | 5,007 | 1,877 | 1,760 |
| R-squared | 0.456 | 0.431 | 0.292 | 0.193 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.11 OLS Estimates of Log Hourly Wages by Age and Gender – 2006

| Year | 2006 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.114*** | 0.054*** | 0.160*** | 0.100*** |
| | (0.017) | (0.016) | (0.031) | (0.037) |
| Exper | 0.018*** | 0.018*** | 0.020*** | 0.022*** |
| | (0.002) | (0.002) | (0.007) | (0.006) |
| Exper2 | -0.000*** | -0.000*** | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.002** | 0.004*** | 0.003 | 0.005 |
| | (0.001) | (0.001) | (0.003) | (0.003) |
| Educ | 0.132*** | 0.138*** | 0.086*** | 0.088*** |
| | (0.012) | (0.010) | (0.017) | (0.018) |
| Max_D_Mark | 0.037 | 0.025 | 0.101** | -0.041 |
| | (0.024) | (0.037) | (0.041) | (0.087) |
| Contract_Type | 0.030 | 0.060*** | 0.055** | 0.039* |
| | (0.019) | (0.019) | (0.022) | (0.023) |
| Work_Climate | -0.006 | 0.001 | -0.004 | -0.017 |
| | (0.008) | (0.008) | (0.013) | (0.014) |
| Work_Stab | 0.005 | 0.010 | 0.000 | 0.014 |
| | (0.007) | (0.007) | (0.011) | (0.012) |
| Work_Time | 0.015* | 0.019** | 0.029** | 0.023** |
| | (0.008) | (0.008) | (0.012) | (0.010) |
| Work_Task | 0.016** | 0.020** | 0.020 | 0.026* |
| | (0.008) | (0.008) | (0.013) | (0.014) |
| Intermed_Prof | 0.077*** | 0.025* | 0.075*** | 0.042** |
| | (0.017) | (0.013) | (0.021) | (0.021) |
| Manager | 0.212*** | 0.203*** | 0.141*** | 0.054 |
| | (0.026) | (0.021) | (0.044) | (0.039) |
| North | 0.005 | 0.024** | 0.067*** | 0.040** |
| | (0.014) | (0.011) | (0.023) | (0.019) |
| Centre | -0.022 | 0.006 | 0.051** | 0.014 |
| | (0.016) | (0.014) | (0.024) | (0.023) |
| Home_Time | 0.005*** | 0.009*** | 0.011*** | 0.018*** |
| | (0.001) | (0.002) | (0.004) | (0.004) |
| Married | 0.033** | 0.042*** | 0.019 | 0.045 |
| | (0.013) | (0.015) | (0.023) | (0.030) |
| Italian | 0.052 | 0.059 | 0.001 | 0.042 |
| | (0.074) | (0.051) | (0.089) | (0.066) |
| Homeowner | 0.028* | 0.022 | 0.021 | 0.018 |
| | (0.015) | (0.013) | (0.022) | (0.028) |
| Educ_Fath_Uni | 0.018 | 0.040 | -0.046 | 0.042 |
| | (0.023) | (0.027) | (0.038) | (0.047) |
| Educ_Moth_Uni | 0.037 | -0.012 | 0.030 | 0.002 |
| | (0.030) | (0.032) | (0.039) | (0.043) |
| Constant | 0.849*** | 0.832*** | 0.888*** | 0.957*** |
| | (0.099) | (0.077) | (0.143) | (0.127) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 4,083 | 5,222 | 1,779 | 1,859 |
| R-squared | 0.457 | 0.408 | 0.269 | 0.147 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.12 OLS Estimates of Log Hourly Wages by Age and Gender – 2008

| Year | 2008 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.083*** | 0.060*** | 0.164*** | 0.107*** |
| | (0.016) | (0.015) | (0.029) | (0.035) |
| Exper | 0.015*** | 0.018*** | 0.013 | 0.004 |
| | (0.002) | (0.002) | (0.008) | (0.012) |
| Exper2 | -0.000*** | -0.000*** | -0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.001) |
| Tenure | 0.004*** | 0.002*** | 0.009** | 0.006 |
| | (0.001) | (0.001) | (0.004) | (0.005) |
| Educ | 0.158*** | 0.133*** | 0.121*** | 0.110*** |
| | (0.012) | (0.010) | (0.020) | (0.017) |
| Max_D_Mark | 0.057** | 0.046 | 0.104*** | 0.034 |
| | (0.028) | (0.034) | (0.037) | (0.066) |
| Contract_Type | 0.006 | 0.054*** | 0.003 | 0.068*** |
| | (0.019) | (0.019) | (0.022) | (0.022) |
| Work_Climate | -0.008 | -0.017** | 0.000 | -0.017 |
| | (0.007) | (0.007) | (0.012) | (0.012) |
| Work_Stab | 0.013* | 0.029*** | 0.007 | 0.030*** |
| | (0.007) | (0.007) | (0.012) | (0.011) |
| Work_Time | 0.032*** | 0.015* | 0.040*** | 0.019 |
| | (0.007) | (0.008) | (0.010) | (0.013) |
| Work_Task | 0.006 | 0.015** | 0.008 | 0.017 |
| | (0.008) | (0.007) | (0.013) | (0.013) |
| Intermed_Prof | 0.091*** | 0.031** | 0.050* | 0.011 |
| | (0.019) | (0.013) | (0.028) | (0.022) |
| Manager | 0.152*** | 0.146*** | 0.021 | 0.010 |
| | (0.029) | (0.022) | (0.050) | (0.043) |
| North | 0.016 | 0.032*** | 0.104*** | 0.044** |
| | (0.014) | (0.012) | (0.023) | (0.021) |
| Centre | 0.015 | 0.015 | 0.105*** | 0.029 |
| | (0.018) | (0.015) | (0.028) | (0.029) |
| Home_Time | 0.005*** | 0.005*** | 0.014*** | 0.011*** |
| | (0.001) | (0.001) | (0.004) | (0.004) |
| Married | 0.039*** | 0.052*** | 0.036** | 0.086*** |
| | (0.012) | (0.015) | (0.018) | (0.025) |
| Italian | 0.074 | 0.009 | 0.079 | -0.028 |
| | (0.085) | (0.058) | (0.117) | (0.069) |
| Homeowner | 0.034* | 0.037** | 0.029 | 0.050* |
| | (0.019) | (0.016) | (0.023) | (0.030) |
| Educ_Fath_Uni | 0.032 | 0.045* | 0.005 | 0.019 |
| | (0.027) | (0.026) | (0.043) | (0.045) |
| Educ_Moth_Uni | 0.081** | -0.103*** | 0.089* | -0.078* |
| | (0.038) | (0.034) | (0.049) | (0.044) |
| Constant | 0.773*** | 1.026*** | 0.744*** | 1.089*** |
| | (0.109) | (0.078) | (0.157) | (0.129) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 3,956 | 4,645 | 1,742 | 1,667 |
| R-squared | 0.469 | 0.397 | 0.300 | 0.157 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.13 OLS Estimates of Log Hourly Wages by Age and Gender – 2010

| Year | 2010 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.108*** | 0.079*** | 0.156*** | 0.106** |
| | (0.017) | (0.017) | (0.030) | (0.042) |
| Exper | 0.016*** | 0.024*** | 0.012 | 0.024** |
| | (0.003) | (0.003) | (0.013) | (0.009) |
| Exper2 | -0.000*** | -0.000*** | -0.000 | -0.001 |
| | (0.000) | (0.000) | (0.001) | (0.001) |
| Tenure | 0.003*** | 0.003*** | 0.017*** | 0.007* |
| | (0.001) | (0.001) | (0.005) | (0.004) |
| Educ | 0.143*** | 0.133*** | 0.098*** | 0.086*** |
| | (0.015) | (0.013) | (0.030) | (0.023) |
| Max_D_Mark | 0.040 | 0.105*** | 0.045 | 0.162*** |
| | (0.026) | (0.030) | (0.041) | (0.052) |
| Contract_Type | 0.005 | 0.069*** | -0.019 | 0.079*** |
| | (0.023) | (0.021) | (0.029) | (0.023) |
| Work_Climate | -0.009 | -0.000 | -0.009 | -0.005 |
| | (0.009) | (0.009) | (0.018) | (0.017) |
| Work_Stab | -0.001 | 0.008 | -0.003 | 0.011 |
| | (0.009) | (0.008) | (0.016) | (0.013) |
| Work_Time | 0.015 | 0.025*** | 0.030 | 0.026* |
| | (0.011) | (0.009) | (0.019) | (0.015) |
| Work_Task | 0.018* | 0.006 | 0.004 | -0.009 |
| | (0.010) | (0.010) | (0.018) | (0.018) |
| Intermed_Prof | 0.058*** | 0.012 | 0.038 | -0.020 |
| | (0.022) | (0.014) | (0.033) | (0.022) |
| Manager | 0.156*** | 0.107*** | 0.087* | 0.006 |
| | (0.028) | (0.021) | (0.045) | (0.040) |
| North | 0.015 | 0.051*** | 0.067** | 0.069*** |
| | (0.017) | (0.014) | (0.033) | (0.024) |
| Centre | 0.007 | 0.004 | 0.024 | 0.013 |
| | (0.018) | (0.017) | (0.033) | (0.030) |
| Home_Time | 0.005*** | 0.006*** | 0.010** | 0.016*** |
| | (0.001) | (0.001) | (0.005) | (0.004) |
| Married | 0.061*** | 0.040** | 0.049** | 0.044* |
| | (0.014) | (0.016) | (0.025) | (0.026) |
| Italian | 0.020 | -0.029 | 0.075 | 0.003 |
| | (0.052) | (0.055) | (0.066) | (0.056) |
| Homeowner | 0.026 | 0.061*** | -0.000 | 0.075** |
| | (0.023) | (0.021) | (0.034) | (0.034) |
| Educ_Fath_Uni | 0.003 | -0.041 | -0.073 | -0.067 |
| | (0.028) | (0.027) | (0.047) | (0.043) |
| Educ_Moth_Uni | -0.024 | 0.051 | -0.001 | 0.069 |
| | (0.039) | (0.041) | (0.051) | (0.055) |
| Constant | 0.981*** | 0.978*** | 0.997*** | 1.003*** |
| | (0.084) | (0.083) | (0.157) | (0.133) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 4,209 | 5,211 | 1,913 | 2,167 |
| R-squared | 0.333 | 0.309 | 0.140 | 0.103 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.14 OLS Estimates of Log Hourly Wages by Age and Gender – 2011

| Year | 2011 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.138*** | 0.087*** | 0.219*** | 0.105*** |
| | (0.019) | (0.015) | (0.035) | (0.032) |
| Exper | 0.005* | 0.018*** | -0.003 | 0.029*** |
| | (0.003) | (0.002) | (0.013) | (0.008) |
| Exper2 | -0.000 | -0.000*** | 0.000 | -0.001*** |
| | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure | 0.005*** | 0.003*** | 0.012*** | 0.003 |
| | (0.001) | (0.001) | (0.004) | (0.003) |
| Educ | 0.149*** | 0.163*** | 0.041 | 0.094*** |
| | (0.016) | (0.013) | (0.030) | (0.024) |
| Max_D_Mark | 0.013 | 0.025 | -0.008 | 0.053 |
| | (0.026) | (0.038) | (0.046) | (0.053) |
| Contract_Type | 0.017 | 0.070*** | -0.006 | 0.058** |
| | (0.024) | (0.021) | (0.033) | (0.023) |
| Work_Climate | 0.006 | 0.008 | 0.019 | -0.007 |
| | (0.009) | (0.010) | (0.019) | (0.018) |
| Work_Stab | -0.004 | 0.005 | -0.004 | 0.005 |
| | (0.009) | (0.008) | (0.016) | (0.011) |
| Work_Time | -0.002 | 0.010 | 0.009 | 0.024 |
| | (0.010) | (0.010) | (0.017) | (0.017) |
| Work_Task | -0.000 | 0.004 | 0.007 | -0.008 |
| | (0.009) | (0.009) | (0.018) | (0.014) |
| Intermed_Prof | 0.072*** | 0.000 | 0.089** | 0.013 |
| | (0.023) | (0.015) | (0.036) | (0.022) |
| Manager | 0.171*** | 0.070*** | 0.150*** | 0.039 |
| | (0.028) | (0.021) | (0.049) | (0.036) |
| North | 0.004 | 0.060*** | 0.031 | 0.113*** |
| | (0.018) | (0.014) | (0.034) | (0.023) |
| Centre | 0.023 | 0.025 | 0.027 | 0.041 |
| | (0.020) | (0.016) | (0.040) | (0.027) |
| Home_Time | 0.006*** | 0.010*** | 0.013*** | 0.011*** |
| | (0.002) | (0.001) | (0.005) | (0.004) |
| Married | 0.041*** | 0.062*** | 0.057** | 0.096*** |
| | (0.015) | (0.018) | (0.028) | (0.035) |
| Italian | 0.167*** | 0.113 | 0.177** | -0.090 |
| | (0.060) | (0.132) | (0.085) | (0.173) |
| Homeowner | 0.034 | 0.021 | 0.003 | -0.015 |
| | (0.023) | (0.020) | (0.040) | (0.029) |
| Educ_Fath_Uni | 0.027 | -0.002 | -0.051 | -0.022 |
| | (0.033) | (0.027) | (0.062) | (0.036) |
| Educ_Moth_Uni | 0.038 | -0.017 | 0.120* | 0.013 |
| | (0.045) | (0.035) | (0.063) | (0.040) |
| Constant | 0.987*** | 0.832*** | 1.211*** | 1.251*** |
| | (0.100) | (0.150) | (0.161) | (0.214) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 4,168 | 4,809 | 1,617 | 2,034 |
| R-squared | 0.303 | 0.316 | 0.137 | 0.101 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.15 OLS Estimates of Log Hourly Wages by Age and Gender – 2014

| Year | 2014 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Aged 18-64 | | Individuals Aged 18-34 | |
| Variables | Women | Men | Women | Men |
| Public_Contest | 0.086*** | 0.076*** | 0.142*** | 0.123*** |
| | (0.013) | (0.014) | (0.026) | (0.032) |
| Exper | 0.009*** | 0.019*** | 0.022** | 0.017* |
| | (0.002) | (0.002) | (0.010) | (0.009) |
| Exper2 | -0.000** | -0.000*** | -0.001** | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.003*** | -0.001 | 0.009** |
| | (0.001) | (0.001) | (0.004) | (0.004) |
| Educ | 0.112*** | 0.127*** | 0.049** | 0.071*** |
| | (0.012) | (0.011) | (0.023) | (0.020) |
| Max_D_Mark | 0.032 | -0.041 | 0.007 | -0.006 |
| | (0.023) | (0.053) | (0.043) | (0.052) |
| Contract_Type | 0.038* | 0.072*** | 0.043 | 0.048* |
| | (0.022) | (0.021) | (0.031) | (0.027) |
| Work_Climate | -0.005 | -0.008 | -0.012 | 0.001 |
| | (0.008) | (0.007) | (0.017) | (0.014) |
| Work_Stab | -0.002 | 0.012 | -0.005 | 0.015 |
| | (0.007) | (0.008) | (0.013) | (0.014) |
| Work_Time | 0.027*** | 0.006 | 0.043*** | 0.006 |
| | (0.008) | (0.007) | (0.015) | (0.012) |
| Work_Task | 0.008 | 0.013 | 0.001 | -0.010 |
| | (0.009) | (0.008) | (0.019) | (0.014) |
| Intermed_Prof | 0.066*** | 0.025* | 0.041 | 0.031 |
| | (0.018) | (0.013) | (0.032) | (0.024) |
| Manager | 0.279*** | 0.170*** | 0.181*** | 0.149*** |
| | (0.025) | (0.020) | (0.047) | (0.033) |
| North | -0.008 | 0.055*** | 0.042 | 0.049** |
| | (0.014) | (0.013) | (0.026) | (0.023) |
| Centre | -0.036** | 0.018 | -0.023 | -0.009 |
| | (0.015) | (0.015) | (0.032) | (0.028) |
| Home_Time | 0.003** | 0.005*** | -0.003 | 0.005 |
| | (0.001) | (0.001) | (0.004) | (0.003) |
| Married | 0.029** | 0.054*** | 0.050** | 0.051** |
| | (0.013) | (0.013) | (0.022) | (0.023) |
| Italian | 0.025 | 0.051 | -0.009 | 0.068 |
| | (0.056) | (0.051) | (0.103) | (0.102) |
| Homeowner | 0.021 | 0.001 | -0.001 | 0.009 |
| | (0.018) | (0.018) | (0.027) | (0.026) |
| Educ_Fath_Uni | -0.013 | 0.065*** | -0.009 | 0.060 |
| | (0.023) | (0.024) | (0.040) | (0.037) |
| Educ_Moth_Uni | 0.006 | -0.027 | 0.018 | -0.013 |
| | (0.031) | (0.031) | (0.051) | (0.044) |
| Constant | 1.219*** | 1.096*** | 1.517*** | 1.312*** |
| | (0.086) | (0.077) | (0.166) | (0.148) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 6,093 | 6,235 | 2,254 | 2,300 |
| R-squared | 0.284 | 0.249 | 0.136 | 0.107 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.16 Bivariate Probit Estimation by Gender – 2005

| Year | 2005 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.061*** | 0.006*** | 0.035*** | -0.009*** |
| | (0.002) | (0.002) | (0.002) | (0.003) |
| Educ | 0.791*** | 0.511*** | 0.506*** | 0.209*** |
| | (0.030) | (0.019) | (0.026) | (0.023) |
| Married | 0.151*** | -0.259*** | 0.261*** | 0.491*** |
| | (0.050) | (0.052) | (0.048) | (0.056) |
| Homeowner | 0.182*** | 0.275*** | 0.255*** | 0.456*** |
| | (0.051) | (0.037) | (0.047) | (0.047) |
| Age5064 | | 1.396*** | | 0.557*** |
| | | (0.057) | | (0.064) |
| Italian | | 0.242** | | -0.026 |
| | | (0.106) | | (0.190) |
| North | | 0.748*** | | 0.797*** |
| | | (0.030) | | (0.037) |
| Centre | | 0.421*** | | 0.420*** |
| | | (0.036) | | (0.043) |
| Partner_Works | | 0.231*** | | 0.091* |
| | | (0.043) | | (0.051) |
| Kids | | -0.469*** | | -0.008 |
| | | (0.044) | | (0.046) |
| Kids_10 | | 0.393*** | | 0.009 |
| | | (0.049) | | (0.086) |
| Work_Climate | -0.009 | | -0.028 | |
| | (0.029) | | (0.027) | |
| Work_Stab | 0.131*** | | 0.228*** | |
| | (0.021) | | (0.021) | |
| Work_Time | -0.025 | | -0.050** | |
| | (0.027) | | (0.024) | |
| Work_Task | -0.003 | | -0.030 | |
| | (0.029) | | (0.026) | |
| Reloc | 0.550*** | | 0.446*** | |
| | (0.062) | | (0.045) | |
| Risp | 0.174*** | | 0.132*** | |
| | (0.047) | | (0.045) | |
| Constant | -6.182*** | -2.560*** | -4.594*** | -0.592*** |
| | (0.184) | (0.139) | (0.148) | (0.219) |
| $\rho$ | 0.571*** | | 1.324*** | |
| | (0.079) | | (0.216) | |
| Observations | 10,744 | | 7,648 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.17 Bivariate Probit Estimation by Gender – 2006

| Year | 2006 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.062*** | 0.005* | 0.038*** | -0.008*** |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Educ | 0.823*** | 0.454*** | 0.490*** | 0.185*** |
| | (0.032) | (0.021) | (0.026) | (0.023) |
| Married | 0.021 | -0.261*** | 0.329*** | 0.367*** |
| | (0.054) | (0.056) | (0.049) | (0.065) |
| Homeowner | 0.275*** | 0.331*** | 0.131*** | 0.350*** |
| | (0.054) | (0.041) | (0.048) | (0.047) |
| Age5064 | | 0.904*** | | 0.475*** |
| | | (0.063) | | (0.066) |
| Italian | | 0.318*** | | 0.071 |
| | | (0.117) | | (0.176) |
| North | | 0.692*** | | 0.653*** |
| | | (0.032) | | (0.036) |
| Centre | | 0.472*** | | 0.468*** |
| | | (0.040) | | (0.046) |
| Partner_Works | | 0.167*** | | 0.129** |
| | | (0.045) | | (0.051) |
| Kids | | -0.121** | | 0.155** |
| | | (0.056) | | (0.065) |
| Kids_10 | | -0.255*** | | 0.129** |
| | | (0.049) | | (0.065) |
| Work_Climate | -0.059** | | -0.094*** | |
| | (0.028) | | (0.025) | |
| Work_Stab | 0.172*** | | 0.307*** | |
| | (0.022) | | (0.023) | |
| Work_Time | 0.100*** | | 0.051** | |
| | (0.028) | | (0.025) | |
| Work_Task | -0.039 | | -0.034 | |
| | (0.030) | | (0.026) | |
| Reloc | 0.360*** | | 0.510*** | |
| | (0.097) | | (0.062) | |
| Risp | 0.104** | | 0.097** | |
| | (0.050) | | (0.046) | |
| Constant | -6.546*** | -2.344*** | -4.997*** | -0.565*** |
| | (0.190) | (0.154) | (0.159) | (0.203) |
| $\rho$ | 0.786*** | | 1.339*** | |
| | (0.114) | | (0.252) | |
| Observations | 8,702 | | 7,703 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.18 Bivariate Probit Estimation by Gender – 2008

| Year | 2008 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.066*** | 0.011*** | 0.045*** | -0.005* |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Educ | 0.760*** | 0.499*** | 0.503*** | 0.243*** |
| | (0.034) | (0.022) | (0.028) | (0.025) |
| Married | 0.123** | 0.011 | 0.308*** | 0.558*** |
| | (0.049) | (0.058) | (0.053) | (0.070) |
| Homeowner | 0.211*** | 0.217*** | 0.113* | 0.164*** |
| | (0.067) | (0.042) | (0.059) | (0.047) |
| Age5064 | | 0.914*** | | 0.562*** |
| | | (0.068) | | (0.073) |
| Italian | | 0.333*** | | 0.278 |
| | | (0.120) | | (0.183) |
| North | | 0.840*** | | 0.814*** |
| | | (0.034) | | (0.039) |
| Centre | | 0.471*** | | 0.538*** |
| | | (0.041) | | (0.049) |
| Partner_Works | | -0.044 | | 0.014 |
| | | (0.049) | | (0.058) |
| Kids | | -0.168*** | | 0.298*** |
| | | (0.059) | | (0.073) |
| Kids_10 | | -0.205*** | | -0.132* |
| | | (0.052) | | (0.072) |
| Work_Climate | 0.005 | | -0.130*** | |
| | (0.027) | | (0.025) | |
| Work_Stab | 0.131*** | | 0.255*** | |
| | (0.023) | | (0.024) | |
| Work_Time | 0.120*** | | 0.124*** | |
| | (0.027) | | (0.026) | |
| Work_Task | -0.054* | | -0.011 | |
| | (0.028) | | (0.026) | |
| Reloc | 0.395*** | | 0.408*** | |
| | (0.084) | | (0.067) | |
| Risp | 0.040 | | 0.061 | |
| | (0.054) | | (0.054) | |
| Constant | -6.758*** | -2.857*** | -5.458*** | -1.220*** |
| | (0.209) | (0.159) | (0.180) | (0.208) |
| $\rho$ | 0.664*** | | 1.114*** | |
| | (0.111) | | (0.239) | |
| Observations | 8,280 | | 7,016 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.19 Bivariate Probit Estimation by Gender – 2010

| Year | 2010 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.064*** | 0.021*** | 0.042*** | 0.003 |
| | (0.002) | (0.003) | (0.002) | (0.002) |
| Educ | 0.787*** | 0.445*** | 0.512*** | 0.260*** |
| | (0.033) | (0.021) | (0.026) | (0.022) |
| Married | 0.125*** | 0.085 | 0.203*** | 0.417*** |
| | (0.046) | (0.055) | (0.046) | (0.061) |
| Homeowner | 0.144** | 0.117*** | 0.173*** | 0.206*** |
| | (0.063) | (0.040) | (0.057) | (0.043) |
| Age5064 | | 0.748*** | | 0.395*** |
| | | (0.066) | | (0.065) |
| Italian | | 0.221** | | 0.340** |
| | | (0.099) | | (0.140) |
| North | | 0.815*** | | 0.652*** |
| | | (0.032) | | (0.033) |
| Centre | | 0.466*** | | 0.397*** |
| | | (0.039) | | (0.040) |
| Partner_Works | | 0.038 | | 0.163*** |
| | | (0.047) | | (0.051) |
| Kids | | -0.251*** | | 0.253*** |
| | | (0.057) | | (0.062) |
| Kids_10 | | -0.064 | | -0.226*** |
| | | (0.051) | | (0.064) |
| Work_Climate | -0.046* | | -0.026 | |
| | (0.028) | | (0.024) | |
| Work_Stab | 0.160*** | | 0.205*** | |
| | (0.022) | | (0.021) | |
| Work_Time | 0.118*** | | 0.067*** | |
| | (0.028) | | (0.025) | |
| Work_Task | -0.054* | | -0.022 | |
| | (0.029) | | (0.027) | |
| Reloc | 0.377*** | | 0.401*** | |
| | (0.083) | | (0.052) | |
| Risp | 0.029 | | 0.116*** | |
| | (0.047) | | (0.042) | |
| Constant | -6.568*** | -2.927*** | -5.374*** | -1.623*** |
| | (0.196) | (0.136) | (0.155) | (0.166) |
| $\rho$ | 0.620*** | | 1.467*** | |
| | (0.122) | | (0.559) | |
| Observations | 9,204 | | 8,579 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.20 Bivariate Probit Estimation by Gender – 2011

| Year | 2011 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.060*** | 0.028*** | 0.044*** | 0.001 |
| | (0.003) | (0.003) | (0.002) | (0.002) |
| Educ | 0.832*** | 0.460*** | 0.559*** | 0.221*** |
| | (0.034) | (0.020) | (0.030) | (0.023) |
| Married | 0.180*** | -0.069 | 0.306*** | 0.452*** |
| | (0.048) | (0.050) | (0.053) | (0.062) |
| Homeowner | 0.063 | 0.126*** | 0.163** | 0.213*** |
| | (0.069) | (0.042) | (0.064) | (0.045) |
| Age5064 | | 0.689*** | | 0.576*** |
| | | (0.065) | | (0.067) |
| Italian | | -0.014 | | 0.344** |
| | | (0.104) | | (0.163) |
| North | | 0.698*** | | 0.664*** |
| | | (0.032) | | (0.034) |
| Centre | | 0.377*** | | 0.453*** |
| | | (0.038) | | (0.041) |
| Partner_Works | | 0.117*** | | 0.161*** |
| | | (0.041) | | (0.052) |
| Kids | | -0.155*** | | 0.254*** |
| | | (0.054) | | (0.067) |
| Kids_10 | | -0.062 | | 0.069 |
| | | (0.048) | | (0.069) |
| Work_Climate | -0.036 | | -0.008 | |
| | (0.030) | | (0.028) | |
| Work_Stab | 0.180*** | | 0.224*** | |
| | (0.023) | | (0.024) | |
| Work_Time | 0.108*** | | 0.039 | |
| | (0.032) | | (0.030) | |
| Work_Task | -0.019 | | -0.014 | |
| | (0.031) | | (0.029) | |
| Reloc | 0.360*** | | 0.426*** | |
| | (0.086) | | (0.068) | |
| Risp | 0.030 | | 0.150*** | |
| | (0.051) | | (0.051) | |
| Constant | -6.689*** | -2.958*** | -5.749*** | -1.594*** |
| | (0.242) | (0.147) | (0.181) | (0.186) |
| $\rho$ | 0.393*** | | 1.077*** | |
| | (0.132) | | (0.269) | |
| Observations | 9,347 | | 8,236 | |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.21 Bivariate Probit Estimation by Gender – 2014

| Year | 2014 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Women | | Men | |
| Variables | Public Contest | Employment | Public Contest | Employment |
| Age | 0.044*** | 0.018*** | 0.031*** | 0.010*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Educ | 0.693*** | 0.566*** | 0.536*** | 0.366*** |
| | (0.031) | (0.017) | (0.024) | (0.019) |
| Married | 0.340*** | 0.047 | 0.405*** | 0.525*** |
| | (0.038) | (0.037) | (0.040) | (0.046) |
| Homeowner | 0.226*** | 0.107*** | 0.159*** | 0.146*** |
| | (0.057) | (0.035) | (0.052) | (0.038) |
| Age5064 | | 0.457*** | | 0.201*** |
| | | (0.048) | | (0.051) |
| Italian | | 0.037 | | 0.435*** |
| | | (0.081) | | (0.098) |
| North | | 0.712*** | | 0.685*** |
| | | (0.026) | | (0.029) |
| Centre | | 0.432*** | | 0.441*** |
| | | (0.032) | | (0.035) |
| Partner_Works | | 0.162*** | | 0.155*** |
| | | (0.029) | | (0.036) |
| Kids | | -0.112*** | | 0.010 |
| | | (0.041) | | (0.049) |
| Kids_10 | | -0.048 | | 0.121** |
| | | (0.038) | | (0.050) |
| Work_Climate | -0.050** | | -0.030 | |
| | (0.022) | | (0.021) | |
| Work_Stab | 0.254*** | | 0.217*** | |
| | (0.019) | | (0.020) | |
| Work_Time | 0.051** | | 0.078*** | |
| | (0.022) | | (0.021) | |
| Work_Task | -0.017 | | -0.033 | |
| | (0.025) | | (0.023) | |
| Reloc | 0.371*** | | 0.439*** | |
| | (0.063) | | (0.050) | |
| Risp | 0.044 | | 0.112** | |
| | (0.045) | | (0.045) | |
| Constant | -5.766*** | -3.104*** | -5.184*** | -2.441*** |
| | (0.200) | (0.116) | (0.148) | (0.133) |
| $\rho$ | 0.343*** | | 1.298*** | |
| | (0.093) | | (0.295) | |
| Observations | 13,129 | | 10,584 | |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.22 OLS Estimates by Gender with Selection Variables – Panel

| Year | Panel | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.020*** | 0.019*** | 0.012*** | 0.018*** |
| | (0.002) | (0.002) | (0.002) | (0.001) |
| Exper2 | -0.000** | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.002*** | 0.003*** | 0.005*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.000) |
| Educ | 0.322*** | 0.203*** | 0.089*** | 0.101*** |
| | (0.018) | (0.017) | (0.019) | (0.011) |
| Max_D_Mark | 0.015 | 0.049*** | 0.045** | 0.021 |
| | (0.012) | (0.017) | (0.019) | (0.021) |
| Contract_Type | -0.011 | 0.065*** | 0.011 | 0.070*** |
| | (0.017) | (0.025) | (0.011) | (0.009) |
| Work_Climate | -0.008 | 0.011* | 0.006 | 0.001 |
| | (0.005) | (0.006) | (0.005) | (0.005) |
| Work_Stab | 0.024*** | 0.016* | -0.014*** | -0.001 |
| | (0.006) | (0.008) | (0.005) | (0.004) |
| Work_Time | 0.015*** | 0.004 | 0.014** | 0.022*** |
| | (0.005) | (0.006) | (0.005) | (0.004) |
| Work_Task | 0.014*** | 0.014** | 0.012** | 0.010** |
| | (0.005) | (0.006) | (0.006) | (0.005) |
| Intermed_Prof | 0.146*** | -0.043*** | 0.071*** | 0.043*** |
| | (0.018) | (0.012) | (0.011) | (0.008) |
| Manager | 0.271*** | 0.135*** | 0.134*** | 0.138*** |
| | (0.019) | (0.015) | (0.015) | (0.012) |
| North | -0.034*** | 0.021* | 0.124*** | 0.120*** |
| | (0.012) | (0.011) | (0.019) | (0.019) |
| Centre | -0.020* | 0.035*** | 0.089*** | 0.060*** |
| | (0.012) | (0.013) | (0.016) | (0.015) |
| Home_Time | 0.012*** | 0.005*** | 0.004*** | 0.007*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Married | 0.050*** | 0.055*** | 0.013 | 0.083*** |
| | (0.009) | (0.016) | (0.011) | (0.016) |

| | | | | |
|---|---|---|---|---|
| Italian | 0.040 | 0.152 | 0.074** | 0.090** |
| | (0.080) | (0.244) | (0.036) | (0.041) |
| Homeowner | 0.037*** | 0.027** | 0.048*** | 0.051*** |
| | (0.011) | (0.013) | (0.010) | (0.009) |
| Educ_Fath_Uni | 0.013 | -0.008 | 0.003 | 0.013 |
| | (0.014) | (0.018) | (0.018) | (0.016) |
| Educ_Moth _Uni | 0.051*** | -0.043 | 0.050** | -0.025 |
| | (0.019) | (0.026) | (0.021) | (0.018) |
| $\lambda_W^{PC}$ | 0.126*** | 0.209* | | |
| | (0.035) | (0.111) | | |
| $\lambda_R^{PC}$ | 0.179*** | -0.009 | | |
| | (0.035) | (0.039) | | |
| $\lambda_W^{NPC}$ | | | -0.041 | -0.013 |
| | | | (0.032) | (0.043) |
| $\lambda_R^{NPC}$ | | | -0.305*** | -0.169*** |
| | | | (0.034) | (0.032) |
| Constant | -0.034 | 0.676** | 0.894*** | 0.848*** |
| | (0.162) | (0.292) | (0.115) | (0.089) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes | Yes |
| Observations | 8,116 | 6,798 | 11,282 | 15,649 |
| R-squared | 0.373 | 0.335 | 0.206 | 0.262 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.23 OLS Estimates by Gender with Selection Variables – 2005

| Year | 2005 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.017*** | 0.011** | 0.013*** | 0.020*** |
| | (0.004) | (0.005) | (0.003) | (0.002) |
| Exper2 | -0.000* | -0.000* | -0.000** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |

| | | | | |
|---|---|---|---|---|
| Tenure | 0.001 | 0.003** | 0.003** | 0.001* |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Educ | 0.271*** | 0.142*** | 0.103*** | 0.095*** |
| | (0.036) | (0.034) | (0.036) | (0.022) |
| Max_D_Mark | 0.014 | 0.041 | 0.093** | -0.043 |
| | (0.025) | (0.035) | (0.040) | (0.042) |
| Contract_Type | -0.045 | 0.115** | 0.032 | 0.040** |
| | (0.032) | (0.047) | (0.021) | (0.019) |
| Work_Climate | 0.001 | 0.013 | 0.006 | 0.014 |
| | (0.011) | (0.013) | (0.011) | (0.009) |
| Work_Stab | 0.030*** | -0.021 | -0.011 | -0.002 |
| | (0.010) | (0.016) | (0.009) | (0.008) |
| Work_Time | -0.011 | 0.009 | 0.020* | 0.028*** |
| | (0.010) | (0.012) | (0.010) | (0.008) |
| Work_Task | 0.007 | 0.020* | 0.010 | 0.022*** |
| | (0.009) | (0.010) | (0.009) | (0.007) |
| Intermed_Prof | 0.081** | -0.035 | 0.076*** | 0.036** |
| | (0.037) | (0.026) | (0.020) | (0.015) |
| Manager | 0.338*** | 0.259*** | 0.143*** | 0.238*** |
| | (0.039) | (0.036) | (0.032) | (0.025) |
| North | -0.033 | 0.017 | 0.134*** | 0.060* |
| | (0.021) | (0.022) | (0.026) | (0.034) |
| Centre | -0.016 | 0.075*** | 0.115*** | 0.039 |
| | (0.021) | (0.025) | (0.025) | (0.025) |
| Home_Time | 0.010*** | 0.001 | 0.004 | 0.006*** |
| | (0.003) | (0.003) | (0.003) | (0.002) |
| Married | 0.032* | 0.003 | 0.022 | 0.043* |
| | (0.019) | (0.029) | (0.021) | (0.026) |
| Italian | 0.003 | 0.981*** | 0.026 | 0.008 |
| | (0.195) | (0.356) | (0.064) | (0.068) |
| Homeowner | 0.011 | 0.034 | 0.004 | 0.008 |
| | (0.021) | (0.029) | (0.022) | (0.024) |
| Educ_Fath_Uni | 0.023 | -0.047 | 0.021 | 0.042 |
| | (0.028) | (0.037) | (0.039) | (0.034) |
| Educ_Moth_Uni | -0.049 | 0.004 | 0.089* | -0.027 |
| | (0.040) | (0.055) | (0.051) | (0.041) |
| $\lambda_W^{PC}$ | 0.019 | 0.375* | | |
| | (0.045) | (0.210) | | |
| $\lambda_R^{PC}$ | 0.118** | -0.085 | | |
| | (0.053) | (0.070) | | |
| $\lambda_W^{NPC}$ | | | 0.009 | -0.050 |
| | | | (0.037) | (0.072) |

| | | | | |
|---|---|---|---|---|
| $\lambda_R^{NPC}$ | | | -0.115** | -0.097* |
| | | | (0.058) | (0.055) |
| Constant | 0.379 | 0.317 | 0.866*** | 1.009*** |
| | (0.316) | (0.455) | (0.182) | (0.145) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 1,987 | 1,692 | 2,526 | 3,315 |
| R-squared | 0.406 | 0.384 | 0.266 | 0.338 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.24 OLS Estimates by Gender with Selection Variables – 2006

| Year | 2006 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.023*** | 0.016*** | 0.013*** | 0.016*** |
| | (0.005) | (0.005) | (0.003) | (0.003) |
| Exper2 | -0.000** | -0.000* | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.000 | 0.003** | 0.004*** | 0.004*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Educ | 0.289*** | 0.215*** | -0.022 | 0.074*** |
| | (0.057) | (0.037) | (0.037) | (0.023) |
| Max_D_Mark | -0.025 | -0.021 | 0.129*** | 0.057 |
| | (0.027) | (0.033) | (0.040) | (0.045) |
| Contract_Type | -0.006 | 0.035 | 0.037* | 0.067*** |
| | (0.038) | (0.051) | (0.020) | (0.020) |
| Work_Climate | -0.012 | -0.005 | 0.004 | 0.009 |
| | (0.011) | (0.013) | (0.011) | (0.010) |
| Work_Stab | 0.029** | 0.039* | -0.016 | -0.016 |
| | (0.015) | (0.021) | (0.010) | (0.010) |
| Work_Time | 0.025** | 0.006 | -0.002 | 0.022** |
| | (0.012) | (0.012) | (0.010) | (0.009) |
| Work_Task | 0.012 | 0.017 | 0.023** | 0.024** |

| | | | | |
|---|---|---|---|---|
| | (0.012) | (0.012) | (0.012) | (0.010) |
| Intermed_Prof | 0.127*** | -0.044* | 0.087*** | 0.062*** |
| | (0.039) | (0.024) | (0.020) | (0.015) |
| Manager | 0.264*** | 0.192*** | 0.150*** | 0.174*** |
| | (0.043) | (0.034) | (0.035) | (0.025) |
| North | -0.046* | -0.022 | 0.029 | 0.078** |
| | (0.025) | (0.021) | (0.032) | (0.032) |
| Centre | -0.077*** | 0.017 | 0.016 | 0.035 |
| | (0.025) | (0.024) | (0.029) | (0.029) |
| Home_Time | 0.009** | 0.010*** | -0.001 | 0.006*** |
| | (0.004) | (0.003) | (0.002) | (0.002) |
| Married | 0.047** | 0.049 | 0.033 | 0.049 |
| | (0.020) | (0.031) | (0.024) | (0.031) |
| Italian | -0.252 | | 0.098 | 0.085 |
| | (0.233) | | (0.067) | (0.076) |
| Homeowner | 0.029 | 0.041 | 0.012 | 0.017 |
| | (0.027) | (0.025) | (0.024) | (0.024) |
| Educ_Fath_Uni | 0.030 | 0.021 | 0.002 | 0.050 |
| | (0.028) | (0.035) | (0.035) | (0.032) |
| Educ_Moth_Uni | 0.048 | -0.045 | 0.025 | -0.005 |
| | (0.038) | (0.054) | (0.045) | (0.039) |
| $\lambda_W^{PC}$ | 0.047 | 0.469** | | |
| | (0.080) | (0.219) | | |
| $\lambda_R^{PC}$ | 0.178** | 0.088 | | |
| | (0.088) | (0.077) | | |
| $\lambda_W^{NPC}$ | | | -0.109** | 0.006 |
| | | | (0.053) | (0.081) |
| $\lambda_R^{NPC}$ | | | -0.220*** | -0.123** |
| | | | (0.059) | (0.061) |
| Constant | 0.406 | 0.599* | 1.440*** | 0.976*** |
| | (0.484) | (0.326) | (0.203) | (0.164) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 1,715 | 1,767 | 2,368 | 3,455 |
| R-squared | 0.376 | 0.359 | 0.234 | 0.317 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.25 OLS Estimates by Gender with Selection Variables – 2008

| Year | 2008 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.017*** | 0.012** | 0.009*** | 0.014*** |
| | (0.005) | (0.006) | (0.003) | (0.002) |
| Exper2 | -0.000 | -0.000* | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.001 | 0.001 | 0.005*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Educ | 0.309*** | 0.178*** | -0.016 | 0.054** |
| | (0.053) | (0.046) | (0.036) | (0.023) |
| Max_D_Mark | 0.021 | -0.015 | 0.069 | 0.093* |
| | (0.030) | (0.040) | (0.043) | (0.054) |
| Contract_Type | 0.018 | 0.106* | 0.001 | 0.051*** |
| | (0.037) | (0.057) | (0.020) | (0.019) |
| Work_Climate | -0.019* | 0.003 | -0.001 | -0.006 |
| | (0.011) | (0.014) | (0.009) | (0.009) |
| Work_Stab | 0.020* | -0.006 | -0.008 | 0.012 |
| | (0.012) | (0.021) | (0.009) | (0.010) |
| Work_Time | 0.049*** | -0.011 | 0.009 | 0.009 |
| | (0.013) | (0.015) | (0.010) | (0.009) |
| Work_Task | -0.001 | 0.007 | 0.015 | 0.018** |
| | (0.011) | (0.012) | (0.010) | (0.009) |
| Intermed_Prof | 0.208*** | -0.035 | 0.087*** | 0.058*** |
| | (0.045) | (0.026) | (0.020) | (0.016) |
| Manager | 0.271*** | 0.112*** | 0.123*** | 0.127*** |
| | (0.049) | (0.033) | (0.032) | (0.026) |
| North | -0.085*** | 0.049** | 0.102*** | 0.059* |
| | (0.028) | (0.024) | (0.035) | (0.035) |
| Centre | -0.058** | 0.026 | 0.090*** | 0.036 |
| | (0.027) | (0.027) | (0.029) | (0.030) |
| Home_Time | 0.008* | 0.002 | -0.002 | 0.002 |
| | (0.004) | (0.004) | (0.003) | (0.002) |
| Married | 0.038* | 0.011 | 0.039** | 0.060* |
| | (0.021) | (0.037) | (0.019) | (0.031) |

| | | | | |
|---|---|---|---|---|
| Italian | 0.065 | -0.160 | 0.122* | 0.031 |
| | (0.140) | (0.331) | (0.069) | (0.072) |
| Homeowner | 0.027 | 0.008 | 0.019 | 0.039** |
| | (0.031) | (0.030) | (0.023) | (0.019) |
| Educ_Fath_Uni | 0.033 | 0.016 | 0.012 | 0.059* |
| | (0.031) | (0.036) | (0.033) | (0.031) |
| Educ_Moth_Uni | 0.117*** | -0.082 | 0.044 | -0.097*** |
| | (0.041) | (0.054) | (0.038) | (0.037) |
| $\lambda_W^{PC}$ | 0.033 | 0.272 | | |
| | (0.084) | (0.252) | | |
| $\lambda_R^{PC}$ | 0.206** | -0.099 | | |
| | (0.082) | (0.098) | | |
| $\lambda_W^{NPC}$ | | | -0.065 | -0.006 |
| | | | (0.054) | (0.074) |
| $\lambda_R^{NPC}$ | | | -0.338*** | -0.168*** |
| | | | (0.060) | (0.061) |
| Constant | 0.167 | 1.609*** | 1.376*** | 1.263*** |
| | (0.458) | (0.557) | (0.214) | (0.172) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 1,586 | 1,392 | 2,370 | 3,253 |
| R-squared | 0.373 | 0.351 | 0.309 | 0.300 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.26 OLS Estimates by Gender with Selection Variables – 2010

| Year | 2010 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.017*** | 0.029*** | 0.009* | 0.017*** |
| | (0.005) | (0.006) | (0.004) | (0.003) |
| Exper2 | -0.000 | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003** | 0.003* | 0.003** | 0.003*** |
| | (0.001) | (0.002) | (0.002) | (0.001) |
| Educ | 0.230*** | 0.228*** | 0.003 | 0.029 |
| | (0.059) | (0.050) | (0.055) | (0.031) |
| Max_D_Mark | 0.050* | 0.080* | 0.030 | 0.099** |
| | (0.027) | (0.041) | (0.043) | (0.046) |
| Contract_Type | 0.015 | 0.029 | -0.001 | 0.077*** |
| | (0.038) | (0.057) | (0.026) | (0.021) |
| Work_Climate | -0.008 | 0.012 | -0.003 | -0.002 |
| | (0.012) | (0.015) | (0.014) | (0.010) |
| Work_Stab | 0.001 | 0.000 | -0.025* | -0.015 |
| | (0.014) | (0.021) | (0.013) | (0.010) |
| Work_Time | 0.016 | 0.008 | -0.000 | 0.024** |
| | (0.014) | (0.017) | (0.014) | (0.010) |
| Work_Task | 0.016 | -0.004 | 0.024* | 0.011 |
| | (0.012) | (0.016) | (0.015) | (0.011) |
| Intermed_Prof | 0.115*** | -0.042 | 0.062** | 0.041** |
| | (0.040) | (0.030) | (0.026) | (0.018) |
| Manager | 0.216*** | 0.090** | 0.136*** | 0.091*** |
| | (0.044) | (0.035) | (0.034) | (0.025) |
| North | -0.027 | 0.039* | 0.118** | 0.083** |
| | (0.034) | (0.024) | (0.055) | (0.036) |
| Centre | 0.003 | -0.005 | 0.064 | 0.028 |
| | (0.030) | (0.029) | (0.042) | (0.030) |
| Home_Time | 0.008** | 0.005 | -0.002 | 0.002 |
| | (0.004) | (0.004) | (0.004) | (0.002) |
| Married | 0.071*** | 0.029 | 0.039 | 0.040 |
| | (0.021) | (0.031) | (0.024) | (0.034) |

| | | | | |
|---|---|---|---|---|
| Italian | -0.014 | 0.459 | 0.057 | -0.044 |
| | (0.245) | (0.392) | (0.083) | (0.081) |
| Homeowner | 0.057* | 0.101** | 0.008 | 0.045* |
| | (0.030) | (0.039) | (0.029) | (0.025) |
| Educ_Fath_Uni | 0.032 | -0.060 | -0.043 | -0.036 |
| | (0.033) | (0.041) | (0.045) | (0.037) |
| Educ_Moth_Uni | 0.022 | 0.035 | -0.041 | 0.068* |
| | (0.044) | (0.064) | (0.051) | (0.041) |
| $\lambda_W^{PC}$ | 0.077 | -0.168 | | |
| | (0.104) | (0.422) | | |
| $\lambda_R^{PC}$ | 0.116 | 0.041 | | |
| | (0.089) | (0.096) | | |
| $\lambda_W^{NPC}$ | | | 0.023 | -0.041 |
| | | | (0.087) | (0.089) |
| $\lambda_R^{NPC}$ | | | -0.313*** | -0.251*** |
| | | | (0.089) | (0.077) |
| Constant | 0.519 | 0.353 | 1.428*** | 1.391*** |
| | (0.524) | (0.606) | (0.323) | (0.234) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 1,621 | 1,416 | 2,588 | 3,795 |
| R-squared | 0.357 | 0.303 | 0.148 | 0.221 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.27 OLS Estimates by Gender with Selection Variables – 2011

| Year | 2011 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.012** | 0.014** | -0.007 | 0.014*** |
| | (0.006) | (0.006) | (0.004) | (0.003) |
| Exper2 | -0.000 | -0.000 | 0.000 | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.002 | 0.004*** | 0.006*** | 0.002* |

| | | | | |
|---|---|---|---|---|
| | (0.001) | (0.001) | (0.002) | (0.001) |
| Educ | 0.285*** | 0.268*** | -0.057 | 0.012 |
| | (0.069) | (0.051) | (0.061) | (0.029) |
| Max_D_Mark | 0.044 | 0.042 | -0.041 | 0.011 |
| | (0.032) | (0.046) | (0.046) | (0.049) |
| Contract_Type | -0.001 | 0.082 | 0.016 | 0.074*** |
| | (0.039) | (0.056) | (0.027) | (0.020) |
| Work_Climate | -0.002 | 0.019 | 0.022 | 0.006 |
| | (0.013) | (0.015) | (0.015) | (0.011) |
| Work_Stab | 0.027* | 0.018 | -0.057*** | -0.022** |
| | (0.017) | (0.021) | (0.013) | (0.010) |
| Work_Time | 0.007 | 0.016 | -0.027* | 0.006 |
| | (0.016) | (0.016) | (0.015) | (0.011) |
| Work_Task | 0.010 | 0.007 | -0.004 | 0.005 |
| | (0.014) | (0.015) | (0.015) | (0.011) |
| Intermed_Prof | 0.133*** | -0.038 | 0.086*** | 0.018 |
| | (0.042) | (0.029) | (0.027) | (0.018) |
| Manager | 0.247*** | 0.068** | 0.134*** | 0.053** |
| | (0.044) | (0.033) | (0.033) | (0.024) |
| North | -0.031 | 0.028 | 0.149*** | -0.005 |
| | (0.036) | (0.028) | (0.053) | (0.035) |
| Centre | 0.024 | 0.032 | 0.100** | -0.029 |
| | (0.031) | (0.030) | (0.039) | (0.030) |
| Home_Time | 0.011** | 0.010*** | -0.003 | 0.002 |
| | (0.005) | (0.004) | (0.004) | (0.002) |
| Married | 0.056** | 0.138*** | -0.018 | -0.065* |
| | (0.025) | (0.041) | (0.026) | (0.036) |
| Italian | -0.013 | | 0.198*** | 0.082 |
| | (0.188) | | (0.073) | (0.083) |
| Homeowner | 0.013 | 0.044 | 0.052* | -0.028 |
| | (0.034) | (0.037) | (0.031) | (0.025) |
| Educ_Fath_Uni | 0.037 | 0.011 | -0.014 | -0.013 |
| | (0.036) | (0.043) | (0.046) | (0.034) |
| Educ_Moth_Uni | -0.035 | 0.000 | 0.129** | 0.001 |
| | (0.046) | (0.061) | (0.056) | (0.039) |
| $\lambda_W^{PC}$ | -0.018 | 0.451* | | |
| | (0.110) | (0.261) | | |
| $\lambda_R^{PC}$ | 0.165 | 0.094 | | |
| | (0.102) | (0.097) | | |
| $\lambda_W^{NPC}$ | | | 0.162 | -0.297*** |
| | | | (0.101) | (0.080) |
| $\lambda_R^{NPC}$ | | | -0.506*** | -0.363*** |

| | | | | |
|---|---|---|---|---|
| | | | (0.090) | (0.070) |
| Constant | 0.515 | 0.368 | 1.535*** | 1.764*** |
| | (0.569) | (0.463) | (0.366) | (0.211) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 1,656 | 1,249 | 2,512 | 3,560 |
| R-squared | 0.314 | 0.359 | 0.141 | 0.208 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 2.B.28 OLS Estimates by Gender with Selection Variables – 2014

| Year | 2014 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Individuals Hired by by Public Contest and Aged 18-64 | | Individuals Not Hired by by Public Contest and Aged 18-64 | |
| Variables | Women | Men | Women | Men |
| Exper | 0.008* | 0.010** | 0.005 | 0.018*** |
| | (0.004) | (0.004) | (0.003) | (0.003) |
| Exper2 | 0.000 | -0.000* | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003** | 0.003* | 0.004*** | 0.004*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Educ | 0.230*** | 0.121*** | -0.042 | 0.103** |
| | (0.056) | (0.043) | (0.056) | (0.040) |
| Max_D_Mark | 0.017 | 0.064 | 0.034 | -0.123** |
| | (0.032) | (0.046) | (0.042) | (0.049) |
| Contract_Type | 0.115*** | 0.202*** | -0.004 | 0.050** |
| | (0.037) | (0.051) | (0.022) | (0.021) |
| Work_Climate | -0.013 | -0.005 | 0.005 | -0.007 |
| | (0.011) | (0.012) | (0.011) | (0.009) |
| Work_Stab | 0.014 | -0.003 | -0.038*** | -0.004 |
| | (0.017) | (0.017) | (0.013) | (0.010) |
| Work_Time | 0.039*** | -0.009 | 0.017* | 0.003 |
| | (0.010) | (0.013) | (0.010) | (0.009) |
| Work_Task | 0.016 | 0.012 | -0.000 | 0.016* |

| | | | | |
|---|---|---|---|---|
| | (0.011) | (0.013) | (0.011) | (0.010) |
| Intermed_Prof | 0.063 | -0.054** | 0.096*** | 0.048*** |
| | (0.043) | (0.027) | (0.021) | (0.016) |
| Manager | 0.326*** | 0.163*** | 0.238*** | 0.152*** |
| | (0.044) | (0.034) | (0.028) | (0.023) |
| North | -0.062 | 0.022 | 0.051 | 0.138*** |
| | (0.038) | (0.026) | (0.048) | (0.052) |
| Centre | -0.076** | 0.060** | 0.009 | 0.057 |
| | (0.030) | (0.027) | (0.036) | (0.039) |
| Home_Time | 0.006* | 0.003 | -0.005 | 0.004* |
| | (0.003) | (0.003) | (0.003) | (0.002) |
| Married | 0.020 | 0.032 | 0.002 | 0.074 |
| | (0.026) | (0.039) | (0.023) | (0.049) |
| Italian | -0.059 | -0.139 | 0.046 | 0.121* |
| | (0.150) | (0.140) | (0.053) | (0.069) |
| Homeowner | 0.089*** | -0.007 | -0.035 | 0.005 |
| | (0.031) | (0.032) | (0.025) | (0.023) |
| Educ_Fath_Uni | -0.001 | 0.023 | -0.030 | 0.081*** |
| | (0.031) | (0.037) | (0.034) | (0.031) |
| Educ_Moth_Uni | -0.031 | -0.008 | 0.042 | -0.035 |
| | (0.039) | (0.049) | (0.038) | (0.037) |
| $\lambda_W^{PC}$ | -0.055 | 0.233 | | |
| | (0.104) | (0.291) | | |
| $\lambda_R^{PC}$ | 0.181** | -0.027 | | |
| | (0.088) | (0.084) | | |
| $\lambda_W^{NPC}$ | | | 0.010 | 0.074 |
| | | | (0.092) | (0.115) |
| $\lambda_R^{NPC}$ | | | -0.346*** | -0.116 |
| | | | (0.087) | (0.086) |
| Constant | 0.631 | 1.593*** | 1.804*** | 0.999*** |
| | (0.473) | (0.415) | (0.332) | (0.312) |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 2,665 | 1,739 | 3,428 | 4,496 |
| R-squared | 0.273 | 0.286 | 0.146 | 0.174 |

Robust standard errors in parentheses

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

# Appendix 2.C Methodological Issues

The probabilities of observing a positive labor income given recruitment through public contests or recruitment through other channels are given below:

$$Pr(Y_W^* > 0, Y_R^* > 0) = Pr(u_W > -Z^{'}\gamma, u_R > -Q^{'}\alpha) = G(Z^{'}\gamma, Q^{'}\alpha, \rho) \tag{2.7}$$

$$Pr(Y_W^* > 0, Y_R^* \leq 0) = Pr(u_W > -Z^{'}\gamma, u_R \leq -Q^{'}\alpha) = G(Z^{'}\gamma, -Q^{'}\alpha, -\rho) \tag{2.8}$$

where $G(\cdot)$ is the standard bivariate normal distribution and $\rho$ is the correlation coefficient between the two selection rules. The subscript $W$ identifies the work decision while $R$ identifies the recruitment decision. Under the assumption that the two selection rules are not independent, that is $\rho \neq 0$, maximum likelihood of the bivariate probit leads to the following selection terms for public-contest selected employees, i.e. with $m = PC$:

$$\lambda_W^{PC} = \frac{f(Z^{'}\gamma)F[\frac{Q^{'}\alpha-\rho Z^{'}\gamma}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, Q^{'}\alpha, \rho)} \tag{2.9}$$

$$\lambda_R^{PC} = \frac{f(Q^{'}\alpha)F[\frac{Z^{'}\gamma-\rho Q^{'}\alpha}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, Q^{'}\alpha, \rho)} \tag{2.10}$$

Similarly, for the subsample of not public-contest selected individuals, i.e. with $m = NPC$, the corresponding selection terms are given by:

$$\lambda_W^{NPC} = \frac{f(Z^{'}\gamma)F[-\frac{Q^{'}\alpha-\rho Z^{'}\gamma}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, -Q^{'}\alpha, -\rho)} \tag{2.11}$$

$$\lambda_R^{NPC} = \frac{-f(Q^{'}\alpha)F[\frac{Z^{'}\gamma-\rho Q^{'}\alpha}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, -Q^{'}\alpha, -\rho)} \tag{2.12}$$

$f(\cdot)$ is the standard normal density function, while $F(\cdot)$ is the standard normal distribution function and $\rho$ is the correlation coefficient between the two selection rules.

# References

Baffoe-Bonnie, J. (2009). Black–White Wage Differentials in a Multiple Sample Selection Bias Model. *Atlantic Economic Journal*, 37(1), 1-16.

Bar, M., Kim, S., and Leukhina, O. (2015). Gender Wage Gap Accounting: The Role of Selection Bias. *Demography*, 52(5), 1729-1750.

Bertrand, M., 2011, New Perspectives on Gender. *Handbook of Labor Economics*, 4, 1543-1590.

Blackaby, D., Booth, A.L. and Frank, J., 2005, Outside Offers And The Gender Pay Gap: Empirical Evidence From the UK Academic Labour Market. *Economic Journal*, 115(501), 81-107.

Blau, F. D., and Kahn, L. M. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*, 15(4), 75-99.

Blau, F. D., and Kahn, L. M. (2003). Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1), 106-144.

Blau, F. D., and Kahn, L. M. (2006). The US Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blau, F. D., and Kahn, L. M. (2007). Changes in the Labor Supply Behavior of Married Women: 1980–2000. *Journal of Labor Economics*, 25(3), 393-438.

Blau, F. D., and Kahn, L. M. (2016). *The Gender Wage Gap: Extent, Trends, and Explanations* (No. W21913). National Bureau of Economic Research.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4), 436-455.

Cappellari, L. (2002). Earnings Dynamics and Uncertainty in Italy: How do They Differ between the Private and Public Sectors?. *Labour Economics*, 9(4), 477-496.

Cardona, M. F. (2006). *Recruitment in Civil Service Systems of EU Members and in some Candidates States*. In Seminar on "Civil Service Recruitment Procedures" Vilnius (pp. 21-22).

Carinci, F., R., De Luca Tomajo, R., Tosi, P., and Treu, T. (2002). *Diritto del Lavoro*. UTET.

Castagnetti, C., and Rosti, L. (2013). Unfair Tournaments Gender Stereotyping and Wage Discrimination among Italian Graduates. *Gender and Society*, 27(5), 630-658.

Centra, M., and Cutillo, A. (2009). *Differenziale Salariale di Genere e Lavori Tipicamente Femminili*. Studi Isfol, 2.

Chang, S. K. (2011). Simulation Estimation of Two-Tiered Dynamic Panel Tobit Models with an Application to the Labor Supply of Married Women. *Journal of Applied Econometrics*, 26(5), 854-871.

Chzhen, Y., and Mumford, K. (2011). Gender Gaps Across the Earnings Distribution for Full-Time Employees in Britain: Allowing for Sample Selection. *Labour Economics*, 18(6), 837-844.

Craveri, P. (2016). *L'Arte del Non Governo: L'Inarrestabile Declino della Repubblica Italiana*. Marsilio Editori.

Devereux, P. J. (2004). Changes in Relative Wages and Family Labor Supply. *Journal of Human Resources*, 39(3), 698-722.

Dickson, M., Postel-Vinay, F., and Turon, H. (2014). The Lifetime Earnings Premium in the Public Sector: The View from Europe. *Labour Economics*, 31, 141-161.

Dobbs, M., and Crano, W. D. (2001). Outgroup Accountability in the Minimal Group Paradigm: Implications for Aversive Discrimination and Social Identity Theory. *Personality and Social Psychology Bulletin*, 27(3), 355-364.

Eurostat (2016). *Gender Statistics*. http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_statistics (accessed 27-5-2016).

Eurostat (2017). *Gender Pay Gap in Unadjusted Form*. http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (accessed 27-05-2017).

Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources*, 43(4), 884-918.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091-1119.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. In *Annals of Economic and Social Measurement*, 5(4), 475-492.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Heilman, M. E. and Okimoto, T. G. (2008). Motherhood: A Potential Source of Bias in Employment Decisions. *Journal of Applied Psychology*, 93(1), 189-198.

Heinze, A., Beninger, D., Beblo, M., and Laisney, F. (2003). *Measuring Selectivity-Corrected Gender Wage Gaps in the EU* (No. 03-74). ZEW Discussion Papers.

Kahn, L. M. (2015). Wage Compression and the Gender Pay Gap. *IZA World of Labor*, 150.

Lazear, E. P., and Rosen, S. (1990). Male-Female Wage Differentials in Job Ladders. *Journal of Labor Economics*, 8(1, Part 2), 106-123.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, 76(3), 1071-1102.

Lee, L. (1979). Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables. *Econometrica*, 47(4), 977-996.

Lee, L. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, 51(2), 507-512.

Lindzey, G., Gilbert, D. T., and Fiske, S. T. (1998). *The Handbook of Social Psychology*, Vol. 2. McGraw-Hill.

Martins, M. F. O. (2001). Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics*, 16(1), 23-39.

Meng, C., Schmidt, P. (1985). On the Cost of Partial Observability in the Bivariate Probit Model. *International Economic Review*, 26(1), 71-85.

Minguez, A. M. (2004). The Persistence of Male Breadwinner Model in Southern European Countries in a Compared Perspective: Familism, Employment and Family Policies. *Marie Curie Annals*, 4.

Mulligan, C. B., and Rubinstein, Y. (2008). Selection, Investment, and Women's Relative Wages Over Time. *The Quarterly Journal of Economics*, 123(3), 1061-1110.

Neuman, S., and Oaxaca, R. L. (2003). Gender versus Ethnic Wage Differentials Among Professionals: Evidence from Israel. *Annales d'Économie et de Statistique*, (71/72), 267-292.

Nicaise, I. (2001). Human Capital, Reservation Wages and Job Competition: Heckman's Lambda Re-Interpreted. *Applied Economics*, 33(3), 309-315.

Niederle, M. and Vesterlund, L. (2007), Do Women Shy Away From Competition? Do Men Compete Too Much?. *The Quarterly Journal of Economics*, 122(3), 1067-1101.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3). 693-709.

Olivetti, C. and Petrongolo, B. (2008). Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26(4), 621-654.

Piazzalunga, D. (2015). Is There a Double-Negative Effect? Gender and Ethnic Wage Differentials in Italy. *Labour*, 29(3), 243-269.

Postel-Vinay, F. (2015). *Does it Pay to Be a Public-Sector Employee?*. IZA World of Labor.

Schein, V. E. (2007). Women in Management: Reflections and Projections. *Women in Management Review*, 22(1), 6-18.

Sorensen, E. (1989). Measuring the Pay Disparity between typically Female Occupations and other Jobs: A Bivariate Selectivity Approach. *Industrial and Labor Relations Review*, 42(4), 624-639.

Tunali, I. (1986). A General Structure for Models of Double-Selection and an Application to a joint Migration/Earnings Process with Remigration. *Research in Labor Economics*, 8 (Part B), 235-282.

Valian, V. (1998). Sex, Schemas, and Success: What's keeping Women back?. *Academe*, 84(5), 50-55.

Worldwide Tax Summaries (2017). *Italy: Individual – Other Tax Credits and Incentives*. http://taxsummaries.pwc.com/ID/Italy-Individual-Other-tax-credits-and-incentives (accessed 01-06-2017).

# Chapter 3

# Overeducation and the Gender Pay Gap – A Double Selectivity Approach

## 3.1 Introduction

This paper contributes to the existing literature by integrating insights from two usually separate research fields: the Gender Pay Gap (GPG) on the one side and overeducation on the other. Workers in occupations that require less schooling than they actually have are labeled 'overeducated' (Sicherman, 1991; Hartog, 2000; Sloane, 2014). According to the literature, overeducation is a pervasive feature of modern labor markets (Groot and Maassen van den Brink, 2000) and its effects represent a serious concern in Italy, where the predicted probability of being overeducated is very high, independently from the educational level or method of assessment (European Commission, 2012). Yet, in Italy, the share of individuals with tertiary education is among the lowest of all European Union (EU) member states.[1] The case of Italy is thus particularly interesting for the study of overeducation, given that, on the one hand, a large share of individuals is overeducated, while on the other hand, the amount of individuals with higher education is very low. This problem is even more important for women, as their share among graduates is high and growing, and as we show that the wage penalty for overeducation is higher for women than for men.[2]

There is a very broad literature on the relationship between overeducation and earnings (Dolton and Vignoles, 2000; Hartog, 2000; McGuinness, 2006; Rubb, 2003; Pecoraro,

[1] In Italy about 20.0% of the population holds a university degree, while 40.0% of individuals with a tertiary degree are overeducated (Meliciani and Radicchia, 2016).

[2] As in Cutillo and Di Pietro (2006).

2016) and an even wider literature on the GPG remaining resilient despite more than thirty years of equal-pay legislation (see e.g. Blau and Kahn, 2000; Goldin, 2014). There is also a considerable amount of empirical work focusing on differences in overeducation risk by gender (Groot, 1996; Boll et al., 2016) and individual features associated to higher overeducation probability (European Commission, 2012). However, very few studies focus on the relationship between overeducation and the GPG (Li and Miller, 2012; Boll and Leppin, 2013). Moreover, these few studies find that overeducation does not matter for explaining the wage gap between men and women. In particular, they do not explicitly control for unobserved heterogeneity of the overeducation decision. Unobserved heterogeneity consists of differences in individual productivity such as innate ability, school quality and on-the-job training as well as motivation and commitment to paid work. Previous studies on the effects of overeducation on wages, however, show the need to control either for sample selection bias (Battu et al., 2000) or for endogeneity of the overducation choice (Dolton and Silles, 2008). We apply a double selectivity model to simultaneously account for both sample selection bias and endogeneity bias (Tunali, 1986; Sorensen, 1989; Cutillo and Di Pietro, 2006), and find that overeducation is an important driver of the GPG in Italy.

In our data, the GPG is significantly higher among overeducated workers than among properly educated workers. Women possess better observed characteristics than men and differences in the wage structure are important for both mis- and properly matched workers in the base model without selection correction. In the adjusted model, all the unexplained component of the GPG among overeducated workers vanishes, and the difference is explained by endowments (a small part) and the selection into overeducation (the big part). In the properly educated sample, a small fraction of the GPG corrected for double selectivity is again due to differences in endowments, while both differences in coefficients and the overeducation choice are main contributors to the differential.

As the discriminatory component of the GPG disappears among overeducated workers but remains significant among properly educated ones, we further examine the question why overeducation can fight gender discrimination in pay whereas a proper match fails to do so. A possible explanation is that by compensating with higher educational attainment lower levels of unobservable personal characteristics, overeducated women signal their true productivity to employers and overcome statistical discrimination. If the educational level is an effective signal of workers' commitment, then discrimination against overeducated females should decrease. Also Boll and Leppin (2013), notwithstanding their finding that overeducation does not matter to explain the GPG among graduates in Germany, report that a noticeable part of the wage gap may be attributed to gender stereotypes assigned by employers and presumably incorporated in women's labor market decisions. We find that overeducated men and women possess worse unobservable characteristics than individuals in the properly educated sample. Moreover, overeducated working women have worse unobservable characteristics than overeducated men. However, overeducated women (not men) are better than out of employment individuals. Hence, this is the signal sent by overeducated women: they possess (for the labor market) valuable, though, unobservable characteristics, and are available to work. Among the properly educated workers, the signaling effect is less clear, as education also features human capital skills required for the job. We draw the conclusion that overeducation is the first-best matching for individuals (both men and women) with lower levels of productive characteristics. In fact, men and women compensate with more education for these lower levels. Moreover, it is a signaling device for women spending their useless-for-the-job diploma to inform employers on their valuable, though, unobservable productive characteristics and fights gender wage discrimination.

The results are important for policy measures. If overeducation signals the incapacity of the labor market to absorb higher levels of education, a higher investment in schooling is a waste of resources by individuals having near the same unobserved characteristics than properly educated ones. Conversely, if overeducation is merely a choice of individuals compensating by more human capital investment their lower unobservable differences in productivity, there is no waste of resources. As our results suggest the latter, the need for more investment in higher education is not more limited in Italy.

The paper is organized as follows. In the next Section, we provide an overview of the literature on the GPG as well as on overeducation. In Section 3.3, we discuss the standard approach to decompose mean wage differences between groups. Section 3.4 describes the problem of double selectivity. In Section 3.5, we outline the data set used. Section 3.6, presents the estimation results. Finally, we conclude.

## 3.2 Background Literature

The aim of this paper is to study the relationship between overeducation and the GPG. The GPG is the difference between the average hourly earnings of men and women expressed as a percentage of average hourly earnings of men. It is usually called raw or 'unadjusted' as it does not take into account factors that influence the GPG, such as differences in education, labor market experience or type of job (Eurostat, 2016). In 2014, the unadjusted GPG was 16.1% at EU-28 level and 6.5% in Italy (Eurostat, 2017b). Besides the GPG and the gender gap in paid hours, it is important to consider gender gaps in employment, as also differences in labor market participation and employment contribute substantially to the difference in average earnings between men and women (Eurostat, 2016). Gender employment gaps are particularly pronounced in Italy. In 2014, 19.4% less women than men were engaged in the Italian labor market (Eurostat, 2017a). For comparison, the difference amounted to 11.5% at EU-28 level in the same year (Eurostat, 2017a).

The risk of overeducation, too, may differ for men and women, either because of gender discrimination or because of gender-specific differences in personal and job characteristics. A meta-analysis of 25 studies on overeducation conducted by Groot and Maassen van den Brink (2000) concludes that the incidence of overeducation varies from 10.0% to 42.0%.[3] On average, 26.0% of all workers in the United States (US) and 22.0% in European countries are overeducated. In our data (ISFOL PLUS 2005-2014), the proportion of individuals working in jobs that require less schooling than they actually have is 33.0% (34.8% in the male sample; 31.5% in the female sample).

As our main statement in this paper is that overeducation is an important driver of the GPG, we summarize, first, briefly the literature about the sources of the GPG, and, second, we review the main theories explaining the overeducation phenomenon. It is worth noting that the main sources of the GPG and the overeducation choice proposed in the literature are near the same. We also discuss implications of different methods of assessment for the phenomenon of overeducation.

### 3.2.1 Theories on the GPG

The literature on the sources of the GPG emphasizes two broad sets of explanations. Explanations focusing on the supply-side of the labor market, and explanations focusing on the demand-side of the labor market. These two sets of explanations are not mutually exclusive, they both play a role in explaining the GPG. However, traditionally, the first set of explanations focuses on the choices made by women, while the second focuses on job-related constraints faced by women. Supply-side explanations mainly refer to work-life preferences and cultural beliefs, the sexual division of labor in the household, and the human capital theory. Demand-side explanations mainly refer to compensating differentials, statistical discrimination and other allocative gender-biased decisions.

We consider the supply-side explanations first. A preference-based explanation posits that gender differences in the career path and earnings derive largely from genuine sex role preferences (Hakim, 2000). However, several scholars indicate that gender stereotypes (that

[3] Groot and Maassen van den Brink (2000) look at average values over a period of 20 years (from the 1970s to the 1990s).

is, non-conscious beliefs that stem from social norms and affect our expectations and our judgments of others) may shape individual's preferences making men and women choose different jobs and different career paths (Correll, 2001; 2004; Ridgeway, 2009). Economists also argue that women earn less than men because of the division of labor within the family. This results in differences between men and women in labor market characteristics as well as in human capital accumulation. Becker (1985) emphasizes the importance of household production in economic theory and highlights that much of this output is produced by women. As a consequence, it is well established in the literature that women are less likely to have successful careers than men in the labor market (Sasser, 2005), and that women with children earn less than other women (Waldfogel 1997; Budig and England, 2001). Lastly, the human capital theory explains women's lower wages with gender differences in the amount and kind of education, on-the-job training and other aspects of labor market experience that affect the individuals' productivity (Mincer and Polachek, 1974). In the past, men typically had better access to university-level institutions, while nowadays female graduates exceed the number of male graduates, and on average female students outperform male students in academic achievements in most OECD[4] countries (OECD, 2009).[5]

We consider now the demand-side explanations of the GPG. Gender inequality in wages may also be due to differences in working conditions. According to the compensating wage theory, jobs with unfavorable conditions receive pecuniary rewards compared to jobs with better working arrangements (England and Folbre, 2005). If female dominated occupations have some benefits making it easier to combine work and family life, these benefits may result in lower wages (Solberg and Laughlin, 1995). Another explanation on the demand-side of the labor market for gender income differences is discrimination against women, i.e. employers' gender-biased decisions on the allocation of individuals across and within occupations. Empirical analysis show that both the possibility of entering an occupation and access to promotion within occupations differ between men and women, all else equal (Anker, 1998). Statistical discrimination occurs, when employers make hiring and promotion decisions based on the average productivity of the individual's gender but not on its personal characteristics (Arrow, 1972; Stiglitz, 1973). For example, based on higher statistical probability for women to quit (England, 1992), employers may prefer to allocate women to positions with low turn-over costs (Bielby and Baron, 1986). Both gender stereotyping and wage discrimination have been well documented in empirical research (e.g.

[4]OECD = Organisation for Economic Co-operation and Development

[5]In fact, the catching-up of women in terms of human capital is identified as a major reason for the convergence of the GPG over time (Goldin, 2006; 2014).

Blau et al., 2013). For example Castagnetti and Rosti (2013) show that stereotyping is clearly related to gender wage discrimination.

### 3.2.2 Theories on Overeducation and Method of Assessment

There are two main (competing) approaches attempting to explain the overeducation phenomenon: the human capital model (Alba-Ramirez, 1993; Büchel and Battu, 2003) and the signaling model (Kroch and Sjoblom 1994; Dolton and Vignoles 2000).

From the human capital perspective (Mincer, 1958), overeducation is a mechanism for labor market adjustment when there is an excess supply of high-skilled workers, and it is considered a second-best employment result. When the increase in the educational level of the work force is accompanied by lower growth rates of jobs for more educated workers, the allocation of skills over jobs may be less than optimal, and some individuals accept jobs for which they are overeducated rather than remaining unemployed. Indeed, there are studies supporting that theory and finding a negative impact of work experience on overeducation risk (e.g. Alba-Ramirez, 1993; Nielsen, 2011).

In the job signaling model, education is used as a screening device to identify higher ability workers (Arrow, 1973; Spence, 1973). Firms are assumed to have imperfect information about the productivity of workers, and in response to this information asymmetry between employers and employees, individuals may use education as a signal of productivity. In this case, overeducation does not imply overqualification. Overeducation arises when there is a signaling equilibrium under which it is optimal for individuals to invest in more education than is strictly required to perform the tasks of their jobs (Spence, 1973). It is worth noting, however, that whilst overeducation can arise in a signaling equilibrium, it is a Pareto-inferior equilibrium in which overeducation persists. Education signals to employers that overeducated workers posses higher levels of individual ability, motivation, commitment and so on, than their unemployed competitors (but lower than their properly educated competitors; Kedir et al., 2012). Statistical discrimination occurs when employers use average characteristics of groups to predict individual worker productivity (Arrow, 1973). In this context, the education level can be seen as a proxy for unobserved positive individual characteristics, such as productivity. As Livanos and Nunez (2012) argue, discrimination arises from an adverse selection problem, where the hidden information is the individual's commitment to a professional career. Education may act as a signal to employers emphasizing the future commitment of workers to their careers.

Bauer (2002) and Chevalier (2003) argue that overeducation may be only apparent, as a consequence of measurement errors due to unobserved heterogeneity. Even though the returns of overeducation are lower than the returns of required schooling, lower return rates do not necessarily imply underutilization of human capital. The negative wage effects of overeducation may be due to self-selection into overqualification. Alba-Ramirez (1993) finds evidence suggesting that the overeducated may use surplus education as a substitute for other forms of human capital that they lack. In this case observed overqualification is simply a measurement error due to the presence of statistically unobserved differences in abilities or motivation, educational quality, unmeasured skills, or worker preferences (e.g. preference for family-friendly work schedules). Most of the difference in earnings between overeducated and properly matched workers identified by the previous literature are caused by a failure to control for unobserved heterogeneity (as for example in Bauer, 2002; Leuven and Oosterbeek, 2011). The career mobility theory presents overeducation as an investment to improve future employment opportunities (Sicherman and Galor, 1990; Sicherman, 1991). The overeducated worker may accept a period of overeducation in order to achieve higher earnings in the future by gathering experience at low entry levels. If the hypothesis of upward mobility holds, overeducation should not be seen as a cause for concern, but merely as a temporary phenomenon consistent with efficiency goals. Sicherman (1991) and Alba-Ramirez (1993) find empirical support for this. In contrast, Battu et al. (2000) and Büchel and Mertens (2004) as well as Baert et al. (2013) do not detect significant wage growth for formerly overeducated workers. The theory of job competition assumes that workers are primary concerned with competition for jobs, not for wages (Thurow, 1975). Among equally educated workers, the higher an individual's ability and experience, the lower the cost of the individual's training. As a consequence, smarter people will have a higher chance of securing better jobs, while less able and less experienced workers will switch to jobs for which they are overeducated. Also the assignment theory views overeducation as an inefficient outcome of a job matching procedure (Sattinger, 1993). Due to the existence of search costs (Jovanovic, 1979), highly educated workers might be satisfied with finding a position at a level below their formal qualification. This case can be extended to include all the overeducation choices arising from heterogeneous preferences among individuals, whether they are genuine or stereotyped. Individuals, both male and female, typically have preferences and dislikes for certain occupations. These preferences can induce them to choose matches in which they are paid below their productivity. A specific application of a job matching framework is the theory of differential overqualification developed by Frank (1978). As in traditional gender role models, the husband optimizes his job search first, the opportunities for a successful career of his wife could be more limited because of higher mobility costs. Empirically, Büchel and Battu (2003) find evidence supporting this theory, while McGoldrick and Robst (1996) reject the hypothesis that married women face a significantly higher probability to be overeducated.

As stated, the literature usually considers workers as overeducated when they have completed more years of education than the current job requires (e.g. Sloane, 2014). However, the literature points out that the concept of overeducation may not have a single meaning and may be open to various interpretations, making the empirical assessment difficult. The specific definition depends on how overeducation is measured in the data. As the exact wording of the question varies across studies, different indicators may classify as overeducation similar, though distinct, phenomena. In particular, it is worth distinguishing between indicators that refer to the level of education required to get the job (overqualification) on the one side, and those that refer to the educational level required to perform the job (overskilling) on the other side. Several methods of overeducation assessment can be identified in empirical studies (Hartog, 2000; McGuinness, 2006). These indicators can be classified in three groups: objective, subjective, and statistical.[6] While objective indicators are based on job analysis, that is, on occupational dictionaries that estimate the required educational level for each occupation, subjective ones are based on workers' self-assessment. Self-assessed procedures may consist of either directly or indirectly formulated questions to the interviewees of a survey. Direct questions ask for example whether the educational level attained is required to obtain (or perform) a certain job, or if the skills acquired during the educational career are actually used. Indirect inquiries ask about the most suitable educational degree (or skills) required to perform the job. In this framework, the presence of overeducation is identified by comparing the reply with the educational level of the interviewee. The statistical method classifies as overeducated those individuals that exceed the mean years of education for their job by more than one standard deviation above the mean. Each of these indicators has merits and drawbacks (Hartog, 2000). Workers' self-assessment deals with the respondent's job precisely, but it usually lacks rigorous instructions. Systematic job analysis is a very attractive source for clear definitions and detailed measurement instructions, but it may be too expensive to carry out on a large scale. Statistical indicators are based on relative terms and can be easily biased by credential inflation. Therefore, overall, the self-assessment indicator is considered the best-available measure for overeducation (Hartog, 2000).

[6] In the data we use, overeducation is based on a subjective measure (direct question).

## 3.3 Decomposing Wage Differences between Men and Women

As we are interested in studying the relationship between overeducation and the GPG, we focus on the adjusted measures of the disparity in hourly wages that persist even when employed women and men are similar with regard to personal and job characteristics. This gap is of special interest for discrimination search, as the measured wage disparity cannot be justified on grounds of productivity.

The most common applied counterfactual procedure for decomposing the GPG is the Oaxaca (1973) and Blinder (1973) decomposition (see Fortin et al., 2011, for a survey). The method divides the wage differential into a part that is 'explained' by group differences in observable labor market characteristics, such as education or work experience, and a part that cannot be accounted for by such differences in wage determinants. The latter is the so-called 'unexplained' part or adjusted GPG and often used as a measure for discrimination. Yet, it also includes effects of group differences in unobserved predictors (Blau and Kahn, 2006). In Appendix 3.A, we provide details on the econometric model applied. Before estimating and decomposing the GPG, i.e. applying the standard Oaxaca-Blinder decomposition for the distinct subsamples (over- and properly educated individuals), we estimate a Mincer-type wage equation separately for men and women in each subsample (Appendix 3.A.1). Then, we describe the decomposition method applied (Appendix 3.A.2).

## 3.4 Accounting for Sample Selection

The outcome of paid work, either for properly educated or overeducated workers, is only observed for a non-random sample. Therefore, the coefficients obtained from Ordinary Least Squares (OLS) regressions are biased. As the origin of the selection could be related to earnings, one needs to explicitly consider this process in the estimation of the wage equation. The selection into the labor market may depend on some positive factors such as individual ability, motivation or educational quality, raising both, the probability of being employed and wages. Yet, it is omitted in the earnings equation as the factors mentioned above are unobservable in the data. The sample selection bias that stems from not considering the participation decision may be particularly relevant in Italy given low female participation in the Italian labor market (Olivetti and Petrongolo, 2008; Centra and Cutillo, 2009). Despite the participation decision in general, individuals are also confronted with the decision whether to accept wage offers for jobs that do not match their educational level. Hence, sorting into the over- or properly educated sample may be a result of differences in unobservable characteristics between the individuals. Neglecting this problem may lead to the conclusion that overeducation signals the incapacity of the labor market to absorb all workers according to their educational level. This would imply that there is an overinvestment in educational attainment and a waste of resources. However, this may not be the case when overeducation mainly reflects unobserved differences in characteristics. In particular, failure to account for these selection choices would lead to inconsistent and biased estimates of both the gender-specific wage equation as well as the components of the GPG. In contrast, provided that the estimated impact of overeducation on wages is free from heterogeneity bias, mismatched workers can expect significant increases in earnings if they were assigned to jobs requiring a qualification level in accordance with their actual educational attainment. Thus, it is important to control for the endogeneity of overeducation in the estimation of the wage equation because the same unobserved characteristics influencing the overeducation choice may also affect wages.

The standard empirical framework that neglects selectivity issues generally tends to overestimate the negative wage effects of overeducation (Bauer, 2002; Chevalier, 2003).[7] In order to fully correct the wage equation, we estimate a model with a double selection process, i.e. we control for both the participation and the overeducation decision. Following the literature, we extend the Heckman (1979) two-stage selection model to include multiple decisions (Dubin and McFadden, 1984; Schmertmann, 1994; Sorensen, 1989; Tunali, 1986). Our setup refers to the case of a censored probit, i.e. partial partial observability according to the definition of Meng and Schmidt (1985). We follow the literature to identify the participation and overeducation decision. In Appendix 3.A.3–3.A.4, we outline both the estimation procedure of the model with double selection as well as the identification strategy of the selection equations. We derive the selection terms that are then included in the wage regressions (in order to obtain consistent parameter estimates) and present the decomposition expression when accounting for double selection into the sample.

[7] In contrast, Cutillo and Di Pietro (2006) find that failure to control for this correlation yields an OLS estimator of the effect of overeducation on wages that is downward biased.

## 3.5 Data and Sample Restriction

We use the complete release of the survey PLUS[8] from the Italian Institute for the Development of Vocational Training for Workers (ISFOL). So far, the following data waves with panel structure have been released: 2005, 2006, 2008, 2010, 2011 and 2014. ISFOL PLUS covers the whole population with focus on the working population. The data was collected by means of Computer Assisted Telephone Interviewing (CATI) and uses only direct answers, i.e. no proxies are used.

Even if the major part of empirical research concerning overeducation refers to graduate workers, overeducation is not a prerogative of tertiary educated individuals only (see e.g. Groot and Maassen van den Brink, 2000). ISFOL PLUS includes a direct question asking whether the individual's level of education is necessary for the working activity performed. Thus, all interviewees categorize themselves via self-assessment as over- or properly educated.[9] In particular, the question allows the categorization not only for graduates but for the whole stock of working individuals in the sample. As stated in Section 3.2.2, estimation results based on a subjective, an objective or a statistical measure of overeducation may differ (McGoldrick and Robst, 1996; Pecoraro, 2016).

We use the complete release of panel dimension to study the effect of overeducation on the GPG. There are new entrants across the releases and through attrition, we loose individuals. Thus, the sample composition changes. The analysis is based on a pooled regression model including dummies for the different releases as explanatory variables. The sample is restricted to individuals that have at least graduated from high school, i.e. enjoyed minimally 13 years of schooling. This sample restriction is justified by a relatively low risk of overeducation for individuals with less than high school diploma (Leuven and Oosterbeek, 2011). In the original sample, there are 159,615 observations of panel dimension. We also exclude students, pensioners and disabled individuals as well as unemployed individuals from the analysis. The aim of this restriction is to form a homogeneous sample of individuals (voluntarily) out of the labor force and employed individuals (Heinze et al., 2003). We drop also missing observations on other variables of interest. This leaves us with a sample size of 43,178 individual labor-market profiles, whereof 23,726 are female (54.9%) and 19,452 are male (45.1%). In the data, 6,775 men and 7,481 women are working in jobs that require less schooling than they actually have (i.e. are overeducated). Thus, more than one third of

[8]PLUS = Participation, Labor, Unemployment Survey

[9]This implies that overeducation is estimated according to a subjective criterion by the workers and it is recorded according to a dichotomous classification, i.e. a positive or negative reply to the following direct question: "Is your level of education necessary for your current job?".

the individuals in the sample is overeducated. We use the logarithm of net hourly wages as dependent variable. The variable is defined as the net monthly wage perceived divided by the number of actual working hours per month.[10]

Table 3.1 reports means and standard deviations for some of the explanatory variables used in the analysis. On average, overeducated workers are younger and (both males and females) have lower schooling, less experience and job tenure. Moreover, overeducated employees are less often married or parents than properly matched employees. The higher average age of properly educated compared to overeducated workers may drive these differences. A full list of the variables used in the analysis along with their definitions and coding is provided in Appendix 3.B, Table 3.B.1.

Table 3.1 Descriptive Statistics

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Overeducated Sample | | Properly Educated Sample | |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. |
| female | 0.525 | 0.499 | 0.562 | 0.496 |
| Age | 35.90 | 11.80 | 41.50 | 12.73 |
| Schooling | 13.66 | 1.244 | 14.26 | 1.480 |
| Exper | 14.52 | 11.62 | 18.94 | 12.55 |
| Tenure | 9.37 | 9.953 | 14.40 | 12.07 |
| Manager | 0.080 | 0.271 | 0.330 | 0.470 |
| Intermed_Prof | 0.466 | 0.499 | 0.549 | 0.498 |
| North | 0.477 | 0.499 | 0.467 | 0.499 |
| Centre | 0.216 | 0.411 | 0.192 | 0.394 |
| Italian | 0.988 | 0.108 | 0.996 | 0.0651 |
| Married | 0.445 | 0.497 | 0.576 | 0.494 |
| Kids | 0.440 | 0.496 | 0.568 | 0.495 |
| Kids_3 | 0.282 | 0.450 | 0.291 | 0.454 |
| Reloc | 0.050 | 0.218 | 0.080 | 0.271 |
| Observations | 14,256 | | 28,922 | |

[10]The survey includes monthly as well as annual gross earnings. However, monthly gross earnings are almost entirely missing (98.0% of all observations in the data are missing values). Gross annual earnings, when divided by the number of months in a calendar year (including a 13th month income) and compared with the monthly measure, differ by more than 800 Euro per month. Therefore, we prefer to use the reported monthly net income.

## 3.6 Estimation Results

In this Section, we present our estimation results. We show that the difference in the GPGs between properly and overeducated individuals is significant, and that the unadjusted as well as adjusted GPG is higher among overeducated workers. We estimate the effect of overeducation on wages and calculate the likelihood of overeducation by gender. Lastly, we discuss the results from the model with double selection.

### 3.6.1 The Effect of Overeducation on the GPG

Table 3.1 reports the log of hourly wages for overeducated and properly educated individuals by gender. The GPG in net hourly wages in the full sample amounts to 4.7%.[11] The data also show that the GPG is much higher among overeducated workers (9.6%) compared to properly educated ones (3.5%).[12]

Table 3.1 Log of Hourly Wages in Euro and Raw GPG

| | (1) | (2) | (3) |
|---|---|---|---|
| | Full Sample | Overeducated Sample | Properly Educated Sample |
| $\overline{ln(W_{M+F})}$ | 2.109 | 1.938 | 2.194 |
| Observations | 43,178 | 14,256 | 28,922 |
| $\overline{ln(W_M)}$ | 2.135 | 1.988 | 2.214 |
| Observations | 19,452 | 6,775 | 12,677 |
| $\overline{ln(W_F)}$ | 2.088 | 1.892 | 2.178 |
| Observations | 23,726 | 7,481 | 16,245 |
| *Raw GPG in %* | 4.7 | 9.6 | 3.5 |

As our purpose in this paper is to analyze the GPG among overeducated workers as well as among properly educated workers, we first verify that a statistically significant gap in pay does not only exist by gender in the respective subsamples (overeducated individuals and properly educated individuals), but also across them. Hence, we test the hypothesis that the difference between the GPG among overeducated individuals and the GPG among properly educated individuals is significantly different from zero. Table 3.2, column (1),

[11]This value is slightly lower than that estimated by (Eurostat, 2016) in the period 2005-2014 (5.6%). This is because we keep also the self-employed, while Eurostat considers only employees in enterprises with more than ten employees.

[12]For example, Cutillo and Di Pietro (2006) find a lower pay gap for properly educated workers relative to overeducated workers in Italy.

shows that the coefficient estimate of *overfem*[13] is negative and statistically significant. The coefficient estimate is the difference of the GPG between properly and overeducated individuals; $-(\Delta^{\text{GPG}_{Over}} - \Delta^{\text{GPG}_{Proper}}) = \Delta^{\text{GPG}_{Proper}} - \Delta^{\text{GPG}_{Over}}$. Given that the difference between the GPGs among properly and overeducated individuals is highly statistically significant, we confirm the hypothesis that there is a statistically significant difference in the GPG across the subsamples and not merely within each subsample.[14] In order to analyze the GPG among overeducated individuals and the GPG among properly educated individuals, we estimate a Mincerian wage equation considering as regressors years of education, actual work experience, as well as experience squared as an indicator of the diminishing marginal utility of work experience, job tenure (years with present employer), controls for the firm size as well as a set of job characteristics (type of contract and non-wage compensatings). Additionally, we include in each wage equation a set of sectoral and occupational dummies, wave or year dummies as well as a set of variables accounting for personal characteristics. The latter includes family status, nationality, regional controls and the educational background of the parents. Table 3.3 reports the effect of overeducation on the log of hourly wages for the entire sample as well as for the overeducated and properly educated samples, respectively.[15] The estimated coefficient of *over* is highly statistically significant and negative, indicating that being overeducated has a negative effect on earnings. The wage penalty for overeducation is 4.9%.[16] The coefficient estimate of the variable *female* being negative and significant confirms the usual result in the literature: being a woman reduces earnings. Here, the female wage penalty amounts to 7.3%. This penalty is higher in the sample of overeducated individuals (9.6%) and lower in the sample of properly educated individuals (6.9%). The coefficient for the interaction term *overfem*, negative and significant, shows that women receive from being overeducated a wage penalty of 2.2%.

As the effect of overeducation on earnings was found to differ for men and women, we analyze in the next Section the incidence of overeducation for both men and women.

[13] *Overfem* is the interaction of the dummies *female* and *over*. The dummy *over* takes the value one if the individual's educational qualification is not a prerequisite to perform his or her current job and zero if the individual holds the level of education required to perform his or her current job.

[14] The coefficient estimates of *female* in column (2) and (3) of Table 3.2 represent the negative of the GPGs, i.e. $\hat{\beta}_i^{female} = -\Delta^{\text{GPG}_i}$, where *i= Over, Proper* and $\Delta^{\text{GPG}_i} = \overline{ln(W_M)}^i - \overline{ln(W_F)}^i$, for the respective subsample.

[15] The full regression output is shown in Table 3.C.1 in Appendix 3.C. Table 3.C.2 in Appendix 3.C shows the regression output by gender and over- or proper education.

[16] Cutillo and Di Pietro (2006) find a wage penalty of 4.4% associated with overeducation in a sample of university graduates; McGuinness and Sloane (2010) find a wage penalty of 4.0% for young university graduates.

Table 3.2 OLS Estimates of Log Hourly Wages with Dummies *female*, *over* and *overfem*

| | (1) | (2) | (3) |
|---|---|---|---|
| Variables | Full Sample | Overeducated Sample | Properly Educated Sample |
| female | -0.036*** | -0.096*** | -0.036*** |
| | (0.006) | (0.008) | (0.006) |
| over | -0.226*** | | |
| | (0.007) | | |
| overfem | -0.060*** | | |
| | (0.010) | | |
| Year Dummies | Yes | Yes | Yes |
| Observations | 43,178 | 14,256 | 28,922 |
| R-squared | 0.065 | 0.023 | 0.005 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table 3.3 OLS Estimates of Log Hourly Wages, Selected Variables

| | (1) | (2) | (3) |
|---|---|---|---|
| Variables | Full Sample | Overeducated Sample | Properly Educated Sample |
| over | -0.050*** | | |
| | (0.006) | | |
| female | -0.076*** | -0.101*** | -0.072*** |
| | (0.005) | (0.008) | (0.006) |
| overfem | -0.022** | | |
| | (0.009) | | |
| Year Dummies | Yes | Yes | Yes |
| Sectoral Dummies | Yes | Yes | Yes |
| Observations | 43,178 | 14,256 | 28,922 |
| R-squared | 0.336 | 0.191 | 0.352 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

### 3.6.2 Probability of Overeducation

The literature shows that the risk of overeducation differs by gender, and in many countries the share of overeducated workers among women is higher than among men.[17] In the following, we explicitly test the incidence of overeducation for men and women via tests of proportions as well as probit regressions. We have found more pronounced wage penalties of overeducation for women than for men. However, this does not necessarily imply that women are more likely to be overeducated than men. Table 3.4, Panel A, shows that in the full sample men are actually more likely to be overeducated. Panel B of Table 3.4 confirms this result: being a woman significantly reduces the probability to be overeducated. The results in the literature concerning the overeducation risk by gender are ambigous. Different estimation techniques as well as different measures for overeducation (statistical, subjective or objective) may contribute to this ambiguity (McGoldrick and Robst, 1996). The method applied for overeducation assessment does not help to explain our result, because women are significantly more likely to report overeducation under the subjective than under the objective measure (McGoldrick and Robst, 1996). Robst (2007) finds that men are more likely to be overeducated due to career-related reasons, while women are more likely to be mismatched due to family-related reasons. Also Büchel and van Ham (2003) document the selection process concerning the labor market participation of overeducated women. On the one hand, a high reservation wage can induce a woman to turn down low pay offers with low qualification requirements, thereby reducing the overeducation probability. Women facing a lack of sufficiently adequate matches might prefer to turn inactive rather than accepting a job below their skill level. Part of the measured lower risk can thus be viewed as the outcome of a selection process concerning labor market participation. On the other hand, especially jobs in the public administration allow for more time flexibility than most high-level positions. Hence, the attractiveness of these jobs is relatively higher for women increasing the overeducation risk. In our data, the former effect is the dominant one leading to a lower overeducation probability for women.

[17] In Section 3.6.4, this non-random selection process is accounted for by adjusting the estimation results for double selectivity into the labor force as well as into overeducation.

Table 3.4 Risk of Overeducation by Gender

(a) Panel A: Tests of Proportions by Overeducation

| | (1) |
|---|---|
| | Full Sample |
| Proportion Male Sample | 0.348 |
| Observations | 19,452 |
| Proportion Female Sample | 0.315 |
| Observations | 23,726 |
| *Difference* | 0.033 |
| H0: diff =0 | |
| Test statistic | 7.252 |
| P-value | 0.000 |
| H1 : *Difference* $> 0$ | |
| P-value | 1.000 |
| H1: *Difference* $< 0$ | |
| P-value | 0.000 |

(b) Panel B: Likelihood of being *female* on Overeducation – Probit Estimation

| | (1) | (2) |
|---|---|---|
| Variables | Full Sample | |
| | Reduced | Full |
| | Regression | |
| female | -0.090*** | -0.114*** |
| | (0.013) | (0.013) |
| Age | | -0.020*** |
| | | (0.001) |
| Schooling | | -0.165*** |
| | | (0.005) |
| North | | 0.019 |
| | | (0.015) |
| Centre | | 0.113*** |
| | | (0.018) |
| Italian | | -0.585*** |
| | | (0.078) |
| Married | | 0.018 |
| | | (0.016) |
| Homeowner | | -0.108*** |
| | | (0.018) |
| Max_D_Mark | | -0.222*** |
| | | (0.033) |
| Work_Climate | | 0.023** |
| | | (0.009) |
| Work_Time | | -0.013 |
| | | (0.009) |
| Work_Task | | -0.199*** |
| | | (0.010) |
| Work_Stab | | -0.056*** |
| | | (0.007) |
| Reloc | | -0.217*** |
| | | (0.027) |
| Constant | -0.394*** | 4.148*** |
| | (0.018) | (0.114) |
| Year Dummies | Yes | Yes |
| Observations | 43,178 | 43,178 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

### 3.6.3 In Search of Discrimination

In Section 3.6.1, we have found evidence that overeducation has a negative effect on earnings and that this negative effect is more pronounced for female workers. In this Section, we use the Oaxaca-Blinder standard methodology to study the GPG and its drivers. Our aim is to estimate the GPG all else equal, and to find evidence of gender discrimination in our data (if any). The two-fold decomposition in Table 3.5 shows that the endowments or explained component is negative and significant among overeducated individuals as well as among properly educated workers. This means that (average) observable female labor market characteristics are actually better than males'. The unexplained or coefficients part shows the hypothetical wage gain for women if their own features were remunerated like men's. As this term is positive and significant for both over- and properly educated individuals but is higher among mismatched workers, it suggests that gender wage discrimination may be more important among overeducated workers. The unexplained component among overeducated employees amounts to 83.4% compared to 66.7% among properly educated individuals. The unexplained part of the GPG is usually attributed to discrimination, but it is important to recall that it also captures differences in unobserved characteristics (Blau and Kahn, 2000). A reason for the high fraction of the GPG due to the unexplained part might be that our data is too poor to capture the differences in observable labor market characteristics that explain the pay gap between groups. Therefore, we check the adequacy of our data to explain differences in wages other than the GPG. The results in Table 3.6 show that the same type of wage decomposition can capture most of the differences in characteristics that explain the pay gap between groups other than gender. For example, available information on individuals and jobs can explain almost 80.0% percent of the difference in pay between over- and properly educated individuals. The comparison between several types of wage differentials shows that the GPG is by far the most unexplained among the considered groups. Hence, the high proportion of the coefficients effect in the GPG in the full sample as well as in the GPG by overeducation is not data-driven.

Table 3.5 Decomposition of the GPG by Overeducation

| | (1) | (2) |
|---|---|---|
| | Overeducated Sample | Properly Educated Sample |
| *Differential* | | |
| $\overline{ln(W_M)}$ | 1.988*** | 2.214*** |
| | (0.006) | (0.005) |
| $\overline{ln(W_F)}$ | 1.892*** | 2.178*** |
| | (0.005) | (0.004) |
| Difference | 0.096*** | 0.035*** |
| | (0.008) | (0.006) |
| *Decomposition* | | |
| Endowments | -0.024*** | -0.035*** |
| | (0.008) | (0.006) |
| Coefficients | 0.121*** | 0.070*** |
| | (0.010) | (0.008) |
| Coefficients in % (Absolute Value) | 83.4 | 66.7 |
| Observations | 14,256 | 28,922 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.14$

Table 3.6 Decomposition of the GPG versus Other Pay Gaps

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Men vs. Women | Properly Educated vs. Overeducated Individuals | Public vs. Private Sector | Full time vs. Part-time |
| *Differential* | | | | |
| $\overline{ln(W_{Group\ 0})}$ | 2.135*** | 2.194*** | 2.334*** | 2.130*** |
| | (0.004) | (0.003) | (0.003) | (0.003) |
| $\overline{ln(W_{Group\ 1})}$ | 2.088*** | 1.938*** | 1.949*** | 2.018*** |
| | (0.003) | (0.004) | (0.003) | (0.006) |
| Difference | 0.047*** | 0.256*** | 0.384*** | 0.111*** |
| | (0.005) | (0.005) | (0.005) | (0.007) |
| *Decomposition* | | | | |
| Endowments | -0.044*** | 0.200*** | 0.315*** | 0.171*** |
| | (0.005) | (0.004) | (0.009) | (0.004) |
| Coefficients | 0.091*** | 0.056*** | 0.070*** | -0.059*** |
| | (0.006) | (0.005) | (0.010) | (0.006) |
| Coefficients in % (Absolute Value) | 67.4 | 21.9 | 18.2 | 25.7 |
| Observations | 43,178 | 43,178 | 43,178 | 43,178 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes: Group 0* is the respective first prediction (male, properly educated, public sector, full-time employees) and *Group 1* is the respective second prediction (female, overeducated, private sector, part-time employees) of the decomposition.

### 3.6.4 Unbiased Estimation Results

In this Section, we analyze the GPG among overeducated workers as well as among properly educated workers controlling for selection decisions. The ignorance of individual selection decisions results in Omitted Variable Bias (OVB) and endogeneity problems. The estimated correlation between the error terms of the two binary choice equations considered, $\rho$, is statistically significant if unobserved characteristics such as individual ability influence both choices. We consider the participation choice and the decision to accept a job that does not match the individual's qualification level. Indeed, sorting into the over- or properly educated sample could be a result of observable as well as unobservable differences in characteristics between the individuals. In our data, $\rho$ is found to have a positive sign and to be significantly different from zero for the female sample but insignificant for the male sample (see Table 3.7).[18] Females choosing to participate in the labor market tend to choose jobs for which they are overeducated more often than individuals actually inactive would do if they had decided to participate. Table 3.7 shows that relocating singificantly lowers the probability of being overeducated for both men and women. Having children or young children lowers the participation probability for women but raises the probability of participating in the labor force for men.

Next, we define and present in Table 3.8 the values of the four selection variables we consider in this study for both men and women: $\lambda_{PA}^{Over}, \lambda_{PA}^{Proper}$ (participation choice) and $\lambda_{OV}^{Over}, \lambda_{OV}^{Proper}$ (overeducation choice), where *Over* identifies the overeducated sample and *Proper* the properly educated sample.[19] The coefficient estimate of $\lambda_{PA}^{Over}$ is positive and significant for both women and men indicating that individuals actually out of the labor force would earn less than their overeducated peers even if they were selected into overeducation. The coefficient estimate of $\lambda_{PA}^{Proper}$ results statistically insignificant for men but statistically significant for women. Females out of the labor force would earn less than properly matched women even if they were selected into a job in line with their level of education. Thus, only women with more favorable unobservable characteristics self-select themselves into the labor force. Men in employment do not receive a wage gain compared to men out of the labor force. Thus, inactive women would earn less than a comparably properly matched individual engaged in the labor market. As expected, the coefficient estimates for the overeducation choice result positive for overeducated workers and negative for their properly educated

[18]Table 3.D.1, Appendix 3.D, shows the full regression output for the bivariate probit of the participation and overeducation selection equations.

[19]We present in Table 3.D.2, Appendix 3.D, the full regression output with the selection correction terms. In the following, for notational simplicity; $\bar{\lambda} = \lambda$ and $\hat{\delta} = \delta$.

counterparts. The coefficient estimate of $\lambda_{OV}^{Over}$ is significant and positive for both women and men, while the estimated coefficient of $\lambda_{OV}^{Proper}$ is significant and negative for both, men and women. For overeducated individuals the same unobserved characteristics that raise the probability to be overeducated also increase wages. In the properly educated sample, the set of unobserved characteristics increasing the overeducation probability impacts negatively on the wage level. The intuition behind this positive selection into overeducation is that properly educated workers would earn more than their overeducated colleagues even if the latter were in a job matching their educational background. Overall, our data show that individuals that select into overeducation obtain lower wages than a randomly chosen individual with a similar set of observable characteristics.

Finally, we calculate the Oaxaca-Blinder decomposition when accounting for double selection using earnings equations that have been corrected for sample selection and endogeneity bias. We provide in Table 3.9 the results of this decomposition. The differential is again divided in the following parts: the endowments part, which is explained by differences in explanatory variables and the coefficients part, which is due to differences in estimated coefficients. Additionally, there are the parts accounting for gender differences in selection; *Participation* and *Overeducation*. The component attributed to gender differences in labor market participation or the participation component is: $(\delta_{M,PA}^{m}\lambda_{M,PA}^{m} - \delta_{F,PA}^{m}\lambda_{F,PA}^{m})$, with $m = Over, Proper$. Analogously, the overeducation component of the GPG is equal to; $(\delta_{M,OV}^{m}\lambda_{M,OV}^{m} - \delta_{F,OV}^{m}\lambda_{F,OV}^{m})$. In the overeducated sample, the selection coefficients for both the participation and the overeducation component are positive. The overeducation part is statistically significant and allows to explain almost the entire GPG. For overeducated individulas, the overeducation decision, exerts a strong positive impact on wages for both men and women (as shown in Table 3.8). However, the corresponding set of unobservables, $\lambda_{OV}^{Over}$, is more favorable for men, i.e. $\lambda_{M,OV}^{Over} > \lambda_{F,OV}^{Over}$. Conseqeuntly, the overeducation component is a net driver of the GPG among overeducated workers. In contrast, the overeducation component is statistically significant but negative for the properly educated sample. The set of unobservables, $\lambda_{OV}^{Proper}$, is more favorable for women than for men. Thus, the component reduces the GPG among properly educated workers significantly. Our results show that controlling for unobserved individual characteristics removes the unexplained component of the GPG among overeducated workers. Yet, it remains a main driver of the GPG among properly educated individuals. The endowments effect is still significant and negative working towards a closure of the gap for both over- and properly educated individuals.

As our results show that the discriminatory component in the Oaxaca-Blinder decomposition of the GPG disappears among overeducated workers but remains significant among properly educated individuals also when controlling for double selection, we further investigate why overeducation can fight gender discrimination in pay whereas a proper match fails to do so. Overeducated female workers compensate with higher educational attainment their lower level of (generally) unobservable characteristics. Their set of unobservables is lower than that of their properly educated colleagues and lower than that of overeducated men. Consequently, overeducation is a signaling device for women spending their useless-for-the-job diploma to inform employers on their true productivity and thereby fights gender wage discrimination. For both men and women, overeducation allows to compensate for differences in unobserved heterogeneity compared to their properly educated peers and thus is a first-best matching for overeducated workers. In contrast, even though among properly educated workers, women have more favorable sets of unobservables compared to their male peers, the discriminatory part remains a main contributor to the wage gap. As the level of education attained is required for the job performed, the signaling effect is less clear and hence does not allow to overcome gender discrimination.

Table 3.7 Bivariate Probit Results: Instruments and Correlation Coefficient for the Participation and Overeducation Decision

| | (1) | (2) | (4) | (5) |
|---|---|---|---|---|
| | Female Sample | | Male Sample | |
| Variables | Overeducation | Participation | Overeducation | Participation |
| Reloc | -0.228*** | | -0.214*** | |
| | (0.041) | | (0.035) | |
| Kids | | -0.449*** | | 0.147** |
| | | (0.037) | | (0.074) |
| Kids_3 | | -0.259*** | | 0.055 |
| | | (0.030) | | (0.092) |
| $\rho$ | 0.165*** | | 0.529 | |
| | (0.064) | | (0.440) | |
| Year Dummies | Yes | Yes | Yes | Yes |
| Sectoral Dummies | No | No | No | No |
| Observations | 31,516 | 31,516 | 21,075 | 21,075 |

Robust standard errors in parentheses

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 3.8 Selection Variables, Definition and Values

(a) Panel A: Overeducated Sample

| | (1) | (2) |
|---|---|---|
| | Female Sample | Male Sample |
| $\lambda_{PA}^{Over}$ measures the selection bias from the *participation* decision for overeducated individuals | 0.094** (0.044) | 1.318** (0.546) |
| $\lambda_{OV}^{Over}$ measures the selection bias from the *overeducation* decision for overeducated individuals | 0.498*** (0.135) | 0.478*** (0.120) |
| Observations | 7,481 | 6,775 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

(b) Panel B: Properly Educated Sample

| | (1) | (2) |
|---|---|---|
| | Female Sample | Male Sample |
| $\lambda_{PA}^{Proper}$ measures the selection bias from the *participation* decision for properly educated individuals | 0.049** (0.022) | 0.044 (0.157) |
| $\lambda_{OV}^{Proper}$ measures the selection bias from the *overeducation* decision for properly educated individuals | -0.351*** (0.083) | -0.260*** (0.095) |
| Observations | 16,245 | 12,677 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table 3.9 Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap adjusted for Double Selection by Overeducation

| | (1) | (2) |
|---|---|---|
| | Overeducated Sample | Properly Educated Sample |
| Difference | 0.096*** | 0.035*** |
| *Decomposition* | | |
| Endowments | -0.024*** | -0.033*** |
| | (0.007) | (0.006) |
| Coefficients | -0.070 | 0.186*** |
| | (0.066) | (0.029) |
| Participation | 0.010 | 0.004 |
| | (0.013) | (0.012) |
| Overeducation | 0.180*** | -0.121*** |
| | (0.060) | (0.029) |
| | | |
| Coefficients in % (Absolute Value) | 24.6 | 54.1 |
| Overeducation in % (Absolute Value) | 63.4 | 42.6 |
| | | |
| Observations | 14,256 | 28,922 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

## 3.7 Conclusion

This paper is the first study that explicitly considers the effect of the overeducation choice on the GPG. It analyzes the GPG among overeducated and properly educated workers from 2005 to 2014 in Italy. The GPG by overeducation is decomposed in an explained and an unexplained part using the Oaxaca-Blinder model. The estimates are then corrected for sample selection and endogeneity bias controlling for two selection decisions: the decision to participate in the labor market and the decision to accept a job for that the individual is overeducated. Neglecting these selection choices would lead to inconsistent and biased estimates of both the gender-specific wage equation as well as the components of the GPG.

Unlike previous studies, our results suggest that overeducation is relevant in explaining gender pay differences. The wage gap is significantly higher among overeducated compared to properly educated workers and the overeducation earnings penalty is higher for women than for men. The GPG may arise from differences in personal and job characteristics of working men and women, or may be the result of disparities in wages that persist when male and female workers have similar personal and job characteristics. In the latter case, the residual gap cannot be justified on grounds of productivity but indicates the presence of gender discrimination. By applying the Oaxaca-Blinder methodology to study the drivers of the GPG, we find that women posses better observable characteristics than men but get lower reward from their characteristics either among mis- or properly matched individuals. In our data, the explained component of the GPG halves among overeducated workers compared to properly educated (16.6% vs. 33.3%) and the unexplained or discriminatory component is higher in the overeducated sample (83.4% vs. 66.7%). Hence, we inquire, whether overeducation leads to an increase of gender discrimination in pay. We know from the literature that most of the difference in earnings between overeducated and properly educated workers are caused by a failure to control for unobserved heterogeneity (Bauer, 2002; Chevalier, 2003; Cutillo and Di Pietro, 2006; Leuven and Oosterbeek, 2011). In order to consistently estimate the gender-specific wage equations and the components of the GPG, we follow the literature and apply a bivariate selectivity model to simultaneously account for both sample selection bias and endogeneity bias (as in Tunali, 1986; Sorensen, 1989; Cutillo and Di Pietro, 2006). The unexplained part of the GPG in the Oaxaca-Blinder decomposition, i.e. the component usually attributed to discrimination, vanishes when the estimates are corrected for sample selection and endogeneity bias in the overeducated sample. The higher GPG among overeducated workers is mainly explained by less favorable sets of unobservable characteristics of overeducated women (relative to overeducated men). By compensating with higher educational attainment their lower level of unobservable personal characteristics, overeducated women indicate their low but actual productivity to employers and overcome statistical discrimination. Even though females self-selecting themselves into proper education have higher motivation and productivity levels or commitment to work compared to their male peers, the discriminatory part remains significant among properly educated workers (though decreases; from 66.7% to 54.1%). In fact, for properly matched workers the signaling effect is less clear, because education testifies human capital skills required for the job. All in all, the decomposition results adjusted for sample selection and endogeneity problems show that a significant part of the GPG can be explained by the overeducation choice.

These results are important for policy issues. In Italy, the share of individuals with tertiary education is among the lowest in the EU. If overeducation indicates the incapacity of the Italian labor market to absorb all graduates, there is an overinvestment in education and a waste of resources. However, this may not be the case if overeducated workers possess lower levels of unobservable characteristics and try to compensate them with more investment in education in order to raise their employment probability. As stated, our results suggest that this is the case in Italy. Overeducation simply compensates for lower unobservable characteristics, there is no waste of human capital, the need for greater investment in higher education is not limited, and the share of individuals with tertiary education may grow. This conclusion is particularly important for women, as their share among graduates is high and growing in Italy, and as the wage penalty for overeducation is higher in the female sample.

# Appendices

## Appendix 3.A Methodological Issues

In this Section, we outline the estimation procedure. Before estimating and decomposing the GPG, i.e. applying the standard Oaxaca-Blinder decomposition, for the distinct subsamples (overeducated individuals and properly educated individuals), we estimate a Mincer-type wage equation separately for men and women. Then, we describe the decomposition method applied. Next, we derive the selection terms, which are then included in the wage regressions and present the decomposition expression when it is accounted for double selectivity (the Oaxaca-Blinder model with double selection).

### 3.A.1 The Wage Model

Consider the following model of wage determination:

$$ln(W_i) = X_i^{'}\beta + \gamma S_i + \varepsilon_i \tag{3.A.1}$$

with $i = 1, \ldots, N$ and where $ln(W_i)$ is the natural logarithm of net hourly wages, $\beta$ is a $K \times 1$ vector of coefficients including the intercept, and $X_i$ is a $K \times 1$ vector of observable individual labor market characteristics such as schooling, work experience or tenure. $S_i$ is a dummy for overeducation[20] and $\gamma$ is the corresponding coefficient. The error term is described by $\varepsilon_i$. In order to analyze the effect of overeducation on wages, the wage model (3.A.1) is evaluated at the mean by OLS, separately for men and women:

$$\overline{ln(W_G)} = \bar{X}_G^{'}\beta_G + \gamma S_G \tag{3.A.2}$$

with $G = M, F$; $G = M$ identifies the male sample and $G = F$ identifies the female sample. $\overline{ln(W_G)}$ is the natural logarithm of net hourly wages evaluated at the mean, $\beta_G$ is a $K \times 1$ vector of coefficient estimates including the intercept and $\bar{X}_G$ is a $K \times 1$ vector of average observable labor market characteristics. $S_G$ is a dummy for overeducation.

In order to estimate the GPG for the different subsamples (overeducated individuals and properly educated individuals), the basic wage model evaluated at the mean becomes:

$$\overline{ln(W_G)}^m = \bar{X}_G^{m'}\beta_G^m \tag{3.A.3}$$

[20]i.e. $S_i$ is equal to one, if the individual $i$ is overeducated and zero otherwise.

with $m = Over, Proper$; $m = Over$ represents the overeducated individuals and $m = Proper$ identifies properly educated individuals.

### 3.A.2 The Oaxaca-Blinder Model

Starting from equation (3.A.3) and using the implicit assumptions in Oaxaca (1973) and Blinder (1973), we decompose the wage differential in two parts; endowments and coefficients:[21]

$$\begin{aligned}\overline{ln(W_M)}^m - \overline{ln(W_F)}^m &= \bar{X}_M^{m'}\hat{\beta}_M^m - \bar{X}_F^{m'}\hat{\beta}_F^m \\ &= (\bar{X}_M^{m'} - \bar{X}_F^{m'})\hat{\beta}_M^m + \bar{X}_F^{m'}(\hat{\beta}_M^m - \hat{\beta}_F^m)\end{aligned}$$

where $\overline{ln(W_G)}^m$ is again the logarithmic net wage evaluated at the mean for the respective subsample, $G = M, F$ and $m = Over, Proper$, with $\bar{X}_G^m$ and $\beta_G^m$ being $K \times 1$ vectors of average characteristics and the corresponding estimated coefficients. The first term is the endowments (or characteristics) effect that evaluates the GPG in terms of characteristics at the rate of return of female characteristics. The second term is the coefficients or wage structure effect evaluating the GPG in terms of differences in returns given female observable labor market characteristics. As the same endowments should have the same effect on earnings for both, men and women, coefficients should not differ by gender, which is why this term is often referred to as the unexplained part of the GPG. If the GPG depends mainly on differences in coefficients, this may indicate the presence of gender discrimination.

### 3.A.3 Selection Rules

Endogeneity arises from correlation of $S_i$ with the error term $\varepsilon_i$. Thus, as long as $Corr(S_i, \varepsilon_i) \neq 0$, unobservable individual characteristics influence the decision to accept a job offer for which the individual is overeducated and OLS techniques lead to inconsistent estimates of the wage model (3.A.1). Despite problems of endogeneity, non-randomness of the sample leads to sample selection bias. A non-random sample may occur, as we observe only those individuals actually participating in the labor market but not those out of the labor market. In order to account for sample selection and endogeneity bias, we set up two selection rules, one for the decision to participate in the labor market and one for the decision to accept a

[21] As Jones and Kelley (1984) show, the use of the pay structure of the higher earnings group as the non-discriminatory norm, i.e. male in the underlying case, in a two-fold model is equivalent to adding the interaction term of the three-fold model to the endowments component. Similarly, the use of the pay structure for the low earnings group in the simple decomposition is equivalent to adding the interaction term for the three-way model to the unexplained component (Li and Miller, 2012).

wage offer for which the individual is overeducated. The selection rules are described by the following relations:

$$\text{Participation Selection:} \qquad Y_{iPA}^{*} = Z_i^{'}\gamma + u_{iPA} \tag{3.A.4}$$

$$\text{Overeducation Selection:} \qquad Y_{iOV}^{*} = Q_i^{'}\alpha + u_{iOV} \tag{3.A.5}$$

where $Y_{iPA}^{*}$ represents the unobserved indices of utility that individual $i$ uses to make the decision to participate in the labor market or not and $Y_{iOV}^{*}$ represents the unobserved indices of utility that individual $i$ uses to make the decision to be overeducated or not; with $Z_i$ and $Q_i$ being $K_Z \times 1$ and $K_Q \times 1$ vectors of explanatory variables, respectively, and $u_i$ is assumed to be $N(0,1)$ with $Cov(u_{PA}, u_{OV}) = \rho$.

Each equation describing the respective selection rule has to include at least one variable that influences the corresponding decision only and, hence, is uncorrelated with wages. Moreover, these instruments have to be mutually independent. The employment decision is identified via the dummy variables *Kids* and *Kids_3*, as women with children and in particular with young children spend a significant amount of time with child-care (Martins, 2001; Mulligan and Rubinstein, 2008; Lee, 2009; Chang, 2011). We assume that these variables affect the individual propensity to participate in the labor market but not marginal productivity. For the identification of the overeducation selection equation, we use the variable *Reloc* establishing whether the individual had to relocate for his or her current job. The intuition behind is that individuals willing to relocate are more likely to find a job appropriate to their educational background (Cutillo and Di Pietro, 2006; Dolton and Silles, 2008). The willingness to relocate thus influences the probability of accepting a job that does not match the individual's educational level but is exogenous to the wage level.

The probabilities of observing a positive labor income given overeducation or proper education are the following:

$$Pr(Y_{PA}^{*} > 0, Y_{OV}^{*} > 0) = Pr(u_{PA} > -Z^{'}\gamma, u_{OV} > -Q^{'}\alpha) = G(Z^{'}\gamma, Q^{'}\alpha, \rho) \tag{3.A.6}$$

$$Pr(Y_{PA}^{*} > 0, Y_{OV}^{*} \leq 0) = Pr(u_{PA} > -Z^{'}\gamma, u_{OV} \leq -Q^{'}\alpha) = G(Z^{'}\gamma, -Q^{'}\alpha, -\rho) \tag{3.A.7}$$

where $G(\cdot)$ is the standard bivariate normal distribution and $\rho$ is the correlation coefficient between the two selection rules. The subscript *PA* identifies the participation decision, while *OV* identifies the overeducation decision. Equation (3.A.6) accounts for the probability of observing a positive wage given overeducation and equation (3.A.7) for the probability of observing a positive wage given proper education. Under the assumption that the two selection rules are not independent, that is $\rho \neq 0$, maximum likelihood of the bivariate probit leads to the following selection terms for overeducated employees, $m = Over$:

$$\lambda_{PA}^{Over} = \frac{f(Z^{'}\gamma)F[\frac{Q^{'}\alpha - \rho Z^{'}\gamma}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, Q^{'}\alpha, \rho)} \tag{3.A.8}$$

$$\lambda_{OV}^{Over} = \frac{f(Q^{'}\alpha)F[\frac{Z^{'}\gamma - \rho Q^{'}\alpha}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, Q^{'}\alpha, \rho)} \tag{3.A.9}$$

Similarly, for the subsample of appropriately educated workers, $m = Proper$, the corresponding selection terms are given by:

$$\lambda_{PA}^{Proper} = \frac{f(Z^{'}\gamma)F[-\frac{Q^{'}\alpha - \rho Z^{'}\gamma}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, -Q^{'}\alpha, -\rho)} \tag{3.A.10}$$

$$\lambda_{OV}^{Proper} = \frac{-f(Q^{'}\alpha)F[\frac{Z^{'}\gamma - \rho Q^{'}\alpha}{\sqrt{1-\rho^2}}]}{G(Z^{'}\gamma, -Q^{'}\alpha, -\rho)} \tag{3.A.11}$$

$f(\cdot)$ is the standard normal density, while $F(\cdot)$ is the standard normal distribution.

Adding the selection terms $\lambda_{PA}^{m}$ and $\lambda_{OV}^{m}$ to the earnings equations in (3.A.3) allows us to consistently estimate the earnings equation for the overeducated and properly educated subsamples, respectively. We obtain the following augmented model of wage determination (Lee, 1983; Tunali, 1986):

$$\overline{ln(W_G)}^{m} = \bar{X}_G^{m'} \hat{\beta}_G^{m} + \hat{\delta}_{G,PA}^{m} \lambda_{G,PA}^{m} + \hat{\delta}_{G,OV}^{m} \lambda_{G,OV}^{m} \tag{3.A.12}$$

where $m = Over, Proper$ and $G = M, F$.

In Section 3.6.4, we discuss the results obtained from the bivariate probit estimation for the participation and overeducation selection equations as well as the estimated correlation between the error terms of these two binary equations, $\rho$.

### 3.A.4 The Oaxaca-Blinder Model with Double Selection

The estimated components of the standard Oaxaca-Blinder decomposition may change when controlling for double selection. When accounting for double selection, the decomposition expression (4.A.1) becomes the following:

$$\begin{aligned}\overline{ln(W_M)}^m - \overline{ln(W_F)}^m &= \bar{\tilde{X}}_M^{m'} \hat{\tilde{\beta}}_M^m - \bar{\tilde{X}}_F^{m'} \hat{\tilde{\beta}}_F^m \\ &= (\bar{X}_M^{m'} - \bar{X}_F^{m'})\hat{\beta}_M^m + \bar{X}_F^{m'}(\hat{\beta}_M^m - \hat{\beta}_F^m) \\ &+ (\hat{\delta}_{M,PA}^m \lambda_{M,PA}^m - \hat{\delta}_{F,PA}^m \lambda_{F,PA}^m) + (\hat{\delta}_{M,OV}^m \lambda_{M,OV}^m - \hat{\delta}_{F,OV}^m \lambda_{F,OV}^m) \quad (3.A.13)\end{aligned}$$

where $\tilde{X}$ and $\tilde{\beta}$ contain, despite the explanatory variables $X$ and the corresponding coefficients, also the selection correction terms and their coefficients. Apart from the endowments and coefficients component, there is now also a component due to differences in the participation and overeducation decision by gender, respectively. The latter two components control for otherwise unobserved factors of the participation and overeducation choice.

# Appendix 3.B Definition of Variables

Table 3.B.1 Definition of Variables

| Variable Name | Definition |
|---|---|
| | Dependent Variables |
| Lhwage | The natural log of net hourly wages; hourly wages in Euros, net of taxes and social security contributions |
| lfp | One if the respective individual chooses to participate in the labor force, zero otherwise |
| over | One if the respective individual is overeducated, zero otherwise *Over* is also used as independent variable |
| | Independent Variables |
| | Dummy and Interaction Effects |
| female | One if the respective individual is a woman, zero otherwise |
| overfem | Interaction term between the dummy *over* and *female* |
| | Human Capital and Labor Market Characteristics |
| Educ | Years of schooling completed |

| | |
|---|---|
| Max_D_Mark | One if individual graduated with the maximum degree *110 e lode* from university, zero otherwise |
| Age | Age of individual (in years) $\in (18,64)$ |
| Age5064 | One if age is between 50 and 64 years, zero otherwise |
| Exper | Number of years of prior work experience |
| Exper2 | *Exper* squared |
| Tenure | Number of years worked for current employer |
| North | One if individual lives and works in the North of Italy, zero otherwise |
| Centre | One if individual lives and works in the Centre of Italy, zero otherwise |
| Homeowner | One if individual owns a house, zero otherwise |
| | This includes bank loan-financed houses |
| Partner_Works | One if partner or spouse of the individual is employed, zero otherwise |
| Married | One if married, zero otherwise |
| Italian | One if individual is Italian, zero otherwise |
| Educ_Moth_Uni | One if mother graduated from university, zero otherwise |
| Educ_Fath_Uni | One if father graduated from university, zero otherwise |
| Kids | One if individual has at least one child, zero otherwise |
| Kids_3 | One if age of youngest child is less or equal to three years, zero otherwise |
| Reloc | One if individual relocated in order to take the current job, zero otherwise |

Job and Firm Characteristics

| | |
|---|---|
| Work_Climate | Level of satisfaction with working climate at current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Stab | Level of satisfaction with stability of current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Time | Level of satisfaction with working time at current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Task | Level of satisfaction with tasks at current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Self_Emp | One if individual is self-employed, zero otherwise |
| Part | One if individual works part-time, zero otherwise |
| Contract_Type | One if individual holds an unlimited contract, zero otherwise |
| Big_Firm | One if individual is employed in a firm with at least 250 employees, zero otherwise |
| Med_Firm | One if individual is employed in a firm with at least 50 employees, zero otherwise |

| | |
|---|---|
| | Occupations and Industries |
| Manager | One if individual is employed in ‘intellectual professions’; scientific and highly specialized occupations |
| Intermediate_Prof | One if individual is employed in ‘intermediary positions’ in commercial, technical or administrative sectors, health services and technicians |
| Sec_02 - Sec_15 | Sectoral dummies for employment in the following sectors or industries: manufacturing, energy, construction, tourism, commerce, transport, communication, financial activities, service industry, public administration, education, health, sciences and family services, respectively |
| | Selection Correction Terms |
| $\lambda_{PA}^{Over}$ | Measures the selection bias from the participation decision for overeducated individuals |
| $\lambda_{OV}^{Over}$ | Measures the selection bias from the overeducation decision for overeducated individuals |
| $\lambda_{PA}^{Proper}$ | Measures the selection bias from the participation decision for properly educated individuals |
| $\lambda_{OV}^{Proper}$ | Measures the selection bias from the overeducation decision for properly educated individuals |

## Appendix 3.C OLS Estimates of Log Hourly Wages

## Table 3.C.1 OLS Estimates of Log Hourly Wages

| | (1) | (2) | (3) |
|---|---|---|---|
| Variables | Full Sample | Overeducated Sample | Properly Educated Sample |
| over | -0.050*** | | |
| | (0.006) | | |
| female | -0.076*** | -0.101*** | -0.072*** |
| | (0.005) | (0.008) | (0.006) |
| overfem | -0.022** | | |
| | (0.009) | | |
| Schooling | 0.055*** | 0.034*** | 0.061*** |
| | (0.002) | (0.003) | (0.002) |
| Max_D_Mark | 0.036*** | 0.022 | 0.037*** |
| | (0.009) | (0.022) | (0.010) |
| Exper | 0.018*** | 0.011*** | 0.022*** |
| | (0.001) | (0.001) | (0.001) |
| Exper2 | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) |
| Tenure | 0.003*** | 0.004*** | 0.002*** |
| | (0.000) | (0.001) | (0.000) |
| Intermed_Prof | 0.019*** | 0.057*** | -0.005 |
| | (0.006) | (0.009) | (0.008) |
| Manager | 0.130*** | 0.139*** | 0.100*** |
| | (0.008) | (0.016) | (0.010) |
| Big_Firm | 0.030*** | 0.050*** | 0.018 |
| | (0.008) | (0.013) | (0.011) |
| Med_Firm | 0.009 | 0.013 | 0.012 |
| | (0.006) | (0.009) | (0.008) |
| Self_Emp | 0.028 | 0.001 | 0.059 |
| | (0.035) | (0.060) | (0.041) |
| Part | 0.085*** | 0.110*** | 0.061*** |
| | (0.006) | (0.010) | (0.009) |
| Contract_Type | 0.066*** | 0.072*** | 0.063*** |
| | (0.007) | (0.011) | (0.009) |
| Work_Climate | -0.002 | 0.010** | -0.009*** |
| | (0.003) | (0.005) | (0.003) |
| Work_Time | 0.019*** | 0.018*** | 0.022*** |
| | (0.003) | (0.005) | (0.003) |
| Work_Task | 0.010*** | 0.009* | 0.010*** |
| | (0.003) | (0.005) | (0.004) |
| Work_Stab | 0.012*** | 0.002 | 0.016*** |
| | (0.003) | (0.004) | (0.003) |
| Married | 0.050*** | 0.046*** | 0.058*** |
| | (0.005) | (0.009) | (0.006) |
| Italian | 0.093*** | 0.090** | 0.087** |
| | (0.028) | (0.038) | (0.042) |
| North | 0.041*** | 0.085*** | 0.028*** |
| | (0.005) | (0.009) | (0.006) |
| Centre | 0.019*** | 0.047*** | 0.012* |
| | (0.006) | (0.010) | (0.007) |
| Homeowner | 0.019*** | 0.026*** | 0.014* |
| | (0.006) | (0.009) | (0.007) |
| Educ_Fath_Uni | 0.010 | -0.008 | 0.014 |
| | (0.009) | (0.019) | (0.010) |
| Educ_Moth_Uni | 0.011 | 0.037 | 0.001 |
| | (0.012) | (0.023) | (0.013) |
| Home_Time | 0.008*** | 0.003*** | 0.010*** |
| | (0.001) | (0.001) | (0.001) |
| Constant | 0.627*** | 0.914*** | 0.533*** |
| | (0.049) | (0.082) | (0.061) |
| Year Dummies | Yes | Yes | Yes |
| Sectoral Dummies | Yes | Yes | Yes |
| Observations | 43,178 | 14,256 | 28,922 |
| R-squared | 0.336 | 0.191 | 0.352 |

Robust standard errors in parentheses

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 3.C.2 OLS Estimates of Log Hourly Wages

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Overeducated Sample | | Properly Educated Sample | |
| Variables | Women | Men | Women | Men |
| Schooling | 0.037*** | 0.024*** | 0.064*** | 0.057*** |
| | (0.004) | (0.005) | (0.003) | (0.003) |
| Max_D_Mark | -0.002 | 0.064 | 0.029** | 0.043** |
| | (0.025) | (0.045) | (0.013) | (0.018) |
| Exper | 0.009*** | 0.014*** | 0.021*** | 0.025*** |
| | (0.002) | (0.002) | (0.001) | (0.002) |
| Exper2 | -0.000* | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.005*** | 0.004*** | 0.003*** | 0.001* |
| | (0.001) | (0.001) | (0.000) | (0.001) |
| Intermed_Prof | 0.062*** | 0.054*** | 0.050*** | -0.029*** |
| | (0.013) | (0.012) | (0.016) | (0.010) |
| Manager | 0.141*** | 0.140*** | 0.141*** | 0.097*** |
| | (0.024) | (0.021) | (0.018) | (0.013) |
| Big_Firm | 0.034* | 0.059*** | 0.016 | 0.013 |
| | (0.020) | (0.017) | (0.017) | (0.015) |
| Med_Firm | 0.015 | 0.008 | 0.005 | 0.019 |
| | (0.014) | (0.012) | (0.011) | (0.012) |
| Self_Emp | 0.026 | -0.024 | 0.083 | 0.022 |
| | (0.079) | (0.085) | (0.052) | (0.062) |
| Part | 0.087*** | 0.189*** | 0.049*** | 0.149*** |
| | (0.011) | (0.020) | (0.009) | (0.025) |
| Contract_Type | 0.058*** | 0.100*** | 0.041*** | 0.103*** |
| | (0.015) | (0.016) | (0.012) | (0.015) |
| Work_Climate | 0.007 | 0.012* | -0.013*** | -0.003 |
| | (0.006) | (0.007) | (0.005) | (0.005) |
| Work_Time | 0.019*** | 0.021*** | 0.026*** | 0.018*** |
| | (0.007) | (0.008) | (0.005) | (0.005) |
| Work_Task | 0.008 | 0.009 | 0.012** | 0.006 |
| | (0.007) | (0.007) | (0.005) | (0.006) |
| Work_Stab | 0.001 | 0.002 | 0.016*** | 0.021*** |
| | (0.006) | (0.007) | (0.004) | (0.005) |
| Married | 0.032*** | 0.076*** | 0.058*** | 0.055*** |
| | (0.011) | (0.013) | (0.007) | (0.011) |
| Italian | 0.093** | 0.040 | 0.062 | 0.132** |
| | (0.045) | (0.070) | (0.052) | (0.066) |
| North | 0.091*** | 0.088*** | 0.024*** | 0.040*** |
| | (0.013) | (0.012) | (0.008) | (0.009) |
| Centre | 0.046*** | 0.055*** | 0.006 | 0.021** |
| | (0.015) | (0.014) | (0.010) | (0.010) |
| Homeowner | 0.040*** | 0.013 | 0.006 | 0.022** |
| | (0.013) | (0.013) | (0.010) | (0.011) |
| Educ_Fath_Uni | -0.009 | -0.001 | 0.017 | 0.008 |
| | (0.028) | (0.025) | (0.014) | (0.015) |
| Educ_Moth_Uni | 0.048 | 0.023 | 0.044** | -0.057*** |
| | (0.032) | (0.033) | (0.018) | (0.020) |
| Home_Time | 0.003** | 0.003** | 0.009*** | 0.010*** |
| | (0.001) | (0.002) | (0.001) | (0.001) |
| Constant | 0.790*** | 1.041*** | 0.455*** | 0.484*** |
| | (0.125) | (0.115) | (0.095) | (0.089) |
| Year Dummies | Yes | Yes | Yes | Yes |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 7,481 | 6,775 | 16,245 | 12,677 |
| R-squared | 0.154 | 0.227 | 0.353 | 0.359 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

# Appendix 3.D Estimation Results when Accounting for Selection Issues

Table 3.D.1 Bivariate Probit Results Overeducation and Participation Selection Equations

| | (1) | (2) | (4) | (5) |
|---|---|---|---|---|
| | Female Sample | | Male Sample | |
| Variables | Overeducation | Participation | Overeducation | Participation |
| Age | -0.025*** | 0.028*** | -0.014*** | -0.006** |
| | (0.001) | (0.002) | (0.001) | (0.003) |
| Schooling | -0.137*** | 0.159*** | -0.197*** | 0.020 |
| | (0.007) | (0.007) | (0.008) | (0.015) |
| North | 0.051** | 0.642*** | 0.028 | 0.159*** |
| | (0.024) | (0.020) | (0.021) | (0.044) |
| Centre | 0.149*** | 0.463*** | 0.097*** | 0.047 |
| | (0.026) | (0.025) | (0.026) | (0.053) |
| Italian | -0.700*** | 0.540*** | -0.199 | 0.302 |
| | (0.084) | (0.068) | (0.143) | (0.227) |
| Married | -0.067** | -0.543*** | 0.036 | 0.253*** |
| | (0.029) | (0.037) | (0.028) | (0.076) |
| Homeowner | -0.087*** | -0.002 | -0.130*** | 0.050 |
| | (0.024) | (0.025) | (0.026) | (0.055) |
| Max_D_Mark | -0.214*** | | -0.226*** | |
| | (0.040) | | (0.056) | |
| Work_Climate | 0.033*** | | 0.012 | |
| | (0.012) | | (0.013) | |
| Work_Time | -0.021* | | -0.010 | |
| | (0.012) | | (0.013) | |
| Work_Task | -0.190*** | | -0.195*** | |
| | (0.012) | | (0.014) | |
| Work_Stab | -0.029*** | | -0.076*** | |
| | (0.009) | | (0.010) | |
| Reloc | -0.228*** | | -0.214*** | |
| | (0.041) | | (0.035) | |
| Age5064 | | 1.170*** | | 0.452*** |
| | | (0.055) | | (0.080) |
| Partner_Works | | -0.031 | | 0.138** |
| | | (0.027) | | (0.067) |
| Kids | | -0.449*** | | 0.147** |
| | | (0.037) | | (0.074) |
| Kids_3 | | -0.259*** | | 0.055 |
| | | (0.030) | | (0.092) |
| Constant | 3.763*** | -2.727*** | 3.972*** | 0.940*** |
| | (0.184) | (0.131) | (0.213) | (0.312) |
| $\rho$ | 0.165*** | | 0.529 | |
| | (0.064) | | (0.440) | |
| Year Dummies | Yes | Yes | Yes | Yes |
| Sectoral Dummies | No | No | No | No |
| Observations | 31,516 | 31,516 | 21,075 | 21,075 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table 3.D.2 OLS Estimates of Log Hourly Wages with Selection Terms

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Overeducated Sample | | Properly Educated Sample | |
| Variables | Women | Men | Women | Men |
| Schooling | -0.017 | -0.044** | 0.096*** | 0.085*** |
| | (0.017) | (0.020) | (0.007) | (0.010) |
| Max_D_Mark | -0.081** | -0.021 | 0.059*** | 0.064*** |
| | (0.034) | (0.050) | (0.015) | (0.020) |
| Exper | 0.001 | 0.010*** | 0.026*** | 0.027*** |
| | (0.003) | (0.002) | (0.002) | (0.002) |
| Exper2 | -0.000* | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.004*** | 0.004*** | 0.003*** | 0.001* |
| | (0.001) | (0.001) | (0.000) | (0.001) |
| Intermed_Prof | 0.063*** | 0.056*** | 0.057*** | -0.027*** |
| | (0.013) | (0.012) | (0.016) | (0.010) |
| Manager | 0.139*** | 0.137*** | 0.145*** | 0.098*** |
| | (0.023) | (0.021) | (0.018) | (0.013) |
| Big_Firm | 0.034* | 0.058*** | 0.017 | 0.013 |
| | (0.020) | (0.017) | (0.017) | (0.015) |
| Med_Firm | 0.015 | 0.009 | 0.006 | 0.020* |
| | (0.014) | (0.012) | (0.011) | (0.012) |
| Self_Emp | 0.031 | -0.022 | 0.087* | 0.022 |
| | (0.080) | (0.085) | (0.051) | (0.061) |
| Part | 0.089*** | 0.189*** | 0.045*** | 0.146*** |
| | (0.011) | (0.020) | (0.009) | (0.025) |
| Contract_Type | 0.058*** | 0.101*** | 0.038*** | 0.103*** |
| | (0.015) | (0.016) | (0.012) | (0.015) |
| Work_Climate | 0.019*** | 0.017** | -0.018*** | -0.004 |
| | (0.007) | (0.007) | (0.005) | (0.005) |
| Work_Time | 0.012* | 0.018** | 0.029*** | 0.019*** |
| | (0.007) | (0.008) | (0.005) | (0.005) |
| Work_Task | -0.057*** | -0.049*** | 0.044*** | 0.033*** |
| | (0.019) | (0.018) | (0.009) | (0.011) |
| Work_Stab | -0.009 | -0.020** | 0.020*** | 0.031*** |
| | (0.007) | (0.009) | (0.004) | (0.006) |
| Married | -0.003 | 0.102*** | 0.049*** | 0.058*** |
| | (0.017) | (0.016) | (0.010) | (0.014) |
| Italian | -0.111 | 0.027 | 0.219*** | 0.175** |
| | (0.082) | (0.073) | (0.061) | (0.069) |

| | | | | |
|---|---|---|---|---|
| North | 0.119*** | 0.109*** | 0.031*** | 0.038*** |
| | (0.017) | (0.013) | (0.009) | (0.010) |
| Centre | 0.105*** | 0.092*** | -0.008 | 0.008 |
| | (0.022) | (0.017) | (0.011) | (0.012) |
| Homeowner | 0.012 | -0.027* | 0.021** | 0.036*** |
| | (0.015) | (0.016) | (0.011) | (0.013) |
| Educ_Fath_Uni | -0.010 | -0.000 | 0.015 | 0.008 |
| | (0.028) | (0.025) | (0.014) | (0.015) |
| Educ_Moth_Uni | 0.047 | 0.023 | 0.047*** | -0.055*** |
| | (0.032) | (0.033) | (0.018) | (0.020) |
| Home_Time | -0.005* | -0.000 | 0.015*** | 0.012*** |
| | (0.003) | (0.002) | (0.001) | (0.001) |
| $\lambda_{PA}^{Over}$ | 0.094** | 1.318** | | |
| | (0.044) | (0.546) | | |
| $\lambda_{OV}^{Over}$ | 0.498*** | 0.478*** | | |
| | (0.135) | (0.120) | | |
| $\lambda_{PA}^{Proper}$ | | | 0.049** | 0.044 |
| | | | (0.022) | (0.157) |
| $\lambda_{OV}^{Proper}$ | | | -0.351*** | -0.260*** |
| | | | (0.083) | (0.095) |
| Constant | 1.609*** | 1.789*** | -0.612*** | -0.292 |
| | (0.287) | (0.279) | (0.233) | (0.273) |
| Year Dummies | Yes | Yes | Yes | Yes |
| Sectoral Dummies | Yes | Yes | Yes | Yes |
| Observations | 7,481 | 6,775 | 16,245 | 12,677 |
| R-squared | 0.155 | 0.230 | 0.355 | 0.360 |

Robust standard errors in parentheses

$***p < 0.01, **p < 0.05, *p < 0.1$

# References


Alba-Ramirez, A. (1993). Mismatch in the Spanish Labor Market: Overeducation?. *Journal of Human Resources*, 28(2), 259-278.

Anker, R. (1998). *Gender and Jobs: Sex Segregation of Occupations in the World*. International Labour Organization.

Arrow, K. J. (1972). Models of Job Discrimination. In Pascal, A. H. (Ed.), *Racial Discrimination in Economic Life*, Lexington Books.

Arrow, K. J. (1973). Higher Education as a Filter. *Journal of Public Economics*, 2(3), 193-216.

Baert, S., Cockx, B., and Verhaest, D. (2013). Overeducation at the Start of the Career: Stepping Stone or Trap?. *Labour Economics*, 25(C), 123-140.

Battu, H., Belfield, C. R., and Sloane, P. J. (2000). How well Can We Measure Graduate Over-education and its Effects?. *National Institute Economic Review*, 171(1), 82-93.

Bauer, T. K. (2002). Educational Mismatch and Wages: A Panel Analysis. *Economics of Education Review*, 21(3), 221-229.

Becker, G. S. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics*, 3(1, Part 2), 33-58.

Bielby, W. T., and Baron, J. N. (1986). Men and Women at Work: Sex Segregation and Statistical Discrimination. *American Journal of Sociology*, 91(4), 759-799.

Blau, F. D., and Kahn, L. M. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*, 15(4), 75-99.

Blau, F. D., and Kahn, L. M. (2006). The US Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blau, F. D., Ferber, M. A., and Winkler, A. E. (2013). *The Economics of Women, Men and Work*. Pearson Higher Ed.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4), 436-455.

Boll, C., and Leppin, J. S. (2013). *Equal Matches are Only Half the Story: Why German Female Graduates Earn 27% Less than Males* (No. 138). HWWI Research Paper.

Boll, C., Leppin, J. S., Rossen, A., and Wolf, A. (2016). Magnitude and Impact Factors of the Gender Pay Gap in EU Countries. *Report prepared for and financed by the European Commission–Directorate-General for Justice, European Union (ed.).*

Büchel, F., and Battu, H. (2003). The Theory of Differential Overqualification: Does it Work?. *Scottish Journal of Political Economy*, 50(1), 1-16.

Büchel, F., and van Ham, M. (2003). Overeducation, Regional Labor Markets, and Spatial Flexibility. *Journal of Urban Economics*, 53(3), 482-493.

Büchel, F., and Mertens, A. (2004). Overeducation, Undereducation, and the Theory of Career Mobility. *Applied Economics*, 36(8), 803-816.

Budig, M. J., and England, P. (2001). The Wage Penalty for Motherhood. *American Sociological Review*, 66(2), 204-255.

Castagnetti, C., and Rosti, L. (2013). Unfair Tournaments Gender Stereotyping and Wage Discrimination among Italian Graduates. *Gender and Society*, 27(5), 630-658.

Chang, S. K. (2011). Simulation Estimation of Two-Tiered Dynamic Panel Tobit Models with an Application to the Labor Supply of Married Women. *Journal of Applied Econometrics*, 26(5), 854-871.

Centra, M., and Cutillo, A. (2009). *Differenziale Salariale di Genere e Lavori Tipicamente Femminili*. Studi Isfol, 2.

Chevalier, A. (2003). Measuring Over-education. *Economica*, 70(279), 509-531.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role ofBiased Self-Assessments. *American Journal of Sociology*, 106(6), 1691-1730.

Correll, S. J. (2004). Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review*, 69(1), 93-113.

Cutillo, A., and Di Pietro, G. (2006). The Effects of Overeducation on Wages in Italy: A Bivariate Selectivity Approach. *International Journal of Manpower*, 27(2), 143-168.

Dolton, P. J., and Silles, M. A. (2008). The Effects of Over-education on Earnings in the Graduate Labour Market. *Economics of Education Review*, 27(2), 125-139.

Dolton, P. J., and Vignoles, A. (2000). The Incidence and Effects of Overeducation in the U.K. Graduate Labour Market. *Economics of Education Review*, 19(2), 179-198.

Dubin, J. A., and McFadden, D. L. (1984). An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, 52(2), 345-362.

England, P. (1992). *Comparable Worth: Theories and Evidence*. Transaction Publishers.

England, P., and Folbre, N. (2005). Gender and Economic Sociology. *The Handbook of Economic Sociology*, 2, 627-649.

European Commission (2012). Employment and Social Developments in Europe 2012. *JCR Science and Policy Report, Luxembourg: Publication Office of the European Union, Directorate-General for Employment, Social Affairs and Inclusion*.

Eurostat (2016). *Gender Statistics*. http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_statistics (accessed 27-5-2016).

Eurostat (2017a). *Employment and Activity by Sex and Age - Annual Data*. http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do (accessed 30-06-2017).

Eurostat (2017b). *Gender Pay Gap in Unadjusted Form*. http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (accessed 27-05-2017).

Fortin, N. M., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. *Handbook of Labor Economics*, 4, 1-102.

Frank, R. (1978). Why Women Earn Less: The Theory and Estimation of Differential Overqualification, *American Economic Review*, 68(3), 360-373.

Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. *American Economic Review*, 96(2), 1-21.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091-1119.

Groot, W. (1996). The Incidence of, and Returns to Overeducation in the UK. *Applied Economics*, 28(10), 1345-1350.

Groot, W., and Maassen van den Brink, H. (2000). Overeducation in the Labor Market: A Meta-Analysis. *Economics of Education Review*, 19(2), 149-158.

Hakim, C. (2000). *Work-lifestyle choices in the 21st Century: Preference Theory*. OUP Oxford.

Hartog, J. (2000). Over-education and Earnings: Where are We, Where Should We Go?. *Economics of Education Review*, 19(2), 131-147.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Heinze, A., Beninger, D., Beblo, M., and Laisney, F. (2003). *Measuring Selectivity-Corrected Gender Wage Gaps in the EU* (No. 03-74). ZEW Discussion Papers.

Jones, F. L., and Kelley, J. (1984). Decomposing Differences Between Groups: A Cautionary Note on Measuring Discrimination. *Sociological Methods and Research*, 12(3), 323-343.

Jovanovic, B. (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy*, 87(5, Part 1), 972-990.

Kedir, A., Kyrizi, A., and Martinez-Mora, F. (2012). *Signalling and Productivity Effects of Overeducation: Is It Really a Waste of Resources?*. Dept. of Economics, University of Leicester.

Kroch, E. A., and Sjoblom, K. (1994). Schooling as Human Capital or a Signal: Some Evidence. *Journal of Human Resources*, 29(1), 156-180.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, 76(3), 1071-1102.

Lee, L. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, 51(2), 507-512.

Leuven, E., and Oosterbeek, H. (2011). Overeducation and Mismatch in the Labor Market. *Handbook of the Economics of Education*, 4, 283-326.

Li, I., and Miller, P. W. (2012). *Gender Discrimination in the Australian Graduate Labour Market*. IZA Discussion Paper No. 6595.

Livanos, I., and Nunez, I. (2012). The Effect of Higher Education on the Gender Wage Gap. *International Journal of Education Economics and Development*, 3(1), 33-47.

Martins, M. F. O. (2001). Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics*, 16(1), 23-39.

McGoldrick, K., and Robst, J. (1996). Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification. *The American Economic Review*, 86(2), 280-284.

McGuinness, S. (2006). Overeducation in the Labour Market. *Journal of Economic Surveys*, 20(3), 387-418.

McGuinness, S., and Sloane, P. J. (2010). «Educational/Skill Mismatch» nel Mercato del Lavoro dei Laureate: Un'Analisi Comparata. *AlmaLaurea, XII Rapporto sulla condizione occupazionale dei laureati. Investimenti in capitale umano nel future di Italia ed Europa, Il Mulino, Bologna*, 101-115.

Meliciani, V., and Radicchia, D. (2016). Informal Networks, Spatial Mobility and Overeducation in the Italian Labour Market. *The Annals of Regional Science*, 56(2), 513-535.

Meng, C., Schmidt, P. (1985). On the Cost of Partial Observability in the Bivariate Probit Model. *International Economic Review*, 26(1), 71-85.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), 281-302.

Mincer, J., and Polachek, S. (1974). Family Investments in Human Capital: Earnings of Women. *Journal of Political Economy*, 82(2, Part 2), 76-108.

Mulligan, C. B., and Rubinstein, Y. (2008). Selection, Investment, and Women's Relative Wages Over Time. *The Quarterly Journal of Economics*, 123(3), 1061-1110.

Nielsen, C. P. (2011). Immigrant Over-education: Evidence from Denmark. *Journal of Population Economics*, 24(2), 499-520.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

OECD (2009). *Education at a Glance 2009, OECD Indicators*. www.oecd.org/edu/eag2009 (accessed 27-05-2015).

Olivetti, C. and Petrongolo, B. (2008). Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26(4), 621-654.

Pecoraro, M. (2016). The Incidence and Wage Effects of Overeducation using the Vertical and Horizontal Mismatch in Skills: Evidence from Switzerland. *International Journal of Manpower*, 37(3), 536-555.

Ridgeway, C. L. (2009). Framed Before We Know It: How Gender Shapes Social Relations. Gender and Society, 23(2), 145-160.

Robst, J. (2007). Education and Job Match: The Relatedness of College Major and Work. *Economics of Education Review*, 26(4), 397-407.

Rubb, S. (2003). Post-College Schooling, Overeducation, and Hourly Earnings in the United States. *Education Economics*, 11(1), 53-72.

Sasser, A. C. (2005). Gender Differences in Physician Pay Tradeoffs between Career and Family. *Journal of Human Resources*, 40(2), 477-504.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. *Journal of Economic Literature*, 31(2), 831-880.

Schmertmann, C. P. (1994). Selectivity Bias Correction Methods in Polychotomous Sample Selection Models. *Journal of Econometrics*, 60(1-2), 101-132.

Sicherman, N., and Galor, O. (1990). A Theory of Career Mobility. *Journal of Political Economy*, 98(1), 169-192.

Sicherman, N. (1991). "Overeducation" in the Labor Market. *Journal of Labor Economics*, 9(2), 101-122.

Sloane, P. J. (2014). *Overeducation, Skill Mismatches, and Labor Market Outcomes for College Graduates*, IZA World of Labor.

Solberg, E., and Laughlin, T. (1995). The Gender Pay Gap, Fringe Benefits, and Occupational Crowding. *Industrial Labor Relations Review*, 48(4), 692-708.

Sorensen, E. (1989). Measuring the Pay Disparity between typically Female Occupations and other Jobs: A Bivariate Selectivity Approach. *Industrial and Labor Relations Review*, 42(4), 624-639.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Stiglitz, J. E. (1973). Approaches to the Economics of Discrimination. *The American Economic Review*, 63(2), 287-295.

Thurow, L.C. (1975). *Generating Inequality*. Basic Books, New York.

Tunali, I. (1986). A General Structure for Models of Double-Selection and an Application to a joint Migration/Earnings Process with Remigration. *Research in Labor Economics*, 8 (Part B), 235-282.

Waldfogel, J. (1997). The Effect of Children on Women's Wages. *American Sociological Review*, 62(2), 209-217.

# Chapter 4

# Detailed RIF Decomposition with Selection – The Gender Pay Gap in Italy

## 4.1 Introduction

Gender differences in pay are a well-known phenomenon of modern labor markets. Despite the promotion of equal-pay legislation and equal-pay opportunities, differences in pay between men and women persist (Blau and Kahn, 1992; 2003; 2006; 2016; Goldin, 2014; Kahn, 2015). Adding to the broad literature on the Gender Pay Gap (GPG), research has recently focused on the estimation of the wage gap beyond the mean (Albrecht et al., 2003; 2009; Machado and Mata, 2005; Melly, 2005a; 2005b; Lucifora and Meurs 2006; Arulampalam et al., 2007; Longhi et al., 2012; Xiu and Gunderson, 2014). Analyzing the GPG along the wage distribution allows to gain additional insights compared to the mean estimation. In particular, the phenomena of glass ceiling and sticky floors, i.e. more pronounced pay gaps at the top and bottom of the wage distribution, have been revealed using quantile-regression approaches. Hence, estimation beyond the mean allows to study gender wage inequality across the wage distribution. The standard approach in modern labor economics when it comes to decomposition methods is the Oaxaca (1973) and Blinder (1973) method. Advantages of the Oaxaca-Blinder decomposition are its relatively simple implementation and intuitive approach.[1] In the literature, the unexplained component, i.e.

[1] The method is easy to implement as it is estimated via Ordinary Least Squares (OLS) and by decomposing the pay gap in an explained and unexplained part, it provides an intuitive interpretation of the results.

the part due to differences in coefficients, is thereby identified as a major contributor to the wage gap (e.g. Blau and Kahn, 2016). A detailed decomposition allows to gain information on the contribution of various personal, labor market or job characteristics to the GPG. However, it implies additional functional form restrictions to identify the various elements of the detailed decomposition. This holds in particular, when the decomposition is conducted beyond the mean (Fortin et al., 2011; Longhi et al., 2012). A popular approach is the method proposed by Machado and Mata (2005) based on Conditional Quantile Regression (CQR). The detailed decomposition, however, is path dependent, i.e. the order of the decomposition matters (Fortin et al., 2011). Moreover, the method is computationally intense as it calculates the entire conditional wage distribution and uses simulation techniques to calculate the counterfactuals. Most importantly, the method based on standard CQRs does not allow for the unconditional mean interpretation. The latter, however, is used in Oaxaca-Blinder type decompositions. Despite the estimation beyond the mean, it is important to control for group-specific sample selection. Indeed, gender differences occur when it comes to labor market participation (Heckman, 1979). Biases due to differences between men and women in the propensity to work may be important in determining the GPG and failure to account for this bias may result in inaccurate and biased estimation of the gender-specific wage equations. Consequently, also the components of the pay differential may be biased. The underlying study is applied on Italian microdata. The Italian case is particularly interesting for the study of sample selection as gender differences in labor market participation are particularly pronounced in Italy. The female labor force participation in Italy amounted to 50.6% in 2015, while the EU-28 average was at more than 64.0% in the same year (Eurostat, 2017a). Albrecht et al. (2009) extend the method by Machado and Mata (2005) to account for sample selection showing that sample selection along the wage distribution is important when considering pay differences between full- and part-time female employees in the Netherlands.

This paper uses linear Recentered-Influence-Function Ordinary Least Squares (RIF-OLS) regressions to estimate Mincer-type wage equations for men and women (Firpo et al., 2009b). Estimation at specific-quantiles is thereby based on Unconditional Quantile Regression (UQR).[2] The method has the advantage that it allows to compute a detailed decomposition in a path-independent way and that it allows for the unconditional mean interpretation of the coefficient estimates.[3] In case of concerns of nonlineartiy, the method may be combined with a reweighting scheme. For robustness, we apply the reweighting scheme proposed by DiNardo et al. (1996) and show that the results do not change significantly.

[2]Firpo et al. (2009b) despite RIF-OLS propose also RIF-Logit and a fully nonparametric estimator RIF-NP.
[3]Contrary to the approach by (e.g. Machado and Mata, 2005).

The main part of the paper focuses on the extension of the quantile-specific RIF-OLS decomposition to account for selection into employment. Thereby, consistent estimates of the components of the GPG along the wage distribution are obtained. The method by Firpo et al. (2009b) is one of the very few approaches allowing to conduct detailed Oaxaca-Blinder type decompositions beyond the mean. Thereby, the model allows to calculate the significance of gender differences in endowments as well as in coefficients at different points of the wage distribution. In the extended model, the selection effect can be attributed to the endowments and coefficients part, respectively, explicitly showing the contribution of (otherwise) unobservable characteristics to the different parts of the GPG. The selection correction terms are estimated using both parametric and semiparametric single-index models. The semiparametric binary choice models applied are the Ichimura (1993) and Klein and Spady (1993) models. The parametric model estimates the incidence of employment via probit estimation. Sample-selection bias correction is generally conducted via parametric regression models such as maximum likelihood probit or logit, which assume normally distributed errors. However, distributional assumptions may play an important role in sample selection models (Martins, 2001). Semiparametric binary choice estimators, such as the Ichimura and Klein-Spady estimator, do not require any distributional assumptions. The semiparametric selection correction terms obtained are then, via polynomial regression, included in the respective wage equations correcting for selection bias at the specific quantiles. As the semiparametric models applied are computationally costly, a two-point wild-bootstrap test based on Horowitz and Härdle (1994) compares the estimation results of the parametric and semiparametric binary choice models. The test rejects the probit specification against the semiparametric models.

In line with the literature, differences of the GPG throughout the wage distribution are found. The results suggest glass ceiling and less pronounced sticky floors. Gender wage inequality across the wage distribution is measured by the change in the GPG across the wage distribution. In this paper, we focus on the $90^{\text{th}}$, $50^{\text{th}}$ and $10^{\text{th}}$ percentiles. Thus, for the gender wage inequality measure, we estimate the 90-10, 90-50 and 50-10 wage gaps.[4] Despite measuring wage inequality between men and women using the change of the GPG at different quantiles, estimation of the variance or gini is also possible (see for example Fortin et al., 2011). The present work focuses on the difference between quantile-specific wage gaps as the phenomena of glass ceiling and sticky floors are particularly relevant when studying gender pay differences across the wage distribution. Indeed, they have been heavily

[4]For example, the 90-10 wage gap is the difference between the GPG at the $90^{\text{th}}$ and the GPG at the $10^{\text{th}}$ percentile.

discussed in the literature (Albrecht et al., 2003; Arulampalam, 2007; Xiu and Gunderson, 2014; Cardoso et al., 2016). The detailed decomposition reveals that different categories of covariates such as educational attainment, labor market characteristics or socio-demographic characteristics contribute in distinct ways to the gender gaps at as well as to the change of the gaps between different quantiles. Similarly, the respective categories contribute differently to the explained (endowments effect) or unexplained part (coefficients effect) of the respective pay gaps. The results show that selection effects explain a substantial part of the GPG that would otherwise remain unobserved or be attributed to discrimination. Moreover, the contribution of the selection component to the GPG varies across the wage distribution. The selection-corrected decomposition suggests that sample selection substantially contributes to gender differences in pay along the earnings distribution.

The major contribution of this paper is the extension of the method proposed by Firpo et al. (2009b) to control for sample selection bias and secondly the empirical application of the method to Italy. The paper is organized as follows. In Section 4.2, the estimation strategy is presented. Section 4.3 outlines the model extension with sample selection correction. Section 4.4 describes the data set used in the analysis and provides the empirical results. Section 4.5 concludes.

## 4.2 Estimation Strategy

The decomposition proposed by Oaxaca (1973) and Blinder (1973) is very popular when it comes to analyzing mean wage differences by groups. In fact, the Oaxaca-Blinder decomposition has become one of the work horses in applied economics (Fortin et al., 2011). Using assumptions of linearity and zero-conditional mean, the approach allows to decompose pay gaps between groups in detail. The method is relatively easy to implement and estimated via OLS. However, the method allows only the estimation at the mean.[5] In the literature, various extensions have been proposed in order to be able to decompose the GPG beyond the mean (e.g. Juhn et al., 1993; Donald et al., 2000; Machado and Mata, 2005). The majority of the literature focuses on CQR models (e.g. Buchinsky, 1998; Machado and Mata, 2005). Several of these approaches calculate the aggregate decomposition only and rely on various assumptions as well as are computationally intense. The latter is in particular an issue of the detailed (and not for the aggregate decomposition) beyond the mean. The method proposed by Machado and Mata (2005) that is reconsidered or applied inter alia by Albrecht et al. (2003) and Melly (2005a; 2005b) allows to conduct a detailed Oaxaca-

[5]The Oaxaca (1973) and Blinder (1973) decomposition is outlined in detail in Appendix 4.A.

Blinder type decomposition beyond the mean.[6] The approach is based on CQRs and assigns to the decomposition the effect of each single covariate for a distributional statistic (quantile, variance or gini) other than the mean. However, the method is generally path dependent, i.e. the decomposition outcome depends on the order in which the decomposition is performed (Fortin et al., 2011). In the underlying investigation, UQRs of RIFs are used to obtain a Oaxaca-Blinder type detailed decomposition beyond the mean (Firpo et al., 2009b). In the RIF-OLS model applied here, similar to the assumptions in the classical Oaxaca-Blinder method and the Machado-Mata approach, linearity is assumed. The method based on RIF regressions is, as well as the standard Oaxaca-Blinder decomposition, path independent in the sense of Gelbach (2016). The main advantage of the UQR model over the CQR model is that it allows for the unconditional mean interpretation. The latter is used in Oaxaca-Blinder decompositions and is particularly interesting for policy evaluation as it estimates the effect of regressors on the entire (unconditional) wage distribution (Borah and Basu, 2013). CQRs analyze effects over the conditional wage distribution and hence are applicable only to subgroups of the target population.[7] In cases of concerns of model misspecification due to nonlinearity, the analysis can be repeated with a semiparametric reweighting scheme.[8] The combination of the RIF-OLS decomposition with a semiparametric reweighting estimator proposed by DiNardo et al. (1996) allows to solve the problem of potential misspecification of the RIF-OLS model if linearity does not hold. The analysis shows only small differences when using the Oaxaca-Blinder type decomposition based on RIF-OLS without or with reweighting. In particular, the specification and reweighting errors are small. In Section 4.4.3, we illustrate that the decomposition outcome with and without reweighting are similar and that the the specification error due to potential nonlinearity is small. This implies that using the RIF-OLS yields a good estimate of the Unconditional Partial Quantile Effect (UQPE).[9] As the main focus of this paper is quantile-specific selection correction and as the estimates do not change significantly in the linear or non-linear model, in the following the estimation approach using RIF-OLS is outlined and then extended to allow for sample selectivity. The paper extends the RIF-OLS model such that it corrects the wage model for selection bias at the corresponding quantile. It is accounted for sample selectivity issues using three distinct

[6]Albrecht et al. (2003) and Melly (2005a; 2005b) refer to the working paper version of Machado and Mata (2005).

[7]For illustration, we compare estimates of the gender wage penalty obtained from CQRs and UQRs in Section 4.4.2.

[8]Indeed, if the assumption of linearity in the RIF-OLS does not hold, the model is misspecified. The estimation procedure with reweighting is outlined in Appendix 4.B. The results of the method without and with reweighting are summarized in Section 4.4.3.

[9]Firpo et al. (2009b) find that RIF-OLS estimates compared to RIF-Logit and the completely nonlinear RIF-NP estimates are very similar for the effect of union membership on log wages.

binary choice models; probit, Ichimura (1993) and Klein and Spady (1993). Even though the semiparametric Ichimura and Klein-Spady models are computationally costly, the paper focuses on these models for sample correction as distributional assumptions may be important in sample-selection processes (Martins, 2001; Frölich, 2006). Indeed, a specification test rejects the parametric selection model and the semiparametric approaches are found to be, especially at lower quantiles, more informative.[10] Using the proposed extension allows to compute the selection-adjusted quantile-specific Oaxaca-Blinder type decomposition of the GPG showing explicitly the contribution of sample selectivity to the quantile-specific GPGs.

### 4.2.1 RIF Regressions at Quantiles

The RIF-OLS regression model allows to estimate the effect of explanatory variables, $X$, on the unconditional quantile, $Q_\tau$, of an outcome variable, $Y$. The RIF is estimated in quantile regressions by first calculating the sample quantile $\hat{Q}_\tau$ and computing the density at $\hat{Q}_\tau$, that is $f(\hat{Q}_\tau)$ using kernel methods (Firpo et al., 2009b). Moreover, this approach relies on the indicator function $\mathbb{1}\{Y \leq Q_\tau\}$ taking value one if the condition in $\{\cdot\}$ is true, zero otherwise. Estimates for each observation $i$ of the RIF, $\widehat{RIF}(Y_i; Q_\tau)$, are then obtained by inserting $\hat{Q}_\tau$ and $f(\hat{Q}_\tau)$ in the aggregate RIF-function, defined as:

$$
\begin{aligned}
RIF(Y;Q_\tau) &= Q_\tau + IF(Y;Q_\tau) \\
&= Q_\tau + \frac{\tau - \mathbb{1}\{Y \leq Q_\tau\}}{f_Y(Q_\tau)} \\
&= \frac{1}{f_Y(Q_\tau)}\mathbb{1}\{Y > Q_\tau\} + Q_\tau - \frac{1}{f_Y(Q_\tau)}(1-\tau) \qquad (4.1)
\end{aligned}
$$

where the RIF is the first order approximation of the quantile $Q_\tau$. $IF(Y;Q_\tau)$ represents the influence function for the $\tau$th quantile. It measures the influence of an individual observation on the $\tau$th quantile. Adding the quantile $Q_\tau$ to the influence function yields the RIF. The probability density of $Y$ evaluated at $Q_\tau$ is $f_Y(Q_\tau)$.

Firpo et al. (2009b) model the conditional expectation of the RIF-regression function, $E[RIF(Y;Q_\tau)|X]$, as a function of explanatory variables, $X$, in the UQR:

$$E[RIF(Y;Q_\tau)|X] = g_{Q_\tau}(X) \qquad (4.2)$$

where a linear function $X\beta_\tau$ is specified for $g_{Q\tau}(X)$ as for example in Borah and Basu (2013). The average derivative of the UQR, $E_X\left[\frac{dg_{Q\tau}(X)}{dX}\right]$, captures the marginal effect of a

[10] The specification test is outlined in Section 4.4.4.

small location shift in the distribution of covariates on the $\tau$th unconditional quantile of $Y$ keeping everything else constant. Therefore, the coefficients, $\beta_\tau$, can be unconditionally interpreted, as $E[RIF(Y;Q_\tau)] = E_X\left[E\left(RIF(Y;Q_\tau)|X\right)\right] = E(X)\beta_\tau$. That is the unconditional expectations $E[RIF(Y;Q_\tau)]$ using the Law of Iterated Expectations (LIE) allow for the unconditional mean interpretation. In contrast, only the conditional mean interpretation is valid in the context of CQRs; $Q_\tau(Y|X) = X\beta_\tau^{CQR}$, where $\beta_\tau^{CQR}$ can be interpreted as the effect of $X$ on the $\tau$th conditional quantile of $Y$ given $X$. The LIE does not apply here; $Q_\tau \neq E_X[Q_\tau(Y|X)] = E(X)\beta_\tau^{CQR}$, where $Q_\tau$ is the unconditional quantile. Hence, $\beta_\tau^{CQR}$ cannot be interpreted as the effect of increasing the mean value of $X$ in the unconditional quantile $Q_\tau$. This is one pitfall of CQRs in decomposition methods. The unconditional mean interpretation is important for decompositions in the sense of Oaxaca (1973) and Blinder (1973). Indeed, Oaxaca-Blinder decompositions use the unconditional mean interpretation of $\beta_\tau$, i.e. the interpretation of $\beta_\tau$ as the effect of increasing the mean value of $X$ on the mean value of $Y$. In UQR, the coefficients $\beta_\tau$ can thus be estimated by OLS in the following way:

$$Q_\tau = E[RIF(Y;Q_\tau)] = E_X\left[E\left(RIF(Y;Q_\tau)|X\right)\right] = E(X)\beta_\tau \tag{4.3}$$

The basic wage equation of the RIF-OLS model at quantile $\tau$, with $\tau \in (0,1)$, is then:

$$RIF(Y;Q_\tau) = X\beta_\tau + u_\tau \tag{4.4}$$

where $Y$ is the natural logarithm of hourly earnings and $X$ is a vector of $K$ explanatory variables (including the constant), $\beta_\tau$ is the corresponding coefficient vector and $u_\tau$ is the corresponding error term. The coefficient vector of the unconditional quantile is defined as:

$$\hat{\beta}_\tau = (\sum_{i=1}^{N} X_i'X_i)^{-1}\sum_{i=1}^{N} X_i'\widehat{RIF}(Y_i;Q_\tau) \tag{4.5}$$

UQRs estimate the effect of covariates on all parts of the earnings distribution and are thus particularly interesting for policy implications or evaluation. CQRs do not allow to draw conclusions about the impact of a variable on the overall earnings distribution but rather provide insights about the dispersion of earnings within different subgroups of the target population (Borah and Basu, 2013).

### 4.2.2 Decomposition

Given the assumptions that the mean of the RIF-function is equal to the actual quantile as well as to the mean of the conditional expectation given $X$ shown in equation (4.3), we have:

$$\begin{aligned} E[RIF(Y_M;Q_\tau)|X_M] - E[RIF(Y_F;Q_\tau)|X_F] &= \bar{X}_M\beta_{M\tau} - \bar{X}_F\beta_{F\tau} \\ &= \Delta_\tau \end{aligned}$$

where $\hat{\Delta}_\tau$ is the GPG at the $\tau$th quantile and $M = Male$ and $F = Female$.

The GPG is, as in the standard two-fold Oaxaca-Blinder decomposition, decomposed in an endowments (explained) and a coefficients (unexplained) component. The decomposition has then the following form:

$$\begin{aligned} \hat{\Delta}_\tau &= \hat{\Delta}_{E\tau} + \hat{\Delta}_{C\tau} \\ &= (\bar{X}_M - \bar{X}_F)\hat{\beta}_{F,\tau} + \bar{X}_M(\hat{\beta}_{M,\tau} - \hat{\beta}_{F,\tau}) \end{aligned} \tag{4.6}$$

where the index $E$ indicates the *Endowments Effect* and the index $C$ the *Coefficients Effect*.

To perform a detailed decomposition, the contribution of each element of the vector of explanatory variables $\bar{X}$ on both components is estimated. For identification, a detailed decomposition underlies thus stronger assumptions such as functional form restrictions as well as potentially further restrictions on the distribution of the error term. An example is the assumption of independence of the set of covariates and the dummy identifying group membership (Fortin et al., 2011). In the RIF-OLS model, the detailed components can be estimated in the same way as in the detailed Oaxaca-Blinder decomposition at the mean. However, as in the Oaxaca-Blinder decomposition at the mean, the decomposition based on RIF-OLS changes according to the choice of the reference category (Reimers, 1983; Cotton, 1988; Neumark, 1988; Oaxaca and Ransom, 1994). We follow the standard case and use male coefficients as the non-discriminatory wage structure. As in standard detailed Oaxaca-Blinder decompositions at the mean, the contribution of the single regressors to the components of the GPG are path independent also in the RIF-OLS framework.

## 4.3 Accounting for Selection

The estimation strategy outlined in Section 4.2.1 yields inconsistent and biased estimates of the wage equation and hence of the decomposition parts due to non-randomness of the sample (Heckman, 1979; 1990; Buchinsky, 1998; Albrecht et al., 2009). Indeed, the observed individuals with a positive labor income may be a non-random subsample of the individuals in the population. As the origin of the selection could be related to earnings, it is essential to explicitly consider the selection process in the estimation of the wage equation. The selection into wage work may depend on some positive factors such as individual ability, motivation or educational quality, raising both, the probability of being employed and wages. However, it is omitted in the earnings equation as these factors are unobservable in the data. The incidence of receiving a wage offer may not only be non-random but also different for men and women. Using the proposed extension of the quantile-specific wage model allows to obtain consistent estimates as well as to attribute the selection effect to the endowments and coefficients part of the quantile-specific GPGs. The estimation procedure consists, similar to Heckman (1979), of two steps. In a first-step estimation, the semiparametric estimator of the selection parameter is estimated. In a second-step regression, the selectivity-corrected model is estimated. The main difference compared to Heckman (1979) is that here the estimated selection terms are estimated with distribution-free approaches rather than by a parametric method (Newey, 2009). Moreover, instead of using only the traditional Inverse Mills Ratio (IMR), an approximation function is used.

The selection decision of interest is the employees' work decision. The decision is identified by the indicator variable $E_i$, which is equal to one if individual $i$ is in employment and zero otherwise. The reservation wage, $Y_i^{res}$, is not observed but we observe, whether the difference between the market wage, $Y_i$, and $Y_i^{res}$ is positive or not. Hence, $E_i = 1$ if $Y_i - Y_i^{res} > 0$, $E_i = 0$ otherwise. In a first-step estimation, the selection equation of the single-index model is estimated with Semiparametric Least Squares (SLS) and reads as:[11]

$$E_i = m(Z_i\gamma) + v_i \tag{4.7}$$

where $Z_i$ is a $1 \times T$ vector of regressors influencing the employment decision with $t = 0, \ldots, T$. The corresponding parameters are contained in the $T \times 1$ column vector $\gamma$ and $v_i$ is the usual additive error term, which is assumed to be uncorrelated with $Z_i$. The function $m(\cdot)$ is an unknown link or smooth function. Contrary to parametric models, in semiparametric single-index models, not only $\gamma$ but also $m(\cdot)$ must be estimated.[12] The set of covariates $Z$ includes at least one variable not included in $X$ and uncorrelated with the log of hourly

[11] The parametric selection equation has the following form: $E_i = Z_i\gamma + v_i$.

[12] The general form of single-index models is: $E_i = m\big(\phi(Z_i, \gamma)\big) + v_i$, where $m(\cdot)$ is the unknown smooth function and $\phi(\cdot)$ is a known parametric function with $T$ regressors, and coefficient vector $\gamma$ having dimension $P \times 1$ (Ichimura, 1993; Henderson and Parmeter, 2015). As $\phi(Z_i, \gamma)$ is a scalar, it is necessarily single index. Similar to many other studies, we assume a linear single-index, and thus the number of regressors and parameters are equal, i.e. $T = P$.

wages $Y$ (the underlying dependent variable) but influencing the work decision. This is important for identification of the selection decision. Moreover, if the regressors in $Z$ are not different from the variables in $X$, the selectivity-corrected regression will be highly collinear. Semiparametric single-index models (such as the Ichimura and Klein-Spady model) are quite popular in nonparametric estimation as they work similar to parametric models (Henderson and Parmeter, 2015). However, no distributional assumptions are required to set up these models, while in the probit model, the standard normal distribution is assumed. Using the semiparametric single-index models that do not require any distributional assumptions allows to circumvent potential bias of the selection-correction terms due to non-normality of the selection process. Indeed, distributional assumptions may be important when considering sample-selection processes (Martins, 2001).

The semiparametric single-index models used to estimate the selection equation are iterative procedures and hence are computationally heavy given that nonparametric kernel estimation is conducted at each iteration. For the estimation, the second-order Gaussian kernel is used and the bandwidth is selected by likelihood cross-validation. The SLS estimator is consistent and independent of the distribution of $v$ (Buchinsky, 1998). The Klein-Spady model achieves the semiparametric efficiency bound for binary choice models, while the Ichimura estimator is inefficient if the model suffers from heteroskedasticity (Ichimura, 1993).[13] Buchinsky (1998), as well as Albrecht et al. (2009) and Chzhen and Mumford (2011) use power series estimation in order to estimate the correction term in the CQR model. We replace the power series by polynomials of order $j$.[14] The following polynomial of order $j$ is estimated:

$$\hat{h}_\tau(\hat{m}) = \hat{\delta}_\tau PS(\hat{m}) \tag{4.8}$$

where $PS(\hat{m})$ is a polynomial vector in $m$:

$$PS(\hat{m}) = \Big[PS_1(\hat{m}), \ldots, PS_J(\hat{m})\Big]$$

and $PS_j(\hat{m}) = \lambda(Z_A\hat{\gamma})^j$ with $j = 1, 2, \ldots, J$. The correction term $\hat{h}_\tau(\hat{m})$ is an approximation of the unknown function for selection correction; $\hat{h}_\tau(\hat{m}) \to h_\tau(m)$ as the number of parameters goes to infinity. The nonlinear function $\lambda$ is the standard IMR[15] and $\hat{\delta}$ are the corresponding

[13] The probit estimate is efficient under normally distributed errors (e.g. Buchinsky, 1998).

[14] Using orthogonal polynomials allows to rule out multicollinearity issues (see Newey, 2009, for further details).

[15] with $\lambda = \frac{\phi(\cdot)}{\Phi(\cdot)}$ being the usual IMR; $\phi(\cdot)$ is the probability density function, $\Phi(\cdot)$ the cumulative distribution function.

coefficient estimates, which vary with the specific quantile $\tau$. The index $A$ denotes individuals accepting a wage offer. The parameter estimates $\hat{\gamma}$ are estimated via semiparametric single-index methods (Ichimura and Klein-Spady). The correction was shown to be asymptotically normal (Newey, 2009). In this study second-order polynomials are used as polynomials allow for more flexibility than standard parametric selection models (Carneiro et al., 2011; Cornelissen et al., 2016). Even though second-order polynomials rule out a nonmonotonic shape of $\hat{h}(\cdot)$, we focus on polynomials of order two as higher order polynomials may become instable at the boundaries of the data space (Harrell, 2015).

Estimation of semiparametric selection models in the way described above does not allow for identification of the level of the constant and the first reported continuous variable (Buchinsky, 1998).[16] Therefore, we normalize the respective coefficients from the semiparametric single-index estimations to the corresponding parameter estimates obtained from the probit model.[17] This way of normalizing the coefficients allows also for a better comparison of the Ichimura and Klein-Spady estimation outcome with the probit estimation outcome (Albrecht et al., 2009; Chzhen and Mumford, 2011). We have then the following expression:

$$\hat{h}^*_\tau(\hat{m}^*) = \hat{\delta}^*_\tau PS(\hat{m}^*) \tag{4.9}$$

with $PS_j(\hat{m}^*) = \lambda^*(Z^*_A\hat{\gamma}^*)^j$, where $Z^*_A = (1, Z_{A,1}, Z_{A,T-2})$ and $\hat{\gamma}^* = (\hat{\gamma}^*_0, \hat{\gamma}^*_1, \hat{\gamma}_{T-2})^T$ include the normalized constant $\hat{\gamma}^*_0$ as well as the normalized coefficient estimate of the first continuous variable $\hat{\gamma}^*_1$ from the selection decision. The coefficient vector $\hat{\gamma}_{T-2}$ includes all the remaining parameter estimates from the single-index models. The (nonlinear) function or the IMR, $\lambda^*(Z^*_A\hat{\gamma}^*)$, is estimated and depends on the normalized constant, the normalized coefficient estimate of the first continuous variable in $Z$ as well as on the other variables in $Z$, $\hat{\delta}^*_\tau$ contains the corresponding coefficient estimates.

In the second-step estimation, the function for selection correction $\hat{h}^*_\tau(\cdot)$ is included in the basic wage equation, i.e. equation (4.4), correcting for selection at the $\tau$th quantile. Thereby, $\hat{h}^*_\tau(\cdot)$ acts as the IMR does in the Heckman (1979) two-step procedure but is quantile-specific and does not require any distributional assumptions of the error terms of the selection process.

[16] The semiparametric estimators require scale and local normalization (Buchinsky, 1998; Newey, 2009).

[17] For an overview of normalization in single-index models see for example Cameron and Trivedi (2009).

The wage equation corrected for selectivity bias at the $\tau$th quantile with $j = 2$ looks as follows:

$$\begin{aligned}
\widehat{RIF}(Y;Q_\tau) &= X\hat{\beta}_\tau + \hat{h}^*_\tau(\hat{m}^*) + \hat{\varepsilon}_\tau \\
&= X\hat{\beta}_\tau + \hat{\delta}^*_\tau PS(\hat{m}^*) + \hat{\varepsilon}_\tau \\
&= X\hat{\beta}_\tau + \hat{\delta}^*_{1\tau}\lambda^*(Z^*_A\hat{\gamma}^*)^1 + \hat{\delta}^*_{2\tau}\lambda^*(Z^*_A\hat{\gamma}^*)^2 + \hat{\varepsilon}_\tau
\end{aligned} \tag{4.10}$$

where $Y$ is the natural logarithm of hourly earnings and $X$ is a vector of $K$ explanatory variables, the selection correction term $\lambda^*(Z^*_A\hat{\gamma}^*)^j$ is a function evaluated at $Z^*_A\hat{\gamma}^*$. The corresponding coefficient vectors are $\hat{\beta}_\tau$ and $\hat{\delta}^*_{j\tau}$ with $j = 1,2$. For equation (4.10) to hold, the following assumptions are made. The reservation and the market wage depend on unobservables, the joint distribution of $u$ and $v$ is continuous and the probability of observing a positive difference $(Y - Y^{res})$ given $Z$, depends only on $Z\gamma$.[18] The selectivity-corrected coefficient estimates are consistent and asymptotically normal distributed. This holds under the assumption that the second-stage estimation successfully corrects for the selection bias (see Appendix 4.C for further details). The consistent coefficient estimates are then obtained from RIF-OLS regression of $\widehat{RIF}(Y;Q_\tau)$ on $X$, $\lambda^*(Z^*_A\hat{\gamma}^*)^1$ and $\lambda^*(Z^*_A\hat{\gamma}^*)^2$.

The parametric selection correction is conducted as in the standard OLS model adjusted for sample selectivity, i.e. the IMR is added as a regressor to the earnings equation (Heckman, 1979). The RIF-OLS model with parametric selection correction is presented in Appendix 4.D.

# 4.4 Empirical Application

## 4.4.1 Data and Descriptive Statistics

The empirical analysis is based on the survey PLUS[19] from the Italian Institute for the Development of Vocational Training for Workers (ISFOL). The survey is particularly relevant for the study of wage inequality by gender as it delivers broad information on the personal working profiles and individual motivation of the interviewees.

The underlying study uses the complete release of panel dimension.[20] The estimation is based on a pooled regression model including wave or year dummies as explanatory

[18]Similar to the assumptions made by Buchinsky (1998) for sample correction in CQRs.

[19]PLUS = Participation, Labor, Unemployment Survey

[20]Up to now, ISFOL has released the following data waves of the survey PLUS: 2005, 2006, 2008, 2010, 2011, 2014.

variables. Individuals enter as well as leave the sample (through attrition). Hence, the composition of the sample changes. The analysis is restricted to the private sector only as there has been a general 'wage freeze' in the Italian public sector at the beginning of the 21st century (Bordogna, 2012; Piazzalunga and Di Tommaso, 2015). This policy disproportionately affected women as women are more likely to work in the public sector. Consequently, the policy influenced the GPG. The analysis focuses on employees working at least 15 and maximally 45 hours per week. Self-employed, students, pensioners as well as other inactive and involuntarily unemployed individuals are excluded from the analysis. The selection decision of interest is thus the employment or work decision from search or voluntary unemployment. We consider only labor income from the main job (defined as the job that pays the highest wage). After deleting observations with missing values on other variables of interest, we are left with a sample size of 24,267 individual wage observations in the private sector[21], of which 11,390 are female and 12,877 are male. This study uses the log of hourly wage as dependent variable. It is defined as the net monthly wage perceived divided by the number of actual working hours. An alternative are monthly gross earnings, which, however, are almost entirely missing (98.0% of all observations are missing). As a second alternative, gross annual earnings could be used. However, dividing gross annual earnings by the number of months in a calendar year (plus an additional 13th month), gives a difference amounting on average to more than 800 Euros per month between the artificially created monthly gross income and the reported monthly gross income. Therefore, we prefer to use the monthly-based net income as dependent variable. Individuals with children are granted tax credits in Italy.[22] As the tax credit is granted yearly, it does not impact on the monthly perceived net income and hence having children does not directly affect monthly perceived net wages in Italy. The explanatory variables used in the regression analysis are grouped in the following categories: *Education*, *Experience*, *Job Characteristics*, *Occupations and Industries*, *Socio-Demographic Background* and *Selection*. The set of regressors labeled *Education* contains variables controlling for the level of educational attainment as well as for excellence in education. The category *Experience* includes labor market experience and labor market experience squared as well as job tenure. *Job Characteristics* include job-specific variables such as wage compensatings (the level of satisfaction with the working climate, with work place stability, with the working time as well as with the tasks at the current job). These job characteristics influence the level of (net hourly) wages as employers offering lower

[21] 17,798 observations of the public sector have been dropped. The sample initially consists of 159,615 observations in total.

[22] In order to be eligible to the grant, annual gross earnings need to be below 95,000 Euro (see Worldwide Tax Summaries (2017) for further information).

wages, may compensate their employees with more satisfactory job characteristics (Filer, 1985). Additionally, dummies controlling for the kind of contract (part-time and unlimited) are included. The set of explanatory variables *Occupations and Industries* contains sectoral and occupational dummies, while the category *Socio-Demographic Background* contains geographic controls as well as a dummy accounting for whether the individual holds the Italian citizenship. The category controls also for the family status (married or not) and the educational background of the parents (whether they have graduated from university). This category controls for any potential labor-market favoritism or discrimination coming from informal social networks. Indeed, informal networks may be important in Italy and may directly influence the wage level (Pistaferri, 1999).[23] Time-varying characteristics are caught by wave dummies and are included in this category.[24] The category *Selection* contains the selection correction terms. A complete list of variables used in the study along with their categories and definitions can be found in Appendix 4.E, Table 4.E.1.

Table 4.1 reports means and standard deviations for some of the variables included in the analysis. Women have on average higher educational attainment than men, while men have more years of labor market experience (*Exper*) and work on average longer for the same firm (*Tenure*) than women. The underlying sample shows no huge differences in the level of satisfaction with particular job characteristics between men and women. However, differences in the type of contract are found. Women have much more often than men a part-time contract, while male employees have more often an unlimited contract than female employees. There are no significant differences in geographic indicators between women and men (*North* and *Centre)*. Most of the individuals observed are Italian citizens (*Italian*). Men and women are relatively equal in terms of marriage (*Married*) as well as in having children at all (*Kids*). Yet, female employees have more often children with less than ten years (*Kids_10*) compared to male employees. Female workers engaged in the labor market are about four years younger than male employees (*Age*). The variables *Age*, *Kids* and *Kids_10* are included in the selection equation only.

[23]Individuals with access to these networks are more likely to obtain more attractive and thus generally better-paying jobs.

[24]If not stated differently, the category *Occupations and Industries* contains sectoral dummies and the category *Socio-Demographic Background* contains year or wave dummies.

Table 4.1 Descriptive Statistics

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Women | | Men | |
| | Mean | Std.Dev. | Mean | Std.Dev. |
| *Education* | | | | |
| Elementary_School | 0.015 | 0.120 | 0.016 | 0.127 |
| High_School | 0.584 | 0.493 | 0.590 | 0.492 |
| University_Degree | 0.251 | 0.433 | 0.165 | 0.371 |
| Max_D_Mark | 0.039 | 0.194 | 0.020 | 0.139 |
| | | | | |
| *Experience* | | | | |
| Exper | 13.942 | 11.307 | 17.813 | 13.396 |
| Tenure | 8.380 | 8.861 | 11.992 | 11.636 |
| | | | | |
| *Job Characteristics* | | | | |
| Work_Climate | 3.079 | 0.852 | 3.055 | 0.825 |
| Work_Stab | 2.937 | 0.982 | 2.985 | 0.949 |
| Work_Time | 3.022 | 0.849 | 3.021 | 0.806 |
| Work_Task | 3.043 | 0.777 | 3.009 | 0.771 |
| Part | 0.251 | 0.434 | 0.054 | 0.227 |
| Contract_Type | 0.761 | 0.426 | 0.818 | 0.386 |
| | | | | |
| *Occupations* | | | | |
| Manager | 0.111 | 0.314 | 0.111 | 0.314 |
| Intermed Prof | 0.555 | 0.497 | 0.405 | 0.491 |
| | | | | |
| *Socio-Demographic Background* | | | | |
| North | 0.554 | 0.497 | 0.522 | 0.500 |
| Centre | 0.211 | 0.408 | 0.198 | 0.398 |
| Italian | 0.988 | 0.110 | 0.995 | 0.074 |
| Married | 0.480 | 0.500 | 0.446 | 0.497 |
| | | | | |
| Age | 34.920 | 10.508 | 37.866 | 12.901 |
| Kids | 0.481 | 0.500 | 0.461 | 0.498 |
| Kids_10 | 0.300 | 0.458 | 0.230 | 0.421 |
| | | | | |
| Observations | 11,390 | | 12,877 | |

### 4.4.2 The Effect of Women on Earnings and Raw GPGs

It is a well-known result in the literature that women perceive lower wages than men – other things equal (e.g. Blau and Kahn, 1992; 2003; England, 2006; Grove et al., 2011). Table 4.2 shows the unadjusted GPG at the mean and at different quantiles (Panel A) as well as the 90-10, 90-50 and 50-10 wage gap (Panel B). The raw mean GPG in log hourly wages in the underlying sample amounts to 11.8 percentage points. Arulampalam et al. (2007) find for the Italian private sector in the period 1995-2001 a mean wage gap between men and women equal to 15.3 percentage points. They find a GPG amounting to 14.5 percentage points at the $10^{th}$ percentile, to 13.0 percentage points at the $50^{th}$ percentile and to 19.4 percentage points at the $90^{th}$ percentile.[25] In the underlying sample quantile-specific GPGs are equal to 11.7, 10.0 and 17.9 percentage points at the $10^{th}$, $50^{th}$ and $90^{th}$ percentile, respectively. Glass ceiling and sticky floors are assumed to exist in an economy, when the $90^{th}$ and $10^{th}$ percentile GPG, respectively, exceeds the reference percentile wage gap by at least two percentage points (e.g. Arulampalam et al., 2007). Indeed, Table 4.2 shows that class ceiling is found in the underlying study for the Italian private sector; the $90^{th}$ percentile wage gap exceeds the $10^{th}$ percentile GPG by 6.2 percentage points and the $90^{th}$ percentile wage gap exceeds the median pay gap by almost 8.0 percentage points. The $10^{th}$ percentile pay gap lies slightly above the $50^{th}$ pay gap (the 50-10 wage gap is slightly negative). As the 50-10 wage gap is slightly lower than 2 percentage points (in absolute terms), only weak evidence for sticky floors is found. This result is in line with the finding of Arulampalam et al. (2007) finding a 50-10 wage gap of -1.9 percentage points for the Italian private sector. Hence, the pay gap between men and women varies significantly between the top and bottom or median and the bottom and median of the wage distribution. Yet, in the latter case, the difference is less pronounced. This finding underlines the importance of considering the GPG at different quantiles and not only at the mean. Indeed, policy implications may change according to whether the gap at different quantiles or at the mean is considered. In particular, not only the magnitude of the raw GPG but also the decomposition may vary across the wage distribution. Similarly, selection effects may change across the distribution. Even when assuming that men and women have the same set of observable labor market characteristics, i.e. considering the unexplained component[26], there is a substantial (adjusted) GPG at the mean as well as along the wage distribution due to differences in returns to observable labor market characteristics

[25] Eurostat finds for the period considered in this study (2005-2014) an average raw GPG in hourly wages equal to 5.6% for Italy as a whole, i.e. for the public and private sector (Eurostat, 2017b).

[26] Following the standard set-up of the Oaxaca-Blinder decomposition, the female set of labor market characteristic, $\bar{X}_F$, is used.

(see Table 4.2, Panel A).[27] This implies that the Italian private sector suffers from a wage gap that is mainly due to the unexplained component, also referred to as discrimination. Even though the coefficients component, i.e. the portion of the GPG not due to gender differences in observed characteristics, is generally taken to be an estimate of gender discrimination, the unexplained portion of the GPG may include effects of unobserved productivity, innate ability or other unobserved characteristics (Blau and Kahn, 2006). Hence, the unexplained component or adjusted GPG may change, when it is accounted for sample selection.

Table 4.2 GPG at Different Quantiles and across the Wage Distribution

(a) Panel A: GPG at the Mean and at Different Quantiles

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Mean | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG | 0.118*** | 0.117*** | 0.100*** | 0.179*** |
| (Unadjusted Gap) | (0.005) | (0.010) | (0.004) | (0.009) |
| Adjusted GPG | 0.124*** | 0.115*** | 0.097*** | 0.160*** |
| | (0.006) | (0.010) | (0.004) | (0.009) |

Bootstrapped standard errors in parentheses, 100 replications
$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes:* The unadjusted GPG is equal in magnitude to the raw GPG. The adjusted GPG is the unexplained or coefficients part of the decomposition. The wage gaps have been estimated using the decomposition model outlined in Section 4.2.2.

(b) Panel B: GPG across the Wage Distribution

| | (1) | (2) | (3) |
|---|---|---|---|
| | 90-10 | 90-50 | 50-10 |
| Unadjusted Change | 0.062*** | 0.079*** | -0.018*** |
| | (0.011) | (0.009) | (0.008) |

Bootstrapped standard errors in parentheses, 100 replications
$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes:* The unadjusted change is the change in the unadjusted GPG from the top to the bottom and median, column (1) and (2), as well as from the median to the bottom quantile, column (3).

[27]The full estimation outcome from the standard decomposition at the mean is shown in Table 4.A.1, while the regression output from the RIF-OLS decomposition is presented in Table 4.4.

Table 4.3 reports coefficient estimates for the dummy variable *female* of a Mincer-type wage model for the $10^{th}$, $50^{th}$ and $90^{th}$ percentile using standard OLS, RIF-OLS and CQR.[28] The effect of being a women is, as expected, strictly negative all along the wage distribution as well as at the mean. According to the OLS estimate, for female employees earnings are reduced by approximately 11.5 percentage points. The UQR and CQR show as well that being a women decreases earnings in the corresponding quantile of the conditional or unconditional earnings distribution. The unconditional (negative) effect of women on log hourly earnings decreases in absolute terms from the bottom to the median and increases thereinafter sharply. The conditional effect decreases slightly from the $10^{th}$ to the $50^{th}$ percentile and increases thereafter. Figure 4.1 plots the effect of being female on log hourly wages for both quantile methods.[29] The partial effect from the UQR is highly nonmonotonic, while the partial effect from the CQR shows almost a linear pattern from the $20^{th}$ percentile onwards. Both, Table 4.3 and Figure 4.1 show that the magnitude of the estimation results changes depending on which approach (UQR or CQR) is used.

Table 4.3 OLS, UQR and CQR of Log Hourly Wages – Gender Wage Penalty

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | $10^{th}$ Percentile | | $50^{th}$ Percentile | | $90^{th}$ Percentile | |
| | OLS | UQR | CQR | UQR | CQR | UQR | CQR |
| female | -0.122*** | -0.115*** | -0.117*** | -0.106*** | -0.112*** | -0.181*** | -0.150*** |
| | (0.005) | (0.012) | (0.007) | (0.005) | (0.004) | (0.012) | (0.009) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Robust standard errors in parentheses for OLS
Bootstrapped standard errors in parentheses, 100 replications, for UQR and CQR

In the following, the detailed decomposition results at specific quantiles as well as across the wage distribution using the 90-10, 90-50 and 50-10 wage gaps as inequality measures are presented. Then, the estimation results of the parametric and semiparametric binary choice models are outlined and the set-up of the test for equality of the parametric and semiparametric models as well as the results from the test are discussed. Finally, the decomposition outcome with selection adjustment is shown.

[28]The full regression output of the Mincer-type wage model using OLS, UQR and CQR, respectively, is shown in Appendix 4.F, Table 4.F.1. For all three model specifications, the same set of regressors is used.

[29]The CQ and UQ partial effects are evaluated at the 0.05, 0.1, 0.15,…,0.90, 0.95 quantile, respectively.

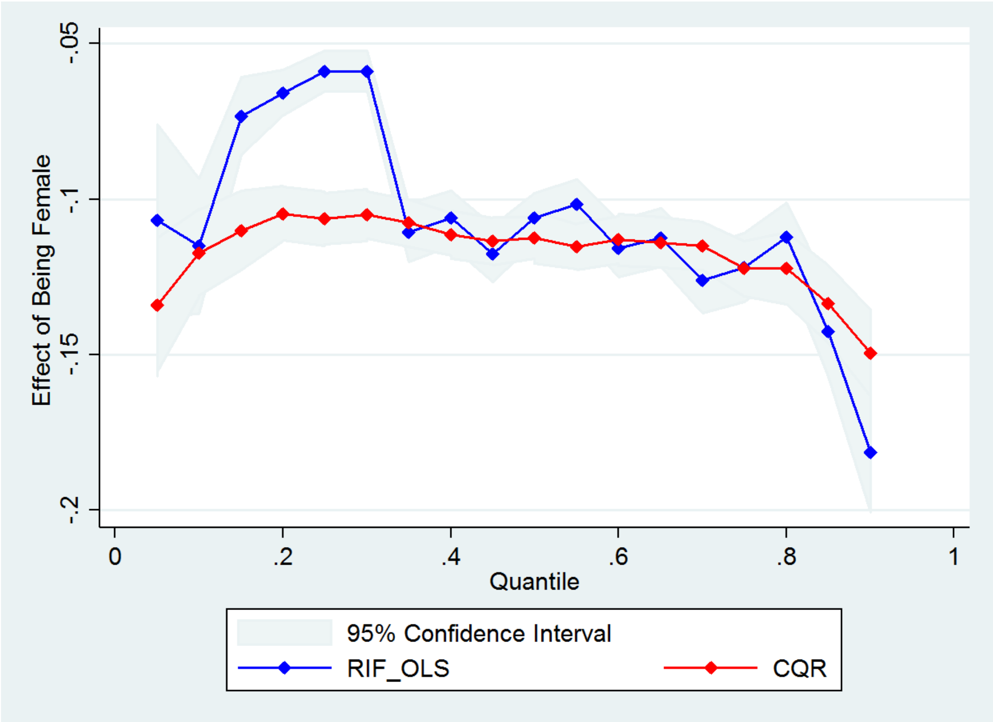


Figure 4.1 UQR and CQR Estimates of the Effect of Women on Log Hourly Wages

### 4.4.3 RIF-OLS Decomposition along the Wage Distribution without Selection Adjustment

Table 4.4 shows the decomposition outcome at specific quantiles. Women are found to have higher observable educational characteristics than men. The difference between men and women is highest at the top of the wage distribution. In contrast, male employees have higher experience and job tenure. Again, the difference is highest at the 90$^{th}$ percentile. Differences in job characteristics as well as in occupations and industries are insignificant at the bottom but negative at the median and top of the wage distribution. The endowments effect of socio-demographic background characteristics reduces the GPG slightly all along the wage distribution. Hence, employed women generally are more often located in the North or Centre of Italy, come from families with higher educational background and are more often married compared to men. The total explained part is positive for low-income earners but negative for median- and top-income earners. However, differences in observable labor market characteristics between men and women statistically significantly reduce only the 90$^{th}$ percentile GPG. In terms of the coefficients effect, educational differences between men and women are insignificant at the bottom, negative at the median and positive at the top of the wage distribution. Gender differences in coefficients to experience and job tenure are positive throughout the wage distribution. Different remuneration to job characteristics between men and women significantly raises the GPG only at the 90$^{th}$ percentile. Gender-specific distributional differences in specific occupations or industries have a statistically significant impact on the coefficients effect all along the wage distribution. Occupational and industrial gender differences in coefficients are negative at the bottom but positive at the median and top of the wage distribution. An intuition of the negative coefficient effect due to distributional differences in occupations and industries between men and women at the 10$^{th}$ percentile GPG is that women are relatively more likely to self-select themselves in low-income jobs and hence to receive the adequate formal education for these jobs (e.g. Brekke and Nyborg, 2010). In contrast, men working in the corresponding sector or occupation at the bottom of the wage distribution have higher probability of not having the adequate formal training compared to their female colleagues. The consequences are negative coefficient effects due to distributional differences in occupations and industries. Negative coefficient estimates due to gender differences in occupations and industries at the bottom of the wage distribution are also found by other studies (e.g. Xiu and Gunderson, 2014). The coefficients effect of socio-demographic background characteristics is generally insignificant all along the income distribution. Consequently, no evidence for gender-based discrimination or favoritism in the labor market based on informal networks is found in this study. The total unexplained part is statistically significant and positive throughout the distribution. In particular, it is a main driver of the GPG at all quantiles considered, while the total explained part is rather small or even working towards a closure of the gap. The coefficient component includes the constant term.[30]

Table 4.5 shows the detailed decomposition results of the different wage inequality measures (90-10, 90-50 and 50-10, respectively). By looking at the different components of the inequality measures in terms of the endowments effect, gender differences in educational attainment is found to reduce wage inequality between the top and bottom or median of the wage distribution. Statistically significant and positive endowments effects of experience and tenure are found for the top-bottom and top-median wage gaps. Hence, in terms of the explained component gender differences in labor market experience and job tenure increase the 90-10 and 90-50 wage gaps, respectively. Job characteristics as well as occupational and sectoral differences reduce the difference in the GPG across the wage distribution. Differences by gender in socio-demographic characteristics have a relevant but small impact on glass ceiling. All in all, differences in the explained component decrease the difference

[30]At the bottom and median there is a premia for simply being male. Contrary, on the top there is a premia for being female.

between the GPGs at the top and bottom or median of the wage distribution. In line with the relatively small 50-10 wage gap, gender differences between the median and bottom of the earnings distribution are found to be rather negligible. By looking at the unexplained component, positive coefficients of education are found to be a main driver of the gender pay disparity between high- and low- or median-income earners. Positive gender differences in returns to experience are found between the 90-10 and 90-50 wage gaps. Similarly, gender differences in job characteristics contribute statistically significantly to the difference between the GPG at the 90$^{th}$ and 10$^{th}$ or 50$^{th}$ percentile. Gender differences in coefficients due to job- and industry-sorting are another driver of gender wage inequality in the Italian private sector. In contrast, the coefficients effect due to changes in differences in socio-demographic characteristics between men and women are found to have no statistically significant impact on the 90-10, 90-50 and 50-10 pay gaps. The total unexplained component is an important driver between the pay gaps at the top and bottom or median of the wage distribution.[31]

In Appendix 4.B, the reweighted decomposition approach is outlined and the decomposition outcome with reweighting for both the quantile-specific GPG and the gender wage inequality measures is shown (Tables 4.B.1–4.B.2). The total reweighting error, $(\bar{X}_M - \bar{X}_M)\hat{\beta}_{\tau}^{rew}$, corresponds to the difference between the *Total Explained* across the UQ Oaxaca-Blinder decomposition and the reweighted-regression decomposition. The reweighting error reflects the fact that the endowments effect in the decomposition with reweighting is not exactly equal to the standard endowments effect, i.e. without reweighting. This occurs, when the reweighted $\bar{X}^{rew}$ is not exactly equal to $\bar{X}$. Figure 4.2 shows the reweghting error and Figure 4.3 the specification error graphically along the wage distribution. The (total) specification error is the difference between the *Total Unexplained* component from the model without and with reweighting; $\bar{X}_M^{rew}(\hat{\beta}_{M,\tau} - \hat{\beta}_{M,\tau}^{rew})$. The specification error is zero if the base model is truly linear. Both errors are rather small, therefore, we expect the RIF-OLS model without reweighting not to be misspecified.

[31]As stated before, the wage structure component contains the constant term. Differences in the constant term decrease wage inequality from the top to the bottom.

Table 4.4 RIF-OLS Detailed Decomposition at Different Quantiles

| | (1) | (2) | (3) |
|---|---|---|---|
| | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG (Unadjusted Gap) | 0.117*** | 0.100*** | 0.179*** |
| | (0.010) | (0.004) | (0.009) |
| *Endowments Effect* | | | |
| Education | -0.013*** | -0.013*** | -0.029*** |
| | (0.002) | (0.001) | (0.003) |
| Experience | 0.024*** | 0.026*** | 0.066*** |
| | (0.003) | (0.002) | (0.004) |
| Job Characteristics | 0.005 | -0.004* | -0.030*** |
| | (0.006) | (0.002) | (0.006) |
| Occupations and Industries | 0.000 | -0.001 | -0.025*** |
| | (0.004) | (0.002) | (0.006) |
| Socio-Demographic Background | -0.008*** | -0.009*** | -0.013*** |
| | (0.002) | (0.001) | (0.002) |
| Total Explained | 0.009 | -0.001 | -0.031*** |
| | (0.008) | (0.004) | (0.009) |
| *Coefficients Effect* | | | |
| Education | -0.040 | -0.027*** | 0.102*** |
| | (0.026) | (0.010) | (0.023) |
| Experience | 0.012 | 0.025** | 0.123*** |
| | (0.026) | (0.010) | (0.024) |
| Job Characteristics | -0.009 | 0.009 | 0.114** |
| | (0.052) | (0.020) | (0.048) |
| Occupations and Industries | -0.288*** | 0.052** | 0.223*** |
| | (0.070) | (0.026) | (0.062) |
| Socio-Demographic Background | -0.012 | -0.030 | -0.059 |
| | (0.111) | (0.043) | (0.110) |
| Total Unexplained | 0.108*** | 0.100*** | 0.210*** |
| | (0.012) | (0.005) | (0.012) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 4.5 Gender Wage Inequality – RIF-OLS Decomposition Results

| | (1) | (2) | (3) |
|---|---|---|---|
| | 90-10 | 90-50 | 50-10 |
| Unadjusted Change | 0.062*** | 0.079*** | -0.018** |
| | (0.011) | (0.009) | (0.008) |
| *Endowments Effect* | | | |
| Education | -0.016*** | -0.011*** | -0.001 |
| | (0.003) | (0.003) | (0.002) |
| Experience | 0.041** | 0.039*** | 0.002 |
| | (0.004) | (0.005) | (0.003) |
| Job Characteristics | -0.035*** | -0.026*** | -0.009* |
| | (0.007) | (0.006) | (0.005) |
| Occupations and Industries | -0.026*** | -0.025*** | -0.001 |
| | (0.006) | (0.005) | (0.004) |
| Socio-Demographic Background | -0.004** | -0.004* | -0.001 |
| | (0.002) | (0.002) | (0.002) |
| Total Explained | -0.040*** | -0.030*** | -0.010 |
| | (0.009) | (0.008) | (0.007) |
| *Coefficients Effect* | | | |
| Education | 0.142*** | 0.128*** | 0.014 |
| | (0.027) | (0.023) | (0.020) |
| Experience | 0.111*** | 0.098*** | 0.013 |
| | (0.027) | (0.023) | (0.020) |
| Job Characteristics | 0.123** | 0.106** | 0.017 |
| | (0.056) | (0.048) | (0.041) |
| Occupations and Industries | 0.511*** | 0.172*** | 0.340*** |
| | (0.071) | (0.061) | (0.052) |
| Socio-Demographic Background | -0.047 | -0.029 | -0.018 |
| | (0.127) | (0.109) | (0.092) |
| Total Unexplained | 0.102*** | 0.110*** | -0.008 |
| | (0.013) | (0.011) | (0.010) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

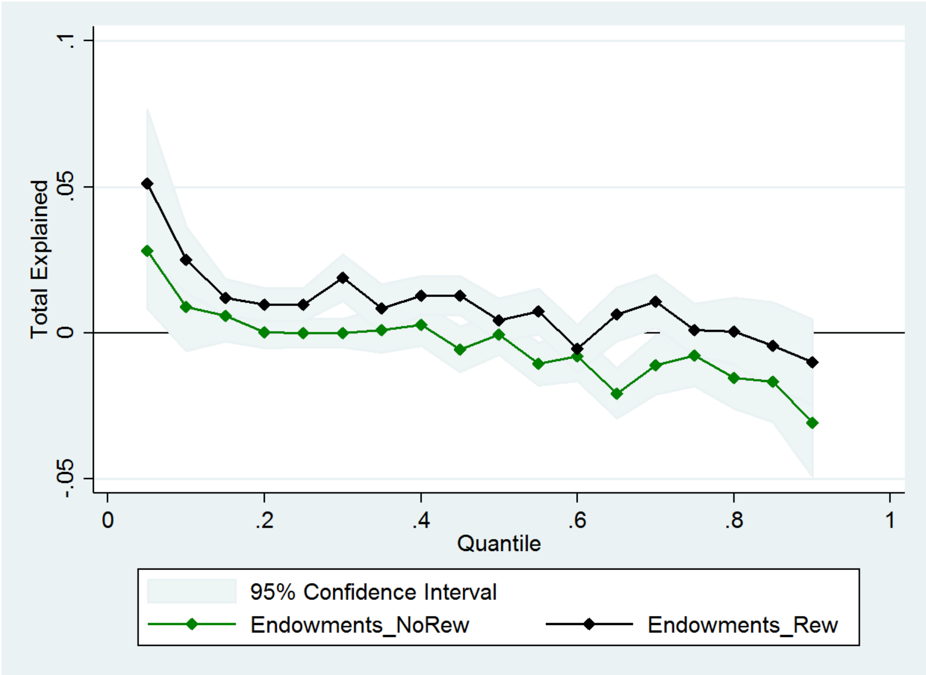


Figure 4.2 Endowments Effect with and without Reweigthing

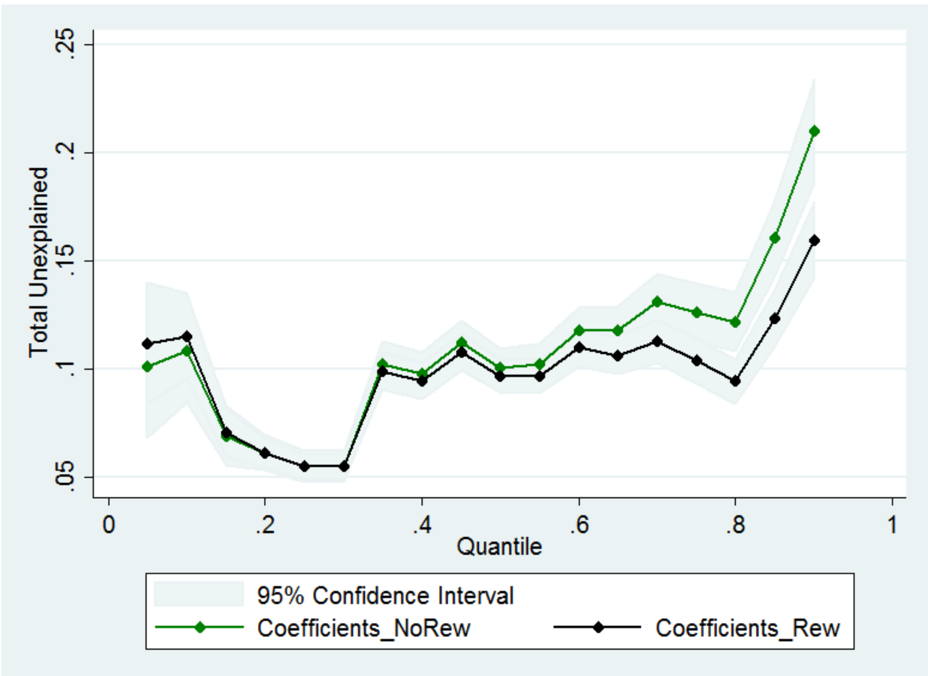


Figure 4.3 Coefficients Effect with and without Reweighting

### 4.4.4 Estimation of the Incidence of Employment

Table 4.6 shows the estimation results of the single-index models (probit, Ichimura and Klein-Spady, respectively). The set of regressors in the selection equations, $Z$, contains at least one variable not included in $X$.[32] The following variables are included in the selection equation only: *Age*, *Age5064*, *Partner_Works* as well as *Kids* and *Kids_10*. These regressors are excluded from the wage equation as they should not influence the wage level directly but reservation earnings. The controls for having children, *Kids*, or young children, *Kids_10*, are used to identify the employment decision.[33] The variables are assumed to affect individual propensity to be employed but not the level of (log) hourly wages. The logic behind is that women with children and in particular young children are less likely to accept wage offers due to child-rearing. In the empirical literature, most studies on the relationship between fertility and female labor market participation find a negative correlation among child-care and female labor force participation (Martins, 2001; Mulligan and Rubinstein, 2008; Lee, 2009; Chang, 2011). Contrary to mothers, fathers have higher employment probability. This is due to the persistence of the male-breadwinner and mother-caretaker model in particular in Southern European countries like Italy (Mínguez, 2004). The dummy variable *Age5064* is equal to one if the individual's age is between 50 and 64 years and zero otherwise. Thus, *Age5064* is a proxy for the last career stage. After child-care and -rearing, the employment probability may increase especially for women. The variable *Partner_Works* is equal to one if the spouse or the partner of the individual is employed and zero otherwise. Devereux (2004) and Bar et al. (2015) find a strong relationship between spousal income and individual labor market participation or employment decision. Therefore, omitting a control for the spouse's or partner's labor market status from the selection equation would potentially lead to inconsistent estimates of the wage equation.

The results in Table 4.6 suggest, on the one hand, that with increasing age, women are more likely to be employed. This may be driven by career breaks due to child-care at earlier career stages. On the other hand, men's incidence of employment decreases slightly with increasing age. Yet, at the final stage of their career both men and women are more likely to accept wage offers. Higher education raises the probability to work for both men and women. Individuals living and working in the North or Centre of Italy have higher probability to be in employment. The positive impact on the probability is highest for employees in Northern

[32] The set of regressors $Z$ for the employment decision is the same in each binary choice estimation, i.e. in the probit, Ichimura and Klein-Spady model, respectively.

[33] For example Chzhen and Mumford (2011) assume that the age of children in the household does not affect the wage level and use it (inter alia) to identify selection of women in full-time employment in Great Britain.

Italy. Married women are less likely to be in employment, while married men are more likely to be employed.[34] Holding the Italian citizenship, if significant, has a positive effect on the incidence of employment for female as well as for male employees. Owning a house significantly raises employees' incidence of employment.[35] The control for owning a house includes bank-financed houses. Consequently, individuals paying-off mortgages are more likely to accept a wage offer. For other house owners, the variable *Homeowner*, as a proxy for wealth, controls for wealthier individuals having better networks and hence are more likely to receive job offers. This increases in turn their likelihood to accept a job offer. Having a partner or a spouse in employment significantly increases the employment probability for men in all three models. For women, the effect is negative in the semiparametric models and positive in the probit estimation. Having children reduces as expected the employment probability of women, while it raises the employment probability for men. Having young children is statistically significant and negative for women, while it impacts positively but statistically insignificantly on the probability of accepting a wage offer for men.[36] The coefficient estimates from the semiparametric single-index models, are comparable to each other in terms of magnitude. The coefficient estimates of the probit model are relatively higher compared to the semiparametric binary choice models in absolute terms. Yet, the signs of the coefficient estimates point generally in the same directions in all three models. The difference in magnitude in the point estimates in the probit estimation compared to the outcome from the semiparametric specifications is in line with results obtained by Buchinsky (1998) or Albrecht et al. (2009) and Chzhen and Mumford (2011).

In order to check whether running the computationally cumbersome semiparametric methods is worth it, in Section 4.4.4, the estimation results of the semiparametric selection models are compared with the regression outcome from the parametric selection model using a two-point wild-bootstrap test based on the idea in Horowitz and Härdle (1994).

[34] In the semiparametric models, no significant effect of being married on the employment probability for men is found.

[35] Except for women in the Klein-Spady model, where owning a house has a negative effect on females' employment probability.

[36] In the semiparametric binary choice models, the effect of having young children on the employment probability for men is statistically significant and negative. Yet, the total effect of having children (*Kids* and *Kids_10*) is positive.

Table 4.6 Estimation Outcome Incidence of Employment

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | Women | | | Men | |
| | Probit Employment | Ichimura Employment | Klein-Spady Employment | Probit Employment | Ichimura Employment | Klein-Spady Employment |
| Constant | -1.335*** | -1.335 | -1.335 | -1.002*** | -1.002 | -1.002 |
| | (0.092) | (0.092) | (0.092) | (0.109) | (0.109) | (0.109) |
| Age | 0.003** | 0.003 | 0.003 | -0.006*** | -0.006 | -0.006 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Age5064 | 0.598*** | 0.334*** | 0.320*** | 0.259*** | 0.035** | 0.025*** |
| | (0.038) | (0.010) | (0.015) | (0.043) | (0.017) | (0.003) |
| Elementary_School | -0.070 | 0.023*** | -0.024 | -0.354*** | -0.021 | -0.018*** |
| | (0.059) | (0.009) | (0.023) | (0.063) | (0.026) | (0.006) |
| High_School | 0.410*** | 0.153*** | 0.141*** | 0.189*** | 0.007 | 0.001 |
| | (0.023) | (0.002) | (0.007) | (0.022) | (0.005) | (0.002) |
| University_Degree | 0.410*** | 0.166*** | 0.208*** | 0.088*** | -0.003 | -0.014*** |
| | (0.026) | (0.003) | (0.010) | (0.031) | (0.006) | (0.003) |
| North | 0.961*** | 0.156*** | 0.146*** | 0.888*** | 0.022*** | 0.011*** |
| | (0.019) | (0.002) | (0.007) | (0.020) | (0.005) | (0.002) |
| Centre | 0.641*** | 0.151*** | 0.100*** | 0.584*** | 0.001 | 0.003 |
| | (0.021) | (0.003) | (0.008) | (0.023) | (0.006) | (0.003) |
| Married | -0.036 | -0.132*** | -0.134*** | 0.489*** | -0.004 | -0.001 |
| | (0.033) | (0.004) | (0.010) | (0.038) | (0.007) | (0.004) |
| Italian | 0.338*** | -0.006 | 0.107*** | 0.406*** | 0.001 | 0.001 |
| | (0.070) | (0.007) | (0.011) | (0.106) | (0.037) | (0.004) |
| Homeowner | 0.046** | 0.010*** | -0.031*** | 0.213*** | 0.022*** | 0.007** |
| | (0.019) | (0.003) | (0.008) | (0.022) | (0.006) | (0.003) |
| Partner_Works | 0.051* | -0.010*** | -0.031*** | 0.167*** | 0.012* | 0.010*** |
| | (0.026) | (0.003) | (0.009) | (0.030) | (0.006) | (0.003) |
| Kids | -0.220*** | -0.164*** | -0.096*** | 0.162*** | 0.049*** | 0.018*** |
| | (0.029) | (0.005) | (0.010) | (0.029) | (0.007) | (0.004) |
| Kids_10 | -0.081*** | -0.003 | -0.037*** | -0.012 | -0.027*** | -0.016*** |
| | (0.027) | (0.004) | (0.010) | (0.039) | (0.006) | (0.004) |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 30,283 | 30,283 | 30,283 | 22,406 | 22,406 | 22,406 |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes:* The constant and the *Age* coefficients in the semiparametric binary choice models have been normalized to the corresponding values from the parametric probit model.

### Testing for Equality of the Parametric and Semiparametric Binary Choice Model

The results obtained from the different selection models are compared using a modified version of the Horowitz and Härdle (1994) test proposed in Henderson and Parmeter (2015). The test compares the parametric with the semiparametric choice model of the employment decision $E$ on $Z\gamma$. The null hypothesis tests whether the parametric model is the correct specification. In the underlying case, the parametric form is the probit model and hence $H_0 = E(Y|X) = F(Z\gamma)$, where $F(\cdot)$ is the standard normal cumulative distribution function. The alternative hypothesis is $H_1 = E(Y|X) = H(Z\gamma)$, where $H(\cdot)$ is the unknown smooth function. The feasible test statistics looks as follows:

$$T_{HH} = \{E - F(Z\hat{\gamma})\}\{H(Z\hat{\gamma}) - F(Z\hat{\gamma})\} \tag{4.11}$$

The intuition behind the test is the following: given that $H(\cdot) = F(\cdot)$ holds, the parametric model is specified correctly and, therefore, should not differ from the semiparametric estimate of the function. In this case, cumbersome computation of the semiparametric models is not necessary. Horowitz and Härdle (1994) pre-multiply the right-hand side of equation (4.11) by a non-negative weighting function that punishes extreme observations. Yet, the test is sensitive to the choice of the weighting function (Proenca, 1993). In the underlying analysis bootstrapping is used what makes the weighting unnecessary (Proenca, 1993; Henderson and Parmeter, 2015). A two-point wild bootstrap in order to calculate the upper-tail bootstrap $p$-value is used. The $p$-values in Table 4.7 reject the parametric model at a 10.0% significance level in all cases. In comparison with the Ichimura estimation, the probit model is even rejected at a 1.0% significance level for both men and women.

Table 4.7 Results of the Horowitz-Härdle Test

| | (1) | (2) |
|---|---|---|
| | p-Value | |
| | Female Sample | Male Sample |
| Probit – Ichimura | 0.002 | 0.002 |
| Probit – Klein and Spady | 0.067 | 0.006 |

Following Martins (2001), Figure 4.4 represents the parametric and semiparametric fit for men and women, Figure 4.5 shows the respective fits for the full sample. The Figures show that the probit specification does not capture the behavior of individuals with low index numbers very well. This is particularly pronounced for women. Hence, the semiparametric models provide more information on the selection behavior of the individuals in the sample.

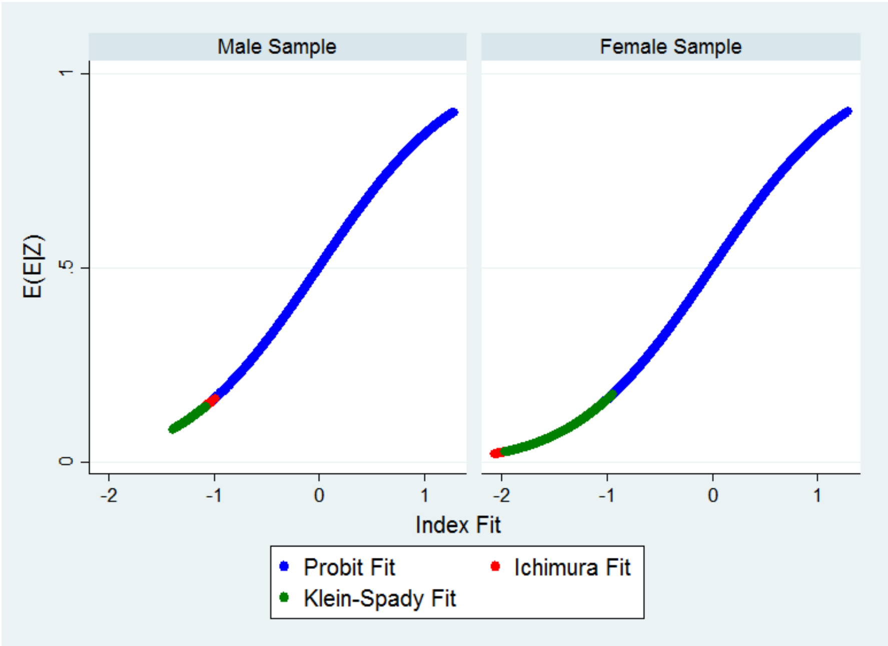


Figure 4.4 Probit and Semiparametric Fit for the Estimated Index by Gender

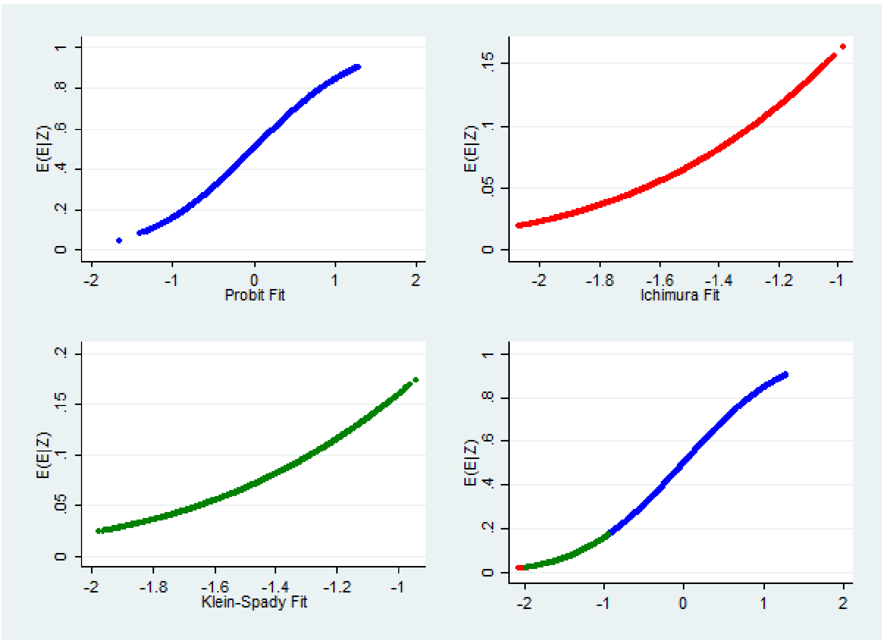


Figure 4.5 Probit and Semiparametric Fit for the Estimated Index

### 4.4.5 RIF-OLS Decomposition along the Wage Distribution with Selection Adjustment

In this Section, we estimate the selectivity-corrected wage model using second-order polynomials.[37] Tables 4.8–4.9 show the detailed decomposition outcome at specific quantiles when it is accounted for sample selectivity. Table 4.10 summarizes the main result; the unexplained component of the GPG being the main driver of the pay disparity at the mean as well as along the wage distribution changes in the case of sample-selection adjustment. The part generally attributed to discrimination is reduced at the bottom but increased at the top of the wage distribution. This implies that without selection correction, we overestimate the part attributed to gender-wage discrimination for low-income earners and underestimate it for high-income earners.

In fact, gender differences in unobservables are main drivers of the GPG at the $10^{\text{th}}$ and $50^{\text{th}}$ percentile. Not accounting for sample selection would therefore significantly underestimate the total explained part at the $10^{\text{th}}$ and $50^{\text{th}}$ percentile. In terms of the coefficients effect, the results suggest that women are paid more from the same set of generally unobservable characteristics at the median and top of the wage distribution. At the bottom, the sign of the selection component in terms of the coefficients effect works in the opposite direction: men gain more than women from the same set of unobservable characteristics. Without selection adjustment, the unexplained component is underestimated at the top but overestimated at the median and bottom. All in all, the selection component is one of the most important components explaining gender differences in pay along the earnings distribution. Hence, otherwise unobservable characteristics and individual heterogeneity contribute significantly to the quantile-specific GPGs. However, the effect differs both in sign and magnitude at the distinct points of the wage distribution.

Table 4.11 shows that the selection component also significantly contributes to the variation of the GPG across the earnings distribution. Between the top and bottom, differences in the selection correction term increase wage inequality. In contrast, gender differences in unobservable characteristics reduce wage inequality between the top and median and the median and the bottom of the wage distribution. Different coefficients of unobservables between men and women decreases wage inequality all along the wage distribution. This result is driven by higher prices for women given the same set of unobservable characteristics between men and women at the top of the wage distribution.

[37] The polynomials are not orthogonal. However, the selection terms used are not collinear; $Corr(\lambda, \lambda^2) < |0.5|$.

In Appendix 4.D, the contribution of the selection component to the GPG at different quantiles as well as to the change in the GPG across the wage distribution is presented for the model with parametric selection correction (Tables 4.D.1 –4.D.2, respectively).

Table 4.8 RIF-OLS Detailed Decomposition at Different Quantiles with Selection – Ichimura

| | (1) | (2) | (3) |
|---|---|---|---|
| | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG (Unadjusted Gap) | 0.117*** | 0.100*** | 0.179*** |
| | (0.010) | (0.004) | (0.009) |
| *Endowments Effect* | | | |
| Education | -0.011*** | -0.013*** | -0.029*** |
| | (0.002) | (0.001) | (0.003) |
| Experience | 0.020*** | 0.026*** | 0.068*** |
| | (0.004) | (0.002) | (0.005) |
| Job Characteristics | 0.006 | -0.004* | -0.030*** |
| | (0.006) | (0.002) | (0.006) |
| Occupations and Industries | 0.001 | -0.000 | -0.026*** |
| | (0.004) | (0.002) | (0.006) |
| Socio-Demographic Background | -0.011*** | -0.009*** | -0.012*** |
| | (0.002) | (0.001) | (0.002) |
| Selection | 0.050*** | 0.013*** | -0.013 |
| | (0.011) | (0.004) | (0.013) |
| | | | |
| Total Explained | 0.055*** | 0.012** | -0.042*** |
| | (0.013) | (0.006) | (0.015) |
| | | | |
| *Coefficients Effect* | | | |
| Education | -0.047* | -0.032*** | 0.080*** |
| | (0.028) | (0.010) | (0.024) |
| Experience | -0.010 | 0.024** | 0.142*** |
| | (0.028) | (0.010) | (0.026) |
| Job Characteristics | -0.011 | 0.008 | 0.118** |
| | (0.052) | (0.020) | (0.048) |
| Occupations and Industries | -0.290*** | 0.052** | 0.225*** |
| | (0.070) | (0.026) | (0.062) |
| Socio-Demographic Background | -0.031 | -0.038 | -0.060 |
| | (0.112) | (0.043) | (0.110) |
| Selection | 0.599 | -0.114 | -0.971** |
| | (0.423) | (0.165) | (0.429) |
| | | | |
| Total Unexplained | 0.062*** | 0.088*** | 0.221*** |
| | (0.016) | (0.006) | (0.017) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 4.9 RIF-OLS Detailed Decomposition at Different Quantiles with Selection – Klein-Spady

| | (1) | (2) | (3) |
|---|---|---|---|
| | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG (Unadjusted Gap) | 0.117*** | 0.100*** | 0.179*** |
| | (0.010) | (0.004) | (0.009) |
| *Endowments Effect* | | | |
| Education | -0.011*** | -0.013*** | -0.029*** |
| | (0.002) | (0.001) | (0.003) |
| Experience | 0.020*** | 0.025*** | 0.065*** |
| | (0.004) | (0.002) | (0.005) |
| Job Characteristics | 0.006 | -0.004* | -0.030*** |
| | (0.006) | (0.002) | (0.006) |
| Occupations and Industries | 0.001 | -0.000 | -0.025*** |
| | (0.004) | (0.002) | (0.006) |
| Socio-Demographic Background | -0.011*** | -0.010*** | -0.012*** |
| | (0.002) | (0.001) | (0.002) |
| Selection | 0.079*** | 0.019** | -0.006 |
| | (0.023) | (0.010) | (0.027) |
| | | | |
| Total Explained | 0.084*** | 0.018* | -0.037 |
| | (0.023) | (0.010) | (0.026) |
| | | | |
| *Coefficients Effect* | | | |
| Education | -0.054* | -0.033*** | 0.079*** |
| | (0.028) | (0.010) | (0.024) |
| Experience | -0.006 | 0.022** | 0.130*** |
| | (0.028) | (0.011) | (0.026) |
| Job Characteristics | -0.010 | 0.009 | 0.118** |
| | (0.052) | (0.020) | (0.048) |
| Occupations and Industries | -0.289*** | 0.052** | 0.225*** |
| | (0.070) | (0.026) | (0.062) |
| Socio-Demographic Background | -0.045 | -0.042 | -0.071 |
| | (0.112) | (0.043) | (0.110) |
| Selection | 0.275 | -0.040 | -0.512 |
| | (0.419) | (0.162) | (0.415) |
| | | | |
| Total Unexplained | 0.033 | 0.082*** | 0.216*** |
| | (0.025) | (0.010) | (0.028) |
| | 24,267 | 24,267 | 24,267 |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 4.10 GPG and Total Unexplained Component with and without Selection

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Mean | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG | 0.118*** | 0.117*** | 0.100*** | 0.179*** |
| (Unadjusted) | (0.005) | (0.010) | (0.004) | (0.009) |
| Total Unexplained | 0.124*** | 0.108*** | 0.100*** | 0.210*** |
| (No Selection) | (0.007) | (0.012) | (0.005) | (0.012) |
| Total Unexplained | 0.123*** | | | |
| (Selection – Standard Heckman Two-Stage) | (0.006) | | | |
| Total Unexplained | | 0.105*** | 0.100*** | 0.211*** |
| (Selection – Probit) | | (0.012) | (0.005) | (0.012) |
| Total Unexplained | | 0.062*** | 0.088*** | 0.221*** |
| (Selection – Ichimura) | | (0.016) | (0.006) | (0.017) |
| Total Unexplained | | 0.033 | 0.082*** | 0.216*** |
| (Selection – Klein-Spady) | | (0.025) | (0.010) | (0.028) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes:* For the mean, the difference in the *Total Unexplained* component with no selection and with selection adjustment is statistically significant at a 5.0% significance level. In the case of the standard probit model, the difference is not statistically significant at all quantiles. For the semiparametric selection models, the difference is statistically significant at a 5.0% significance level only for the $10^{th}$ percentile. At the $50^{th}$ percentile, the difference is statistically significant at a 10.0% significance level for the component with no selection and the Klein-Spady selection-adjusted component. At the $90^{th}$ percentile, the difference is not statistically different from zero in both cases. The difference between the respective components has been tested using a two-sample t-test.

Table 4.11 Gender Wage Inequality – Selection Component

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | 90-10 | | 90-50 | | 50-10 | |
| | Ichimura | Klein-Spady | Ichimura | Klein-Spady | Ichimura | Klein-Spady |
| Unadjusted Change | 0.062*** | | 0.079*** | | -0.018** | |
| | (0.011) | | (0.009) | | (0.008) | |
| *Endowments Effect* | | | | | | |
| Selection | 0.072*** | 0.034*** | -0.026* | -0.037*** | -0.025 | -0.060** |
| | (0.012) | (0.006) | (0.013) | (0.012) | (0.028) | (0.025) |
| *Coefficients Effect* | | | | | | |
| Selection | -0.311*** | -0.138* | -0.857* | -0.314 | -0.472 | -0.012 |
| | (0.103) | (0.112) | (0.460) | (0.454) | (0.445) | (0.449) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

## 4.5 Conclusion

This paper analyzes the GPG at different points of as well as gender wage inequality across the wage distribution. The empirical application is based on UQR or the RIF-OLS model. This approach allows to decompose the wage equations by gender in detail along the earnings distribution using a Oaxaca-Blinder type decomposition. The method delivers detailed information on the drivers of the pay gap between men and women at specific quantiles. Gender wage inequality in the sample is estimated by the change in the GPG across the wage distribution, i.e. the 90-10, 90-50 and 50-10 wage gap. The estimation is based on linear RIF regressions, as potential misspecification problems of the RIF-OLS are found to be negligible. The method based on UQRs has several advantages compared to CQR models such as its intuitive and computationally easy estimation as well as interpretation. However, CQRs are the standard approach in the quantile-regression literature (Fortin et al., 2011). So far, sample selection correction beyond the mean is only conducted in the CQR framework. Studies controlling for selection effects find that the work decision impacts differently along the earnings distribution. Therefore, in this paper, the RIF-OLS model is extended in order to allow for sample selection. By adding selection terms as second-order polynomials to the earnings equation, the estimation results are adjusted for potential non-random selection into employment (Buchinsky, 1998). The selection correction focuses on semiparametric models as the selection process may be non-normally distributed (Martins, 2001). Indeed, a two-point wild-bootstrap test, based on Horowitz and Härdle (1994) and comparing the parametric and semiparametric binary choice models, rejects the parametric probit specification.

The analysis in this paper shows that different factors, such as educational attainment, labor market presence, job characteristics, employment in different industries or demographic and family background characteristics contribute differently to the GPG along the wage distribution. In particular, by splitting the various categories in an endowments and a coefficients part, differences in the contribution to the GPG at different quantiles are found. Individual heterogeneity, like individual ability or personal motivation, and other unobservable labor market characteristics (as for example differences in educational quality) contribute statistically significantly to pay differences between men and women along the wage distribution. Moreover, we detect glass ceiling, i.e. significant differences in the GPGs between upper and lower quantiles. In line with this, the wage penalty of being female is highest at the top. Wage structure effects of male-female differences in educational attainment are a main driver of wage inequality between the top and bottom or median quantile, while the endowments effect of gender differences in education significantly lowers wage inequality. Endowments effects of the set of regressors accounting for gender differences in labor market presence across the wage distribution are relevant in contributing to wage inequality as well as a positive GPG at all quantiles. The bottom of the wage distribution is relatively more equal in terms of job characteristics and industrial and occupational differences between men and women in terms of endowments. Differences in demographic and family background characteristics between men and women across the wage distribution both in terms of endowments and coefficients effects are less important.

Overall, the results suggest that it is important to consider GPGs throughout the wage distribution and hence to go beyond the mean. This may be particularly relevant, when it comes to policy implications as policy makers want to address potential inequality or distributional effects. Most of the quantile-specific pay gaps is accounted for by how men and women are rewarded, i.e. by the unexplained component. This finding is in conformity with results obtained in other studies on gender differences in pay (see for example Blau and Kahn, 2016). In contrast, net differences in endowments, i.e. the total explained part, work towards a reduction of the phenomenon of glass ceiling as well as of the GPGs at the corresponding quantiles. According to which selection adjustment model (parametric or semiparametric) is chosen, the correction terms contribute differently to the quantile-specific GPGs. Yet, in all model specifications, the main pattern of results remains the same. The unexplained part is overestimated at the bottom and median but underestimated at the top of the wage distribution. The extension proposed, to the author's best knowledge, is the first approach allowing to control for selection issues when conducting detailed Oaxaca-Blinder type decompositions based on UQRs along the wage distribution.

# Appendices

## Appendix 4.A Oaxaca-Blinder Decomposition

The standard approach in decomposing wage differences between groups is the Oaxaca (1973) and Blinder (1973) decomposition approach outlined in the following:

$$\begin{aligned}\bar{Y}_M - \bar{Y}_F &= \bar{X}_M \hat{\beta}_M - \bar{X}_F \hat{\beta}_F \\ &= (\bar{X}_M - \bar{X}_F)\hat{\beta}_M + \bar{X}_F(\hat{\beta}_M - \hat{\beta}_F)\end{aligned} \tag{4.A.1}$$

where $\bar{Y}_M$ and $\bar{Y}_F$ are the log hourly wages for the male and female sample evaluated at the mean, respectively, with $\bar{X}_G$ being a $1 \times K$ vector of average characteristics and $\hat{\beta}_G$ being a $K \times 1$ vector of estimated coefficients for $G = (F, M)$, with $M = Male$ and $F = Female$. Define $\bar{Y}_M - \bar{Y}_F = \hat{\Delta}$ and $(\bar{X}_M - \bar{X}_F)\hat{\beta}_M = \hat{\Delta}_E$ as well as $\bar{X}_F(\hat{\beta}_M - \hat{\beta}_F) = \hat{\Delta}_C$, with $E$ identifying the *Endowments Effect* and $C$ the *Coefficients Effect*. The endowments effect, $\hat{\Delta}_E$, evaluates the GPG in terms of differences in observable characteristics given male prices. The standard case that is applied here uses male coefficients, $\hat{\beta}_M$, as the non-discriminatory wage structure and hence assumes no discrimination against men. The second term, $\hat{\Delta}_C$, i.e the coefficients part or the adjusted GPG, evaluates the pay gap in terms of different returns for female characteristics.

Table 4.A.1 shows the result from the standard Oaxaca-Blinder decomposition at the mean without and with sample selection correction (column (1) and (2), respectively). For selection correction, the standard Heckman (1979) two-step procedure is applied. The selection component is significant only in terms of the endowments effect and adjusts both the total explained and total unexplained part only slightly. In absolute terms both parts are corrected downwards.

Table 4.A.1 Oaxaca-Blinder Decomposition at the Mean without and with Selection

| | (1) | (2) |
|---|---|---|
| | No Selection | Selection[a] |
| GPG (Unadjusted Gap) | 0.118*** | |
| | (0.005) | |
| *Endowments Effect* | | |
| Education | -0.016*** | -0.015*** |
| | (0.001) | (0.001) |
| Experience_Tenure | 0.035*** | 0.032*** |
| | (0.002) | (0.002) |
| Job_Char | -0.010*** | -0.009*** |
| | (0.003) | (0.003) |
| Occupations_Industry | -0.007*** | -0.007*** |
| | (0.002) | (0.002) |
| Socio-Demographic_Background | -0.009*** | -0.009*** |
| | (0.001) | (0.001) |
| Selection | | 0.003** |
| | | (0.001) |
| Total Explained | -0.007 | -0.006 |
| | (0.004) | (0.004) |
| *Coefficients Effect* | | |
| Education | 0.004 | 0.005 |
| | (0.011) | (0.013) |
| Experience_Tenure | 0.041*** | 0.039*** |
| | (0.012) | (0.013) |
| Job_Char | 0.016 | 0.016 |
| | (0.023) | (0.023) |
| Occupations_Industry | -0.004 | -0.006 |
| | (0.031) | (0.031) |
| Socio-Demographic_Background | -0.029 | -0.032 |
| | (0.052) | (0.053) |
| Selection | | 0.010 |
| | | (0.040) |
| Total Unexplained | 0.124*** | 0.123*** |
| | (0.006) | (0.006) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

[a] The selection components are estimated via the classical Heckman two-step correction method (Heckman, 1979).

# Appendix 4.B Reweighted-Regression Decomposition

If the assumed linearity in the RIF model does not hold, the model is misspecified and the decomposition components are incorrect. Adding a reweighting scheme solves this problem. The reweighted-regression decomposition using the reweighting approach proposed by DiNardo et al. (1996) consists in performing two Oaxaca-Blinder decompositions; one for the endowments and one for the coefficients effect.[38] In order to use men as the non-discriminatory wage structure, the sample of men is reweighted to the sample of women, indicated by the index *rew*. The method uses a semiparametric reweighting factor and creates a counterfactual framework by reweighting men's characteristics to be as women's. The marginal and unconditional wage distribution $Y_M|D_M$ is derived via the law of iterated probabilities by integrating the conditional distribution of wages observed for men given the set of observable characteristics $X$, $F_{Y_M|X,D_M}$ over the male marginal distribution of $X$:

$$F_{Y_M|D_M}(Y) = \int F_{Y_M|X,D_M}(Y|X=x)dF_{X|D_M}(x) \tag{4.B.1}$$

$$\begin{aligned} F_{Y_M^{rew}:X=X|D_F}(Y) &= \int F_{Y_M|X,D_M}(Y|X=x)\Psi(X)dF_{X|D_M}(x) \\ &= \int F_{Y_M|X,D_M}(Y|X=x)dF_{X|D_F}(x) \end{aligned} \tag{4.B.2}$$

where the dummy variable $D_G$ with $G = M, F$ identifies group membership, for $M = Male$ and $F = Female$ and $\Psi(X)$ is the reweighting factor. It reweights female observations and is derived using the predicted probability of belonging to the male sample, i.e. being male, given $X$, $Pr(X|D_M = 1)$. The reweighting factor $\Psi(X) = \frac{dF_{X_F}(X)}{dF_{X_M}(X)}$ is the ratio of the marginal distributions of $X$ for women $F$ and men $M$. As $\Psi(X)$ is simply a function of $X$, it can be derived using the predicted probability of being a woman given X, $Pr(X|D_F = 1)$, i.e. via standard probit or logit. Since $dF_{X_F}(X) = Pr(X|D_F = 1)$ and $dF_{X_M}(X) = Pr(X|D_F = 0)$, the reweighting factor can be re-written as:

$$\Psi(X) = \frac{Pr(X|D_F = 1)}{Pr(X|D_F = 0)} = \frac{Pr(D_F = 1|X)Pr(D_F = 0)}{Pr(D_F = 0|X)Pr(D_F = 1)} \tag{4.B.3}$$

[38]The application of other reweighting procedures is possible. For example, propensity score reweighting could be used (Hirano et al., 2003). Here the method proposed by DiNardo et al. (1996) is used as it relies minimally on functional form assumptions. Alternatively, the model proposed by Ghosh (2014) could be used as the reweigthing may have relatively poor finite sample performance.

In order to obtain $Pr(X|D_F = 1)$, a probit regression for the pooled sample is run. In the probit estimation, all regressors used in the decomposition, as well as a set of interaction terms between educational dummies, experience and occupations (as a proxy for tasks) are included. In particular, for the detailed decomposition, besides the reweighting factor described in equation (4.B.3), for each covariate $k$ (with $k = 1, 2, \ldots, K$) a reweighting factor using all covariates except $X_k$ is estimated:

$$\Psi_k(X_{K-k}) = \frac{Pr(X_{K-k}|D_F = 1)}{Pr(X_{K-k}|D_F = 0)} = \frac{Pr(D_F = 1|X_{K-k})Pr(D_F = 0)}{Pr(D_F = 0|X_{K-k})Pr(D_F = 1)} \tag{4.B.4}$$

The counterfactual statistic of each covariate $k$ is obtained by using the product of the reweighting factors (4.B.3) and (4.B.4), $\Psi(X)\Psi_k(X_{K-k})$, as weights (instead of using only $\Psi(X)$ as weight). The counterfactual statistic is then subtracted from $\Psi(X)$ yielding the contribution of each covariate $k$ (Fortin et al., 2011). As the effect on the single covariates is estimated conditional on all other covariates, the method is path independent.

In order to obtain a detailed decomposition in the sense of Oaxaca (1973) and Blinder (1973), in a first stage, the distributional changes are estimated separately for an endowments and a coefficients effect. In a second stage, the two effects are further divided into the contribution of each set of covariates (or each covariate) using the RIF-regression model outlined in Section 4.2.1. The endowments effect is obtained by decomposing the wage gap between the male and the reweighted sample:

$$\begin{aligned} \hat{\Delta}_{E,R} &= \hat{\Delta}_{E,p} + \hat{\Delta}_{E,e} \\ &= \underbrace{(\bar{X}_M - \bar{X}_M^{rew})\hat{\beta}_{M,\tau}}_{\text{Pure Endowments Effect}} + \underbrace{\bar{X}_M^{rew}(\hat{\beta}_{M,\tau} - \hat{\beta}_{M,\tau}^{rew})}_{\text{Specification Error}} \end{aligned} \tag{4.B.5}$$

where $p$ indicates the *pure effect*, $e$ the part attributed to the *error term* and $R$ the total effect when reweighting is conducted. The index $E$ identifies again the *Endowments Effect*. The specification error in the linear model is equal to zero, if the model is truly linear; $\bar{X}_M^{rew}(\hat{\beta}_{M,\tau} - \hat{\beta}_{M,\tau}^{rew}) = 0$. Differences between the detailed reweighted RIF-decomposition and the RIF decomposition without reweighting are caught by the specification error. These differences can be measured as the difference between the coefficients effect from the model without and with reweigthing (specification error). The additional term in the (total) endowments component, the specification error, allows to draw conclusions on the goodness of specification of the linear model (the specification error is zero if the model is truly linear). Hence, it adjusts the endowments component, when the linear model is not accurately specified. In another Oaxaca-Blinder type decomposition, the coefficients part is calculated. The decomposition is conducted between the reweighted sample, *rew*, and the female sample, $F$:

$$\begin{aligned}\hat{\Delta}_{C,R} &= \hat{\Delta}_{C,p} + \hat{\Delta}_{C,e} \\ &= \underbrace{\bar{X}_F(\hat{\beta}^{rew}_{M,\tau} - \hat{\beta}_{F,\tau})}_{\text{Pure Coefficients Effect}} + \underbrace{(\bar{X}^{rew}_M - \bar{X}_F)\hat{\beta}^{rew}_{M,\tau}}_{\text{Reweighting Error}} \\ &\approx \bar{X}_F(\hat{\beta}^{rew}_{M,\tau} - \hat{\beta}_{F,\tau})\end{aligned} \tag{4.B.6}$$

where the index $C$ identifies the *Coefficients Effect*. The reweighting error, $(\bar{X}^{rew}_M - \bar{X}_F)\hat{\beta}^{rew}_{M,\tau}$, goes to zero given that the following property of large samples holds: $\text{plim}(\bar{X}^{rew}_F) = \text{plim}(\bar{X}_M)$ leading to $\hat{\Delta}_{C,e} \to 0$ as $N \to \infty$.[39]

For the quantile-specific reweighted decomposition outcome and the reweighted wage inequality measures shown in Table 4.B.1 and Table 4.B.2, respectively, the pure endowments and coefficients effect are referred to as *Total Explained* or *Total Unexplained*. The application of a reweighting approach may be particularly important when considering RIF regressions as they might not be linear for distributional statistics besides the mean (Fortin et al., 2011). Advantages of the reweighting scheme applied here are the low dependence on functional form assumptions of the (flexible) probit for gender effects and that the procedure yields efficient estimates (Fortin et al., 2011).

[39]Given that the reweighting function has been correctly specified.

Table 4.B.1 RIF-OLS Detailed Decomposition at Different Quantiles with Reweighting

| | (1) | (2) | (3) |
|---|---|---|---|
| | $10^{th}$ Percentile F(X) in male sample reweighted to female sample | $50^{th}$ Percentile F(X) in male sample reweighted to female sample | $90^{th}$ Percentile F(X) in male sample reweighted to female sample |
| GPG (Unadjusted Gap) | 0.117*** | 0.100*** | 0.179*** |
| | (0.010) | (0.004) | (0.009) |
| *Endowments Effect* | | | |
| Education | -0.011*** | -0.019*** | -0.028*** |
| | (0.002) | (0.001) | (0.003) |
| Experience | 0.028*** | 0.038*** | 0.062*** |
| | (0.003) | (0.002) | (0.004) |
| Job Characteristics | 0.013*** | -0.007*** | -0.025*** |
| | (0.004) | (0.002) | (0.005) |
| Occupations and Industries | -0.002 | -0.007*** | -0.019*** |
| | (0.004) | (0.002) | (0.005) |
| Socio-Demographic Background | -0.003*** | -0.000 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| Total Explained | 0.025*** | 0.004 | -0.010 |
| | (0.006) | (0.004) | (0.007) |
| *Coefficients Effect* | | | |
| Education | -0.024 | -0.015 | 0.146*** |
| | (0.028) | (0.011) | (0.025) |
| Experience | -0.070*** | -0.009 | 0.043* |
| | (0.027) | (0.010) | (0.024) |
| Job Characteristics | -0.011 | 0.005 | 0.089* |
| | (0.053) | (0.020) | (0.048) |
| Occupations and Industries | -0.364*** | 0.031 | 0.116* |
| | (0.073) | (0.028) | (0.064) |
| Socio-Demographic Background | 0.175* | -0.059 | -0.212** |
| | (0.100) | (0.038) | (0.090) |
| Total Unexplained | 0.115*** | 0.097*** | 0.160*** |
| | (0.010) | (0.004) | (0.009) |
| Specification Error | -0.007 | 0.003 | 0.05 |
| Reweighting Error | 0.016 | 0.005 | 0.041 |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

*Notes:* The *Total Unexplained* parts from the model without and with reweighting are not statistically significantly different from each other at the $10^{th}$ and $50^{th}$ percentile of the wage distribution but statistically significantly different at the $90^{th}$ percentile. Similarly, the *Total Explained* components with and without reweighting are not statistically significantly different at the bottom and median but statistically significantly different at the top. The difference between the respective parts has been tested using a two-sample t-test.

Table 4.B.2 Gender Wage Inequality – RIF-OLS Decomposition Results with Reweighting

| | (1) | (2) | (3) |
|---|---|---|---|
| | 90-10<br>F(X) in male sample reweighted to female sample | 90-50<br>F(X) in male sample reweighted to female sample | 50-10<br>F(X) in male sample reweighted to female sample |
| Unadjusted Change | 0.062*** | 0.079*** | -0.018** |
| | (0.011) | (0.009) | (0.008) |
| *Endowments Effect* | | | |
| Education | -0.017*** | -0.009*** | -0.008*** |
| | (0.002) | (0.002) | (0.001) |
| Experience | 0.034*** | 0.024*** | 0.010*** |
| | (0.002) | (0.003) | (0.002) |
| Job Characteristics | -0.038*** | -0.018*** | -0.020*** |
| | (0.003) | (0.001) | (0.002) |
| Occupations and Industries | -0.017*** | -0.012*** | -0.006*** |
| | (0.002) | (0.002) | (0.001) |
| Socio-Demographic Background | 0.003 | -0.000 | 0.003 |
| | (0.003) | (0.002) | (0.002) |
| Total Explained | -0.035*** | -0.014*** | -0.021*** |
| | (0.005) | (0.004) | (0.003) |
| *Coefficients Effect* | | | |
| Education | 0.170*** | 0.161*** | 0.009 |
| | (0.037) | (0.027) | (0.030) |
| Experience | 0.113*** | 0.052** | 0.061** |
| | (0.036) | (0.026) | (0.029) |
| Job Characteristics | 0.100 | 0.084 | 0.017 |
| | (0.071) | (0.052) | (0.057) |
| Occupations and Industries | 0.480*** | 0.085 | 0.396*** |
| | (0.097) | (0.070) | (0.078) |
| Socio-Demographic Background | -0.387*** | -0.154 | -0.234** |
| | (0.135) | (0.098) | (0.107) |
| Total Unexplained | 0.045** | 0.063*** | -0.018* |
| | (0.013) | (0.010) | (0.011) |
| Specification Error | 0.057 | 0.047 | 0.01 |
| Reweighting Error | 0.005 | 0.016 | -0.011 |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

# Appendix 4.C Asymptotic Normality of the RIF-OLS Model with Semiparametric Estimators for Selection Correction

We assume that the following model using only the observed data yields biased parameters:

$$RIF(Y;Q_\tau) = X\beta_\tau + u_\tau \tag{4.C.1}$$

as

$$E[RIF(Y;Q_\tau)|X, E=1] = X\beta_\tau + E[u_\tau|v_\tau > -Z\gamma] \tag{4.C.2}$$

with $E[u_\tau|v_\tau > -Z\gamma] \neq 0$. Firpo et al. (2009a) derive the asymptotic properties of the RIF-OLS. In the following, the asymptotic normality as derived in Firpo et al. (2009a) is extended for the model with selection correction. Recall that, in the case of semiparametric estimators for selection correction, the RIF-OLS regression model corrected for selection bias at $\tau$ has the following form:

$$\begin{aligned}\widehat{RIF}(Y;Q_\tau) &= X\hat{\beta}_\tau + \hat{h}^*_\tau(\hat{m}^*) + \hat{\varepsilon}_\tau \\ &= X\hat{\beta}_\tau + \hat{\delta}^*_{1\tau}\lambda^*(Z^*_A\hat{\gamma}^*)^1 + \hat{\delta}^*_{2\tau}\lambda^*(Z^*_A\hat{\gamma}^*)^2 + \hat{\varepsilon}_\tau\end{aligned} \tag{4.C.3}$$

where $X$ is a vector of $K$ regressors, $\hat{\varepsilon}_\tau$ is the quantile-specific residual.[40] $\hat{\beta}_\tau$ is the corresponding vector of coefficient estimates at $\tau$. $\lambda^*(Z^*_A\hat{\gamma}^*)$ is the IMR and $\lambda^*(Z^*_A\hat{\gamma}^*)^2$ is the quadratic term of the IMR, $\hat{\delta}^*_{1\tau}$ and $\hat{\delta}^*_{2\tau}$ are the corresponding coefficient estimates. We assume that the wage model in (4.C.3) yields consistent and unbiased parameter estimates and hence successfully corrects for sample selection. For each observation $i$, we have:

$$\widehat{RIF}(Y_i;Q_\tau) = X_i\hat{\beta}_\tau + \hat{\delta}^*_{1\tau}\lambda^*_i(Z^*_A\hat{\gamma}^*) + \hat{\delta}^*_{2\tau}\lambda^*_i(Z^*_A\hat{\gamma}^*)^2 + \hat{\varepsilon}_{i\tau}$$

where $X_i$ has dimension $1 \times K$ and $Y_i$ is a scalar with sample size $i = 1, \ldots, N$. Following Firpo et al. (2009a), the regression of the RIF-function on $X_i$, $\lambda^*_i(Z^*_A\hat{\gamma}^*)$ and $\lambda^*_i(Z^*_A\hat{\gamma}^*)^2$ yields the

[40] For simplicity, $\lambda^*(Z^*_A\hat{\gamma}^*)^1 = \lambda^*(Z^*_A\hat{\gamma}^*)$ in the following.

following $\hat{\beta}_\tau$ coefficient vector[41]

$$\hat{\beta}_\tau = \frac{\hat{\zeta}(\hat{Q}_\tau)}{f_Y(\hat{Q}_\tau)}$$

with dimension $K \times 1$ and where $f_Y(\hat{Q}_\tau)$ is the kernel density estimator and

$$\hat{\zeta}(\hat{Q}_\tau) = \Omega_X^{-1} \frac{1}{N} \sum_{i=1}^{N} \left\{ X_i' \big( \hat{q}_\tau + \mathbb{1}\{Y_i > \hat{Q}_\tau\} - (1-\tau) \big) \right\}$$

with

$$\Omega_X = \frac{1}{N} \sum_{i=1}^{N} X_i' X_i \qquad \text{and} \qquad \hat{q}_\tau = \hat{Q}_\tau f_Y(\hat{Q}_\tau)^{-1}$$

Consequently, as $\hat{\delta}^*_{1\tau}$ and $\hat{\delta}^*_{2\tau}$ are the coefficient estimates obtained from the RIF-OLS regression of $\widehat{RIF}(\cdot)$ on $X_i$, $\lambda_i^*(Z_A^* \hat{\gamma}^*)$ and $\lambda_i^*(Z_A^* \hat{\gamma}^*)^2$, we have:

$$\hat{\delta}^*_{1\tau} = \frac{\hat{\eta}_1(\hat{Q}_\tau)}{f_Y(\hat{Q}_\tau)}$$

$$\hat{\delta}^*_{2\tau} = \frac{\hat{\eta}_2(\hat{Q}_\tau)}{f_Y(\hat{Q}_\tau)}$$

with

$$\hat{\eta}_1(\hat{Q}_\tau) = \Omega_{\lambda^*}^{-1} \frac{1}{N} \sum_{i=1}^{N} \left\{ \lambda_i^*(Z_A^* \hat{\gamma}^*) \big( \hat{q}_\tau + \mathbb{1}\{Y_i > \hat{Q}_\tau\} - (1-\tau) \big) \right\}$$

$$\Omega_{\lambda^*} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^*(Z_A^* \hat{\gamma}^*)' \lambda_i^*(Z_A^* \hat{\gamma}^*)$$

and

$$\hat{\eta}_2(\hat{Q}_\tau) = \Omega_{\lambda^{*2}}^{-1} \frac{1}{N} \sum_{i=1}^{N} \left\{ \lambda_i^*(Z_A^* \hat{\gamma}^*)^2 \big( \hat{q}_\tau + \mathbb{1}\{Y_i > \hat{Q}_\tau\} - (1-\tau) \big) \right\}$$

$$\Omega_{\lambda^{*2}} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^*(Z_A^* \hat{\gamma}^*)^{2'} \lambda_i^*(Z_A^* \hat{\gamma}^*)^2$$

Then, we have:

$$\sqrt{Nh}(\hat{\tilde{\beta}}_\tau - \tilde{\beta}_\tau) \xrightarrow{D} N(0, V_{OLS})$$

[41]Firpo et al. (2009a) consider the regression of the RIF-function on $X_i$ in the model without selection correction.

with

$$\hat{\tilde{\beta}}_\tau = \begin{pmatrix} \hat{\beta}_\tau \\ \hat{\delta}_{1\tau} \\ \hat{\delta}_{2\tau} \end{pmatrix} \qquad \text{and} \qquad \tilde{X} = \Big(X, \lambda^*(Z_A^*\hat{\gamma}^*), \lambda^*(Z_A^*\hat{\gamma}^*)^2\Big)$$

having dimension $K+1+1 \times 1$, where $K+1+1 = K^*$ and $V_{OLS} = V_{OLS}(Q_\tau, \kappa)$ with $\kappa(\cdot)$ being a real-value kernel function and positive bandwidth $h$. Following Firpo et al. (2009a), the asymptotic variance can then be represented as:

$$\begin{aligned} V_{OLS}(Q_\tau, \kappa) = \lim_{h\downarrow 0} \Bigg\{ & \frac{1}{f_y^2(Q_\tau)} \tilde{\beta}_\tau \tilde{\beta}_\tau' E[(f_y(Q_\tau))^2] + \\ & = + \frac{1}{f_y^2(Q_\tau)} Var[\sqrt{h}\tilde{\Omega}_{\tilde{X}}^{-1} \tilde{X} u(Q_\tau) + \tilde{\beta}_\tau (q_\tau + \mathbb{1}\{Y > Q_\tau\} - (1-\tau))] \Bigg\} \end{aligned}$$

where

$$\tilde{\Omega}_{\tilde{X}} = \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_i' \tilde{X}_i$$

$$u(Q_\tau) = q_\tau + \mathbb{1}\{Y > Q_\tau\} - (1-\tau)) - \tilde{X}_i \tilde{\zeta}(Q_\tau)$$

and

$$\tilde{\zeta}(Q_\tau) = \begin{pmatrix} \zeta_\tau(Q_\tau) \\ \eta_{1\tau}(Q_\tau) \\ \eta_{2\tau}(Q_\tau) \end{pmatrix}$$

The kernel density estimator, $\hat{f}_y(\hat{Q}_\tau)$, has an asymptotic squared bias that will go faster to zero than the variance (Firpo et al., 2009a). A possible estimator of $V_{OLS}(Q_\tau, \kappa)$ is $\hat{V}_{OLS}(\hat{Q}_\tau, h\kappa)$ (see Firpo et al., 2009a). Assuming that $E[u(Q_\tau)|\tilde{X}] = 0$ and $\tilde{\beta}_\tau = \text{UQPE}_\tau$, then:

$$\text{plim}_{h\downarrow 0} \hat{V}_{OLS}(\hat{Q}_\tau, h, \kappa) = V_{OLS}(Q_\tau, \kappa)$$

# Appendix 4.D The RIF-OLS Model with Parametric Estimators for Selection Correction

If the selection process is assumed to be normally distributed, the probit model can be used for selection adjustment. Following Heckman (1979), the RIF-OLS model corrected for sample selection using a parametric estimator for sample correction is:

$$\widehat{RIF}(Y;Q_\tau) = X\hat{\beta}_\tau + \hat{\delta}_\tau \lambda(Z_A\hat{\gamma}) + \hat{\varepsilon}_\tau \tag{4.D.1}$$

where $\lambda(Z_A\hat{\gamma})$ is the standard IMR evaluated at $Z_A\hat{\gamma}$, $\hat{\delta}_\tau$ is the corresponding coefficient estimate and $\hat{\varepsilon}_\tau$ is the quantile-specific residual. Asymptotic normality of the RIF-OLS model corrected for sample selection using a parametric estimator follows from the proof provided by Heckman (1979) for the parametric Heckman estimator at the mean.

The components of the quantile-specific GPG adjusted for selection with the parametric selection correction term are provided in Table 4.D.1. The effect of the estimated selection part due to differences in endowments is less strong compared to the results obtained in Section 4.4.5 but points generally in the same direction; positive at the bottom, negative at the top of the earnings distribution. At the median no effect is found in the model with parametric selection correction. Differences in the selection effect in terms of the unexplained part have again smaller point estimates but the same sign. Except for the median, where the selection effect is slightly negative in the parametric selection correction approach. Table 4.D.2 shows that gender differences in unobservables (given same prices) do not differ significantly across the wage distribution when the parametric correction approach is applied. Similarly, differences in prices between men and women to the same set of unobservables from higher to lower quantiles do not significantly impact on the variation of the GPG across the distribution.

Table 4.D.1 RIF-OLS Detailed Decomposition at Different Quantiles with Selection – Probit

| | (1) | (2) | (3) |
|---|---|---|---|
| | $10^{th}$ Percentile | $50^{th}$ Percentile | $90^{th}$ Percentile |
| GPG (Unadjusted Gap) | 0.117*** | 0.100*** | 0.179*** |
| | (0.010) | (0.004) | (0.009) |
| *Endowments Effect* | | | |
| Education | -0.010*** | -0.013*** | -0.029*** |
| | (0.002) | (0.001) | (0.003) |
| Experience | 0.018*** | 0.025*** | 0.067*** |
| | (0.004) | (0.002) | (0.005) |
| Job Characteristics | 0.006 | -0.004* | -0.030*** |
| | (0.006) | (0.002) | (0.006) |
| Occupations and Industries | 0.001 | -0.001 | -0.025*** |
| | (0.004) | (0.002) | (0.006) |
| Socio-Demographic Background | -0.005** | -0.009*** | -0.014*** |
| | (0.002) | (0.001) | (0.003) |
| Selection | 0.004** | 0.000 | -0.001 |
| | (0.001) | (0.001) | (0.002) |
| Total Explained | 0.012 | -0.000 | -0.032*** |
| | (0.008) | (0.004) | (0.009) |
| *Coefficients Effect* | | | |
| Education | -0.024 | -0.026** | 0.092*** |
| | (0.029) | (0.011) | (0.026) |
| Experience | 0.012 | 0.025** | 0.121*** |
| | (0.027) | (0.010) | (0.025) |
| Job Characteristics | -0.011 | 0.008 | 0.116** |
| | (0.052) | (0.020) | (0.048) |
| Occupations and Industries | -0.293*** | 0.051** | 0.225*** |
| | (0.070) | (0.026) | (0.062) |
| Socio-Demographic Background | 0.082 | -0.025 | -0.116 |
| | (0.135) | (0.052) | (0.130) |
| Selection | 0.110 | 0.006 | -0.064 |
| | (0.080) | (0.030) | (0.071) |
| Total Unexplained | 0.105*** | 0.100*** | 0.211*** |
| | (0.012) | (0.005) | (0.012) |

Bootstrapped standard errors in parentheses, 100 replications

$***p < 0.01, **p < 0.05, *p < 0.1$

Table 4.D.2 Gender Wage Inequality – Selection Component using Parametric Selection Correction

| | (1) | (2) | (3) |
|---|---|---|---|
| | 90-10 | 90-50 | 50-10 |
| | | Probit | |
| Unadjusted Change | 0.062*** | 0.079*** | -0.018** |
| | (0.011) | (0.009) | (0.008) |
| *Endowments Effect* | | | |
| Selection | -0.005 | -0.001 | -0.004 |
| | (0.008) | (0.002) | (0.002) |
| *Coefficients Effect* | | | |
| Selection | 0.174 | -0.070 | -0.104 |
| | (0.107) | (0.077) | (0.088) |

Bootstrapped standard errors in parentheses, 100 replications
$***p < 0.01, **p < 0.05, *p < 0.1$

# Appendix 4.E Definition of Variables

Table 4.E.1 Definition of Variables

| Variable Name | Definition |
|---|---|
| | **Dependent Variables** |
| Lhwage | The natural logarithm of net hourly wages; hourly wages in Euros, net of taxes and social security contributions |
| E (Employment) | One if the respective individual is in employment, zero otherwise |
| | **Independent Variables** |
| | **Dummy and Interaction Effects** |
| female | One if the respective individual is a woman, zero otherwise |
| Education × Experience | Quadratic interactions between educational dummies (*Elementary_School*, *High_School*, *University_Degree*) and experience (*Exper*) |
| Education × Occupation | Quadratic interactions between educational dummies (*Elementary_School*, *High_School*, *University_Degree*) and occupational dummies (Manager, Intermediate_Prof) |
| Experience × Occupation | Quadratic interactions between experience (*Exper*) and occupational dummies (*Manager*, *Intermediate_Prof*) |
| | **Education** |
| Elementary_School | One if the highest educational attainment of the individual is *Elementary_School*, zero otherwise |
| High_School | One if the highest educational attainment of the individual is *High_School*, zero otherwise; *High_School* corresponds to 13 years of schooling |
| University_Degree | One if the highest educational attainment of the individual is *University_Degree*, zero otherwise |
| Max_D_Mark | One if the best degree mark was attained (conditional on having a *University_Degree*), i.e. *110 e lode*, zero otherwise |
| | **Experience** |
| *Exper* | Number of years of prior work experience of the individual |
| Exper2 | *Exper* squared |
| Tenure | Number of years the individual has worked for his or her current employer |

Job Characteristics

| | |
|---|---|
| Work_Climate | Individual's level of satisfaction with the working climate at the individual's current job $\in (0,4)$ where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Stab | Individual's level of satisfaction with the stability of the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Time | Individual's level of satisfaction with the working time at the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Task | Individual's level of satisfaction with the tasks at the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Part | One if the individual holds a part-time contract, zero otherwise |
| Contract_Type | One if the individual holds an unlimited contract, zero otherwise |

Occupations and Industries

| | |
|---|---|
| Manager | Intellectual professions; scientific, and highly specialized occupations |
| Intermediate_Prof | Intermediary positions in commercial, technical or administrative sectors, health services and technicians |
| Sec_02 - Sec_15 | Sectoral dummies for employment in the following sectors or industries: manufacturing, energy, construction, tourism, commerce, transport, communication, financial activities, service industry, public administration, education, health, sciences and family services, respectively |

Socio-Demographic Background

| | |
|---|---|
| Age | Age of the individual (in years) $\in (18,64)$ |
| Age5064 | One if the age of the individual is between fifty and sixty-four years, zero otherwise |
| North | One if the individual lives and works in the North of Italy, zero otherwise |
| Centre | One if the individual lives and works in the Centre of Italy, zero otherwise |
| Homeowner | One if the individual owns a house (including houses financed by bank loans), zero otherwise |
| Partner_Works | One if the partner or the spouse of the individual is employed, zero otherwise |
| Married | One if the individual is married, zero otherwise |
| Italian | One if the individual is Italian, zero otherwise |
| Educ_Moth_Uni | One if the mother of the individual holds a university degree, zero otherwise |

| | |
|---|---|
| Educ_Fath_Uni | One if the father of the individual holds a university degree, zero otherwise |
| Kids | One if the individual has at least one child, zero otherwise |
| Kids_10 | One if the age of the youngest child of the individual is less than ten years, zero otherwise |
| Year_1-Year_5 | Year dummies, one if year = 2005, 2006, 2008, 2010, 2011, respectively, and zero otherwise |
| | Selection |
| $\lambda$ | Measures the selection bias from the employment decision |

# Appendix 4.F Regression Output OLS, UQR and CQR

Table 4.F.1 OLS, UQR and CQR of Log Hourly Wages

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | | $10^{th}$ Percentile | | $50^{th}$ Percentile | | $90^{th}$ Percentile | |
| | OLS | UQR | CQR | UQR | CQR | UQR | CQR |
| female | -0.122*** | -0.115*** | -0.117*** | -0.106*** | -0.112*** | -0.181*** | -0.150*** |
| | (0.005) | (0.012) | (0.007) | (0.005) | (0.004) | (0.012) | (0.009) |
| Elementary_School | -0.073*** | -0.009 | -0.040* | -0.065*** | -0.080*** | -0.108*** | -0.042 |
| | (0.019) | (0.044) | (0.023) | (0.016) | (0.014) | (0.031) | (0.036) |
| High_School | 0.109*** | 0.084*** | 0.095*** | 0.067*** | 0.091*** | 0.205*** | 0.126*** |
| | (0.006) | (0.016) | (0.009) | (0.005) | (0.004) | (0.012) | (0.009) |
| University_Degree | 0.202*** | 0.193*** | 0.163*** | 0.170*** | 0.193*** | 0.312*** | 0.258*** |
| | (0.009) | (0.021) | (0.016) | (0.009) | (0.007) | (0.021) | (0.013) |
| Max_D_Mark | 0.032* | 0.054** | 0.086*** | 0.042*** | 0.025** | 0.013 | 0.008 |
| | (0.016) | (0.026) | (0.026) | (0.012) | (0.011) | (0.035) | (0.028) |
| Exper | 0.015*** | 0.029*** | 0.019*** | 0.011*** | 0.014*** | 0.013*** | 0.011*** |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) | (0.002) | (0.001) |
| Exper2 | -0.000*** | -0.001*** | -0.000*** | -0.000*** | -0.000*** | -0.000*** | -0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure | 0.005*** | 0.002*** | 0.004*** | 0.004*** | 0.005*** | 0.010*** | 0.004*** |
| | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) |
| Work_Climate | -0.003 | 0.010 | -0.002 | -0.000 | -0.001 | -0.005 | -0.002 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (0.003) | (0.006) | (0.004) | (0.002) | (0.002) | (0.007) | (0.005) |
| Work_Stab | 0.015*** | 0.021*** | 0.032*** | 0.008*** | 0.008*** | 0.006 | 0.003 |
| | (0.003) | (0.005) | (0.004) | (0.002) | (0.002) | (0.006) | (0.004) |
| Work_Time | 0.013*** | 0.004 | 0.005 | 0.010*** | 0.013*** | 0.022*** | 0.011*** |
| | (0.003) | (0.006) | (0.005) | (0.002) | (0.002) | (0.005) | (0.004) |
| Work_Task | 0.010*** | 0.001 | -0.001 | 0.010*** | 0.009*** | 0.021*** | 0.023*** |
| | (0.003) | (0.007) | (0.005) | (0.003) | (0.003) | (0.007) | (0.007) |
| Part | 0.036*** | -0.046** | -0.040*** | 0.041*** | 0.043*** | 0.056*** | 0.098*** |
| | (0.008) | (0.021) | (0.015) | (0.007) | (0.006) | (0.015) | (0.013) |
| Contract_Type | 0.075*** | 0.204*** | 0.180*** | 0.050*** | 0.040*** | -0.017 | 0.017 |
| | (0.007) | (0.020) | (0.017) | (0.005) | (0.005) | (0.011) | (0.011) |
| Intermed_Prof | 0.055*** | 0.086*** | 0.062*** | 0.052*** | 0.058*** | 0.067*** | 0.047*** |
| | (0.005) | (0.013) | (0.008) | (0.005) | (0.004) | (0.013) | (0.009) |
| Manager | 0.116*** | 0.034* | 0.039** | 0.083*** | 0.120*** | 0.303*** | 0.188*** |
| | (0.010) | (0.018) | (0.018) | (0.007) | (0.007) | (0.025) | (0.016) |
| North | 0.056*** | 0.177*** | 0.103*** | 0.026*** | 0.038*** | 0.034*** | 0.032*** |
| | (0.006) | (0.015) | (0.010) | (0.005) | (0.005) | (0.010) | (0.009) |
| Centre | 0.025*** | 0.132*** | 0.071*** | -0.007 | 0.013** | -0.005 | 0.004 |
| | (0.007) | (0.017) | (0.011) | (0.005) | (0.006) | (0.013) | (0.009) |
| Italian | -0.001 | 0.086 | 0.002 | -0.005 | -0.020 | -0.080* | -0.067 |
| | (0.023) | (0.055) | (0.042) | (0.019) | (0.017) | (0.044) | (0.046) |
| Married | 0.067*** | 0.029*** | 0.053*** | 0.065*** | 0.056*** | 0.092*** | 0.074*** |
| | (0.005) | (0.009) | (0.007) | (0.005) | (0.005) | (0.014) | (0.010) |
| Homeowner | 0.029*** | 0.076*** | 0.024*** | 0.024*** | 0.025*** | 0.031*** | 0.043*** |
| | (0.006) | (0.015) | (0.009) | (0.005) | (0.005) | (0.012) | (0.008) |
| Educ_Fath_Uni | 0.026** | -0.026 | -0.021 | 0.021** | 0.033*** | 0.073*** | 0.039** |
| | (0.013) | (0.020) | (0.024) | (0.009) | (0.011) | (0.027) | (0.018) |
| Educ_Moth_Uni | 0.005 | 0.042 | 0.018 | 0.014 | 0.002 | -0.018 | 0.026 |
| | (0.015) | (0.032) | (0.026) | (0.011) | (0.008) | (0.028) | (0.021) |
| Constant | 1.418*** | 0.616*** | 0.959*** | 1.589*** | 1.500*** | 1.825*** | 1.861*** |
| | (0.032) | (0.082) | (0.053) | (0.023) | (0.025) | (0.066) | (0.061) |
| Sectoral Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Robust standard errors in parentheses for OLS

Bootstrapped standard errors in parentheses, 100 replications, for UQR and CQR

# References


Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a Glass Ceiling in Sweden?. *Journal of Labor Economics*, 21(1), 145-177.

Albrecht, J., van Vuuren, A., and Vroman, S. (2009).Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands. *Labour Economics*, 16(4), 383-396.

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2007). Is there a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial and Labor Relations Review*, 60(2), 163-186.

Bar, M., Kim, S., and Leukhina, O. (2015). Gender Wage Gap Accounting: The Role of Selection Bias. *Demography*, 52(5), 1729-1750.

Blau, F. D., and Kahn, L. M. (1992). The Gender Earnings Gap: Learning from International Comparisons. *The American Economic Review*, 82(2), 533-538.

Blau, F. D., and Kahn, L. M. (2003). Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1), 106-144.

Blau, F. D., and Kahn, L. M. (2006). The US Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blau, F. D., and Kahn, L. M. (2016). *The Gender Wage Gap: Extent, Trends, and Explanations* (No. W21913). National Bureau of Economic Research.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4), 436-455.

Borah, B. J., and Basu, A. (2013). Highlighting Differences between Cconditional and Unconditional Quantile Regression Approaches through an Application to Assess Medication Adherence. *Health Economics*, 22(9), 1052-1070.

Bordogna, L. (2012). Employment Relations and Union Action in the Italian Public Services – Is There a Case of Distortion of Democracy. *Comparative Labor Law and Policy Journa*l, 34(2), 507-530.

Brekke, K. A., and Nyborg, K. (2010). Selfish Bakers, Caring Nurses? A Model of Work Motivation. *Journal of Economic Behavior and Organization*, 75(3), 377-394.

Buchinsky, M. (1998). The Dynamics of Change in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*, 13(1), 1-30.

Cameron, A., and Trivedi, P. (2009). *Microeconometrics: Methods and Application*. Cambridge University Press.

Cardoso, A. R., Guimarães, P., and Portugal, P. (2016). What Drives the Gender Wage Gap? A Look at the Role of Firm and Job-Title Heterogeneity. *Oxford Economic Papers*, 68(2), 506-524.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *The American Economic Review*, 101(6), 2754-2781.

Chang, S. K. (2011). Simulation Estimation of Two-Tiered Dynamic Panel Tobit Models with an Application to the Labor Supply of Married Women. *Journal of Applied Econometrics*, 26(5), 854-871.

Chzhen, Y., and Mumford, K. (2011). Gender Gaps Across the Earnings Distribution for Full-Time Employees in Britain: Allowing for Sample Selection. *Labour Economics*, 18(6), 837-844.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions. *Labour Economics*, 41, 47-60.

Cotton, J. (1988). On the Decomposition of Wage Differentials. *The Review of Economics and Statistics*, 70(2), 236-243.

Devereux, P. J. (2004). Changes in Relative Wages and Family Labor Supply. *Journal of Human Resources*, 39(3), 698-722.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5), 1001-1044.

Donald S. G., Green, D. A. and Paarsch, H. J. (2000). Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates. *The Review of Economic Studies*, 67(4), 609-633.

England, P. (2006). Toward Gender Equality: Progress and Bottlenecks. In Blau F., Brinton M., and Grusky D. (Eds.), *The Declining Significance of Gender?* (pp. 245-264). Russell Sage Foundation.

Eurostat (2017a). *Europe 2020 Employment Indicator*. http://ec.europa.eu/eurostat/documents/2995521/7997105/3-25042017-BP-EN.pdf/377b4834-5a19-42f4-8a2d-36e133ed887d (accessed 27-05-2017).

Eurostat (2017b). *Gender Pay Gap in Unadjusted Form*. http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (accessed 27-05-2017).

Filer, R. K. (1985). Male-Female Wage Differences: The Importance of Compensating Differentials. *Industrial and Labor Relations Review*, 38(3), 426-437.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009a). Supplement to 'Unconditional Quantile Regressions'. *Econometrica Supplemental Material*, 77(3).

Firpo, S., Fortin, N. M., and Lemieux, T. (2009b). Unconditional Quantile Regressions. *Econometrica*, 77(3), 953-973.

Fortin, N. M., Lemieux, T., and Firpo, S. (2011). *Decomposition Methods in Economics*. Handbook of Labor Economics, 4, 1-102.

Frölich, M. (2006). Non-Parametric Regression for Binary Dependent Variables. *The Econometrics Journal*, 9(3), 511-540.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much?. *Journal of Labor Economics*, 34(2), 509-543.

Ghosh, P. K. (2014). The Contribution of Human Capital Variables to Changes in the Wage Distribution Function. *Labour Economics*, 28(C), 58-69.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091-1119.

Grove, W. A., Hussey, A., and Jetter, M. (2011). The Gender Pay Gap Beyond Human Capital: Heterogeneity in Noncognitive Skills and in Labor Market Tastes. *Journal of Human Resources*, 46(4), 827-874.

Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Heckman, J. J. (1990). Varieties of Selection Bias. *The American Economic Review*, 80(2), 313-318.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*, Cambridge University Press.

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score. *Econometrica*, 71(4), 1161-1189.

Horowitz, J. L., and Härdle, W. (1994). Testing a Parametric Model Against a Semiparametric Alternative. *Econometric Theory*, 10(5), 821-848.

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single- Index Models. *Journal of Econometrics*, 58(1-2), 71-120.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, 101(3), 410-442.

Kahn, L. M. (2015). Wage Compression and the Gender Pay Gap. *IZA World of Labor*, 150.

Klein, R. W., and Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*, 61(2), 387-421.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, 76(3), 1071-1102.

Longhi, S., Nicoletti, C., and Platt, L. (2012). Explained and Unexplained Wage Gaps across the Main Ethno-Religious Groups in Great Britain. *Oxford Economic Papers*, 65(2), 471-493.

Lucifora, C., and Meurs, D. (2006). The Public Sector Pay Gap In France, Great Britain and Italy. *Review of Income and Wealth*, 52(1), 43-59.

Machado, J. A., and Mata, J. (2005). Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. *Journal of Applied Econometrics*, 20(4), 445-465.

Martins, M. F. O. (2001). Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics*, 16(1), 23-39.

Melly, B. (2005a). Decomposition of Differences in Distribution using Quantile Regression. *Labour Economics*, 12(4), 577-590.

Melly, B. (2005b). Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression. *Empirical Economics*, 30(2), 505-520.

Mínguez, A. M. (2004). The Persistence of Male Breadwinner Model in Southern European Countries in a Compared Perspective: Familism, Employment and Family Policies. *Marie Curie Annals*, 4.

Mulligan, C. B., and Rubinstein, Y. (2008). Selection, Investment, and Women's Relative Wages Over Time. *The Quarterly Journal of Economics*, 123(3), 1061-1110.

Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *Journal of Human Resources*, 23(3), 279-295.

Newey, W. K. (2009). Two-Step Series Estimation of Sample Selection Models. *The Econometrics Journal*, 12(1), 217-229.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

Oaxaca R. and Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61(1), 5-21.

Piazzalunga, D., and Di Tommaso, M. L. (2015). *The Increase of the Gender Wage Gap in Italy during the 2008-2012 Economic Crisis*. IZA Discussion Papers, No. 9931.

Pistaferri, L. (1999). Informal Networks in the Italian Labor Market. *Giornale Degli Economisti e Annali di Economia*, 58(3-4), 355-375.

Proenca, I. (1993). *On the Performance of the HH Test* (No. 1993005). Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Reimers, C. W. (1983). Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics*, 65(4), 570-579.

Worldwide Tax Summaries (2017). *Italy: Individual – Other Tax Credits and Incentives*. http://taxsummaries.pwc.com/ID/Italy-Individual-Other-tax-credits-and-incentives (accessed 01-06-2017).

Xiu, L., and Gunderson, M. (2014). Glass Ceiling or Sticky Floor? Quantile Regression Decomposition of the Gender Pay Gap in China. *International Journal of Manpower*, 35(3), 306-326.

# Chapter 5

# The Convergence of the Gender Pay Gap – An Alternative Estimation Approach –

## 5.1 Introduction

Gender differentials in the labor market have obtained much attention from policy makers and researchers leading to the implementation of equal-pay legislation and the promotion of equal opportunities. Even though equal-pay legislation and equal opportunities have been promoted in Western industrialized countries for several decades, differences in pay between men and women persist (e.g. Blau and Kahn, 1992; 2003; 2006; 2016; Goldin, 2014). For example, in the European Union (EU) in 2014, women earned on average 16.7% less than men (Eurostat, 2017).

Typically, different Gender Pay Gaps (GPGs) are found across time. In particular, declining GPGs are observed with slower convergence in recent decades (see Blau and Kahn, 2006; England, 2006). The main reasons for the decline of the GPG over time are found to be the catching-up of women in terms of education and labor market experience (Goldin, 2006), technical development (Black and Spitz-Oener, 2010), changes in attitudes towards women in the labor market, less occupational segregation (Cotter et al., 2004; England, 2006) and anti-discrimination laws (Fortin, 2015). Research has shown that the unexplained or coefficients effect of the GPG is reduced subsequently over time (e.g. Mandel and Semyonov, 2014). Differences in pay are revealed also across sectors and especially between the public and the private sector. The Public-Private Sector Wage Gap (PPWG) is found to differ significantly for men and women (Melly, 2005; Lucifora and Meurs, 2006; Arulampalam et al., 2007). In fact, the difference in pay by gender is found to be smaller in the public compared to the private sector (see for example Melly, 2005; Arulampalam et al., 2007). Regardless of gender, pay levels in the public sector are on average higher than in the private sector (Lucifora and Meurs, 2006). The public sector is generally the preferred sector of women due to its fairer recruitment, selection criteria and remuneration as well as better implementation of anti-discrimination laws (Gornick and Jacobs, 1998; Grimshaw, 2000).

However, studies examining changes in the wage gap over time and between groups do not directly estimate the difference of the GPG in year $t$ and year $t+1$ (or the wage gap by sector for men and women for instance), but rather compare the results of the pay gaps in the corresponding subsamples ex post (e.g. Christofides and Michael, 2013; Mandel and Semyonov, 2014). Studies estimating the difference of the pay gaps in different subsamples, often do not even provide standard errors for the decomposition (Mandel and Semyonov, 2014; Bar et al., 2015). Hence, it is not possible to conduct statistical inference. Indeed, this does not allow to draw conclusions on which of the two wage gaps is more statistically significant, i.e. whether the difference between the two pay gaps under investigation is statistically significantly different from zero. Additionally, the conclusion about drivers of the change of pay gaps between groups may be different, when estimated directly compared to analyzing results estimated in different subsamples. The reason is that it is not possible to draw direct inference of the difference of the respective components in the latter case. Moreover, the standard method, i.e. ex-post comparison of the decomposition results, does not allow to catch time- (or sector-) and gender-specific effects that may exist simultaneously, i.e. interactions across gender and time or sector and gender (in the case of the GPG over time and the PPWG by gender, respectively). We slightly extend the method proposed by Gelbach (2016) that is based on the Omitted Variable Bias (OVB) formula to estimate directly the difference between two wage gaps. We are then able to draw inference on the changes of the pay gap by groups across subsamples and to compare the various contributors directly, i.e. we can test whether there has been a significant change of the explained or unexplained part of the gap. Moreover, we can draw conclusions on the relevance of interaction effects across subsamples and groups. The standard method in applied labor economics, when it comes to pay gaps between groups is the Oaxaca (1973) and Blinder (1973) decomposition method (Fortin et al., 2011). The approach, however, suffers from non-invariance with respect to categorical variables and the index-number problem. The intercept-shift approach attempts to solve the latter but suffers, in particular, from the indeterminacy problem (Lee, 2015). We extend our proposed method based on the OVB formula and show that it can be made robust to the above mentioned problems.

We apply our model to two cases. First, we examine the evolution of the GPG over ten years, from 2005 to 2014 in Italy. Second, we analyze the PPWG between men and women in 2014 in Italy. We analyze each case with the standard Oaxaca-Blinder decomposition method and then repeat the examination with our proposed extension of the Gelbach decomposition. We expect to find a statistically significant change in differences in observable characteristics (such as educational attainment, labor market presence as well as job-, industry- or occupational-specific characteristics) by gender over time as well as a statistically significant change in differences in coefficients to these characteristics between men and women over time. In fact, the latter may indicate the effectiveness of anti-discrimination policies. For the second empirical application, the PPWG by gender, we expect, to find in line with the literature larger pay gaps for women between the public and the private sector than for men. Additionally, we expect to find a larger effect of the unexplained component in the PPWG for women; while differences in endowments may be the main driver of the pay differential for men, they may not explain equally the difference in the PPWG for women.

For the first case, the findings of the study reveal interesting differences in results when applying our proposed estimation methodology compared to the 'standard' approach.[1] Changes in gender differences of observable characteristics are found to be the only statistically significant driving force of the convergence of the GPG in the last decade in Italy. In contrast, by comparing the different components of the GPGs following Oaxaca (1973) and Blinder (1973), differences in returns to observable characteristics, often referred to as the unexplained part of the GPG, seem to play a role in closing the gap over the last decade in Italy. In the second case, we can confirm the conclusions drawn from the estimation in the respective subsamples; the higher PPWG for women than for men is due to both differences in the explained and unexplained component.

The paper is organized as follows. Section 5.2 presents the standard Oaxaca-Blinder decomposition. In Section 5.3, we outline the method by Gelbach (2016) as well as our proposed modification. Similarly, we discuss problems of the standard approach and show the robustness of our method to these problems. Next, in Section 5.4, we empirically apply the method proposed to the GPG over time as well as to the PPWG by gender and discuss the results obtained. Section 5.5 concludes.

[1] i.e. the Oaxaca-Blinder decomposition and ex-post comparison of the decomposition results.

## 5.2 Standard Estimation Strategy

The standard methodology to decompose pay differentials between two groups is the Oaxaca (1973) and Blinder (1973) decomposition. The methodology estimates Mincer-type wage regressions separately for a specific group (e.g. men or women, the public or the private sector) and then decomposes the wage differential in different components. We use the three-fold Oaxaca-Blinder approach and thus decompose the pay gap in three components; endowments, coefficients and interactions:[2]

$$\begin{aligned}\overline{ln(w_0)} - \overline{ln(w_1)} &= \hat{\alpha}_0 + \bar{X}_0\hat{\beta}_0 - \hat{\alpha}_1 - \bar{X}_1\hat{\beta}_1 \\ &= (\bar{X}_0 - \bar{X}_1)\hat{\beta}_1 + (\hat{\alpha}_0 - \hat{\alpha}_1) + \bar{X}_1(\hat{\beta}_0 - \hat{\beta}_1) \\ &+ (\bar{X}_0 - \bar{X}_1)(\hat{\beta}_0 - \hat{\beta}_1)\end{aligned}$$

where $\overline{ln(w_G)}$ is the logarithmic hourly wage of group $G$ evaluated at the mean, $\hat{\alpha}_G$ is the intercept of group $G$ and $\bar{X}'_G$ and $\hat{\beta}_G$ are $K \times 1$ vectors of average characteristics and estimated coefficients for $G \in \{0,1\}$. The first term is the effect due to differences in observable characteristics. As different observed characteristics are expected to have different effects on earnings, the difference in observable characteristics is also referred to as the explained component, the quantity or endowments effect of the Oaxaca-Blinder decomposition. The second term is due to differences in the starting point, i.e. differences in the intercept. The third term is the effect due to differences in returns on the same set of observable characteristics. This component is generally referred to as the unexplained part, price or coefficients effect of the gap. Differences in the intercept are attributed to the coefficients component. In the case of the GPG, if the differential is mainly due to the price effect, this may indicate the presence of gender discrimination.[3] The last term is the so-called interaction term. The intuition behind is that differences in endowments and coefficients may exist simultaneously between groups (Jann, 2008).

[2]An alternative to the three-fold decomposition outlined here is the standard two-fold decomposition that decomposes the wage differential in an explained and an unexplained part;

$$\begin{aligned}\overline{ln(w_0)} - \overline{ln(w_1)} &= \hat{\alpha}_0 + \bar{X}_0\hat{\beta}_0 - \hat{\alpha}_1 - \bar{X}_1\hat{\beta}_1 \\ &= (\bar{X}_0 - \bar{X}_1)\hat{\beta}_0 + (\hat{\alpha}_0 - \hat{\alpha}_1) + \bar{X}_1(\hat{\beta}_0 - \hat{\beta}_1)\end{aligned}$$

We focus here on the three-fold decomposition, as we argue that interaction effects may be important when considering differences across pay gaps.

[3]However, as pointed out by Blau and Kahn (2006), the unexplained portion of the GPG may include effects of unobserved characteristics such as individual productivity, motivation or educational quality.

## 5.3 Econometric Model

We propose a slight modification of the decomposition method by Gelbach (2016). The Gelbach approach decomposes cross-specification differences in Ordinary Least Squares (OLS) estimates of the group-dummy coefficient from the wage model in a path-independent way yielding a Oaxaca-Blinder type decomposition. By using the OVB formula, the decomposition is consistently estimated conditional on all covariates used in the regression. This method, similar to the standard estimation approach outlined in Section 5.2, decomposes the sample mean difference in wages between different groups in an explained and an unexplained part (see Gelbach, 2016, for details).

### 5.3.1 Extension of Gelbach (2016)

The model outlined in the following allows not only to obtain information on whether the pay gap has decreased in a statistically significant way on aggregate but also to testify what are the main contributors to the change (if any) of the differential. Consider the case, when we estimate the wage equation separately by $G$ (group) and $Y$ (data wave or a group different from $G$, i.e. $Y \neq G$) for individual $i$, with $i = 1, 2, \ldots N$:

$$ln(w_{iGY}) = \alpha_{GY} + X_{iGY}\beta_{GY} + \varepsilon_{iGY} \tag{5.1}$$

with $G \in \{0,1\}$, $Y \in \{A,B\}$; and where $ln(w_{iGY})$ is individual $i$'s logarithmic wage of $G$ in $Y$, $\alpha_{GY}$ is a constant, $X_{iGY}$ is a $1 \times K$ vector of exogenous regressors, $\beta_{GY}$ is the corresponding $K \times 1$ vector of coefficients and $\varepsilon_{iGY}$ is the error term.[4] When we evaluate the estimation at the mean given the OLS property that OLS estimates must go through the mean of the data, equation (5.1) becomes:

$$\overline{ln(w_{GY})} = \hat{\alpha}_{GY} + \bar{x}_{GY}\hat{\beta}_{GY} \tag{5.2}$$

where $\hat{\alpha}_{GY}$ is the constant, $\bar{x}_{GY}$ is the $1 \times K$ row vector of sample means of observable characteristics in $X$:

$$\bar{x}_{GY} = \left[\bar{x}_{GYk1}, \bar{x}_{GYk2}, \ldots, \bar{x}_{GYK}\right]$$

[4] In the first empirical application in Section 5.4, we set the index $G$ equal to gender and the index $Y$ equal to different years or waves of the data set. Consequently, in case 1 of the empirical implementation, we have for $G \in \{0,1\}$; $0 =$ male and $1 =$ female and for $Y \in \{A,B\}$; $A =$ starting period or 2005 and $B =$ ending period or 2014. In the second empirical example shown in Section 5.4, group $G$ represents different sectors and $Y$ men or women. Thus, in case 2 of the empirical part, we have for $G \in \{0,1\}$; $0 =$ public-sector employment and $1 =$ private sector employment and for $Y \in \{A,B\}$; $A =$ female and $B =$ male.

and $\hat{\beta}_{GY}$ is the corresponding $K \times 1$ vector of parameter estimates. Four different pairs of $(G,Y)$ and thus four regressions of equation (5.2) are possible; $(0,A)$, $(0,B)$, $(1,A)$, $(1,B)$. The corresponding regressions between $G$ and $Y$ are conducted by assuming the same set of regressors for all four cases.

Now, consider estimating the joint model. The first group index $G$ is added to the regression as a dummy variable $G_i$ among the controls on the right-hand side. Analogously, the second group index $Y$ is transformed in a dummy variable $Y_i$ controlling for group $Y$ membership. The indicator variable takes value one, if the observation corresponds to $A$ and takes value zero, if we observe $B$.[5] As in Gelbach (2016), we distinguish between two sets of regressors, $X_{i1}$ and $X_{i2}$, where the set of regressors $X_{i1}$, with dimension $1 \times 4$, is the base specification containing only (for each observation $i$) a constant, an interaction term between the group dummies, $G_iY_i$, as well as the dummies, $G_i$ and $Y_i$, separately. The interaction of the dummies for group membership $G_i$ and $Y_i$ are contained in $G_iY_i$. The base model is therefore defined as follows:

$$\begin{aligned} ln(w_{iGY}) &= X_{i1}\alpha^{base} + \varepsilon_{iGY}^{base} \\ ln(w_{iGY}) &= \alpha_0^{base} + G_iY_i\alpha_1^{base} + G_i\alpha_2^{base} + Y_i\alpha_3^{base} + \varepsilon_{iGY}^{base} \end{aligned} \tag{5.3}$$

where $\alpha_0^{base}$ is the constant and $\alpha_1^{base}, \alpha_2^{base}, \alpha_3^{base}$ are the corresponding coefficients contained in the $4 \times 1$ column vector $\alpha^{base}$, $\varepsilon_{iGY}^{base}$ is the corresponding error term. The second set of regressors, $X_{i2}$, has dimension $1 \times 4K$ and contains the $1 \times K$ vector of explanatory variables $X_i$ as well as the interactions of $X_i$ with $G_i$, $Y_i$ and $G_iY_i$, respectively. The set of regressors $X_{i2}$ will be considered later as omitted variables in order to obtain a decomposition of the change of the wage gap between $G_i$ across $Y_i$. The full model is then defined as:

$$\begin{aligned} ln(w_{iGY}) &= X_{i1}\alpha^{full} + X_{i2}\beta + \varepsilon_{iGY}^{full} \\ ln(w_{iGY}) &= \alpha_0^{full} + G_iY_i\alpha_1^{full} + G_i\alpha_2^{full} + Y_i\alpha_3^{full} + X_i\beta_1 + G_iX_i\beta_2 + Y_iX_i\beta_3 + G_iY_iX_i\beta_4 + \varepsilon_{iGY}^{full} \end{aligned} \tag{5.4}$$

[5] We thus have the index $G \in \{0,1\}$ and the dummy variable $G_i$, with

$$G_i = \begin{cases} 1 & \text{if the index of person } i \text{ is } G = 1 \\ 0 & \text{if the index of person } i \text{ is } G = 0 \end{cases}$$

For the second group, we have the index $Y \in \{A,B\}$ and the dummy variable $Y_i$, with

$$Y_i = \begin{cases} 1 & \text{if the index of person } i \text{ is } Y = A \\ 0 & \text{if the index of person } i \text{ is } Y = B \end{cases}$$

where $\alpha^{full}$ and $\beta$ are the $4 \times 1$ and $4K \times 1$ vectors of coefficients from $X_{i1}$ and $X_{i2}$, respectively. The error term is represented by $\varepsilon_{iGY}^{full}$.

We can recast the parameters of the full model evaluated at the mean from the pair-wise regressions of (5.2):

1. When (the indices) G=1 and Y=A, we get:
   - $\hat{\alpha}_{1A} = \hat{\alpha}_0^{full} + \hat{\alpha}_1^{full} + \hat{\alpha}_2^{full} + \hat{\alpha}_3^{full}$
   - $\hat{\beta}_{1A} = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4$
2. When (the indices) G=0 and Y=A, we get:
   - $\hat{\alpha}_{0A} = \hat{\alpha}_0^{full} + \hat{\alpha}_3^{full}$
   - $\hat{\beta}_{0A} = \hat{\beta}_1 + \hat{\beta}_3$
3. When (the indices) G=1 and Y=B, we get:
   - $\hat{\alpha}_{1B} = \hat{\alpha}_0^{full} + \hat{\alpha}_2^{full}$
   - $\hat{\beta}_{1B} = \hat{\beta}_1 + \hat{\beta}_2$
4. When (the indices) G=0 and Y=B, we get:
   - $\hat{\alpha}_{0B} = \hat{\alpha}_0^{full}$
   - $\hat{\beta}_{0B} = \hat{\beta}_1$

Re-arranging the terms slightly, gives us:

$$\begin{aligned}
\hat{\alpha}_0^{full} &= \hat{\alpha}_{0B} \\
\hat{\alpha}_2^{full} &= \hat{\alpha}_{1B} - \hat{\alpha}_{0B} \\
\hat{\alpha}_3^{full} &= \hat{\alpha}_{0A} - \hat{\alpha}_{0B} \\
\hat{\alpha}_1^{full} &= \hat{\alpha}_{1A} - \hat{\alpha}_{0B} - \hat{\alpha}_{1B} + \hat{\alpha}_{0B} - \hat{\alpha}_{0A} + \hat{\alpha}_{0B} \\
&= (\hat{\alpha}_{0B} - \hat{\alpha}_{1B}) - (\hat{\alpha}_{0A} - \hat{\alpha}_{1A})
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_1 &= \hat{\beta}_{0B} \\
\hat{\beta}_2 &= \hat{\beta}_{1B} - \hat{\beta}_{0B} \\
\hat{\beta}_3 &= \hat{\beta}_{0A} - \hat{\beta}_{0B} \\
\hat{\beta}_4 &= \hat{\beta}_{0B} - \hat{\beta}_{1B} - \hat{\beta}_{0A} + \hat{\beta}_{1A} \\
&= (\hat{\beta}_{0B} - \hat{\beta}_{1B}) - (\hat{\beta}_{0A} - \hat{\beta}_{1A})
\end{aligned}$$

By evaluating the base model at the mean and considering the set of regressors $X_2$ as omitted variables, we obtain the following specification:[6]

$$\hat{\alpha}^{base} = \hat{\alpha}^{full} + (X_1'X_1)^{-1}X_1'X_2\hat{\beta}^{full} \tag{5.5}$$

where

- $(X_1'X_1)^{-1}X_1'X_2\hat{\beta}^{full}$ is the OVB
- The parameter estimates from the base model (5.3) evaluated at the mean are:

  $$\hat{\alpha}^{base} = \left[\hat{\alpha}_0^{base}, \hat{\alpha}_1^{base}, \hat{\alpha}_2^{base}, \hat{\alpha}_3^{base}\right]'$$

  being a $4 \times 1$ column vector.
- $\hat{\alpha}^{full}$ is the $4 \times 1$ column vector containing the coefficient estimates of $X_{i1}$ from the full model (5.4) evaluated at the mean.
- $(X_1'X_1)^{-1}X_1'X_2$ is the linear projection of $X_2$ on $X_1$, with dimension $4 \times 4K$.
- $$\hat{\beta}^{full} = \left[\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4\right]'$$

  is a $4K \times 1$ column vector of coefficients from the full model (5.4) evaluated at the mean. The model specification in equation (5.5) can be decomposed as follows:

  $$\hat{\alpha}^{base} = \hat{\alpha}^{full} + \hat{\delta}^1 + \hat{\delta}^2 + \hat{\delta}^3 + \hat{\delta}^4 \tag{5.6}$$

  with $\hat{\delta} \equiv \hat{\alpha}^{base} - \hat{\alpha}^{full} = (X_1'X_1)^{-1} X_1'X_2\hat{\beta}^{full}$, where

  - $\hat{\delta}^q = \hat{\Gamma}^q\hat{\beta}_q^{full}$, with $\hat{\Gamma}^q = (X_1'X_1)^{-1} X_1'X_{2q}$ of dimension $k_{X_1} \times k_q$ and $X_{2q}$ being the $q$th column of $X_2$, for $q = 1,..,Q$. The column vector $\hat{\beta}_q^{full}$ has dimension $k_q \times 1$, thus $\hat{\delta}_q$ is a $k_{X_1} \times 1$ column vector;
  - $k_{X_1}$ is equal to the number of regressors from $X_1$, i.e. 4 in our case ($X_1$ contains sample means and has dimension $1 \times 4$);
  - $k_q$ is equal to the number of regressors in the $q$th column of $X_2$.

[6]Notably, the set of regressors $X_1$ contains the sample means of $X_{i1}$, while the set of regressors $X_2$ contains the sample means of $X_{i2}$.

### 5.3.2 Decomposition

Recall that we are interested in the estimation and decomposition of the change in the pay gap between group $G$ across group $Y$, i.e.[7]

$$\Delta^{\mathrm{B}} - \Delta^{\mathrm{A}} = \left(\overline{ln(w_{0B})} - \overline{ln(w_{1B})}\right) - \left(\overline{ln(w_{0A})} - \overline{ln(w_{1A})}\right)$$

with $\Delta^{\mathrm{B}}$ being the pay gap by group $G$ given that $Y = B$ and $\Delta^{\mathrm{A}}$ being the wage gap between $G$ given that $Y = A$. From equation (5.2), we know that:

$$\begin{aligned} \Delta^{\mathrm{B}} &= \left(\overline{ln(w_{0B})} - \overline{ln(w_{1B})}\right) \\ &= -\hat{\alpha}_2^{base} \end{aligned}$$

$$\begin{aligned} \Delta^{\mathrm{A}} &= \left(\overline{ln(w_{0A})} - \overline{ln(w_{1A})}\right) \\ &= -\hat{\alpha}_1^{base} - \hat{\alpha}_2^{base} \end{aligned}$$

and hence $\hat{\alpha}_1$ represents the difference of the two wage gaps:

$$\Delta^{\mathrm{B}} - \Delta^{\mathrm{A}} = \hat{\alpha}_1^{base}$$

Given the definition of $\hat{\alpha}^{base}$, we are interested in the second row of $\hat{\alpha}^{base}$, i.e. of equation (5.5), or $\hat{\alpha}_1^{base}$ in order to obtain the change of the wage gaps, $\Delta^{\mathrm{B}} - \Delta^{\mathrm{A}}$. Starting from equation (5.5), we calculate the second row of the $4 \times 4K$ matrix $(X_1'X_1)^{-1}X_1'X_2$ considering the sample means of observable characteristics:

$$\kappa = \left[(\bar{x}_{0B} - \bar{x}_{1B}) - (\bar{x}_{0A} - \bar{x}_{1A}), (\bar{x}_{1A} - \bar{x}_{1B}), (\bar{x}_{1A} - \bar{x}_{0A}), \bar{x}_{1A}\right]$$

with dimension $1 \times 4K$. The second row of equation (5.5) or the difference of the respective wage gap evaluated at the mean is thus:

$$\hat{\alpha}_1^{base} = \hat{\alpha}_1^{full} + \kappa \hat{\beta}^{full} \tag{5.7}$$

[7]For example, the change of the GPG across two years.

and can be re-written as:

$$\begin{aligned}
\hat{\alpha}_1{}^{base} &= \underbrace{(\hat{\alpha}_{0B} - \hat{\alpha}_{1B}) - (\hat{\alpha}_{0A} - \hat{\alpha}_{1A})}_{\hat{\alpha}_1^{full}} + [(\bar{x}_{0B} - \bar{x}_{1B}) - (\bar{x}_{0A} - \bar{x}_{1A})]\underbrace{\hat{\beta}_{0B}}_{\hat{\beta}_1} \\
&+ (\bar{x}_{1A} - \bar{x}_{1B})\underbrace{(\hat{\beta}_{1B} - \hat{\beta}_{0B})}_{\hat{\beta}_2} \\
&+ (\bar{x}_{1A} - \bar{x}_{0A})\underbrace{(\hat{\beta}_{0A} - \hat{\beta}_{0B})}_{\hat{\beta}_3} \\
&+ \bar{x}_{1A}\underbrace{[(\hat{\beta}_{0B} - \hat{\beta}_{1B}) - (\hat{\beta}_{0A} - \hat{\beta}_{1A})]}_{\hat{\beta}_4} \\
&= \Delta^B - \Delta^A
\end{aligned} \tag{5.8}$$

where $\hat{\alpha}_1{}^{base}$ and $\hat{\alpha}_1{}^{full}$ are scalars and $\bar{x}'_{GY}$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ are $K \times 1$ column vectors, respectively. The above expression can be re-written as a 'double' (two-fold) Oaxaca-Blinder decomposition:

$$\begin{aligned}
\hat{\alpha}_1{}^{base} &= (\hat{\alpha}_{0B} - \hat{\alpha}_{1B}) + (\bar{x}_{0B} - \bar{x}_{1B})\hat{\beta}_{0B} + \bar{x}_{1B}(\hat{\beta}_{0B} - \hat{\beta}_{1B}) \\
&- [(\hat{\alpha}_{0A} - \hat{\alpha}_{1A}) + (\bar{x}_{0A} - \bar{x}_{1A})\hat{\beta}_{0A} + \bar{x}_{1A}(\hat{\beta}_{0A} - \hat{\beta}_{1A})]
\end{aligned}$$

Decomposing the change in the wage gap between group $G$ across group $Y$ in the following way allows to better understand the elements that contribute to the earnings differences across $G$ and $Y$: $\Delta^{\mathrm{B}} - \Delta^{\mathrm{A}} = E + U + I1 + I2$, with

$$E = [(\bar{x}_{0B} - \bar{x}_{1B}) - (\bar{x}_{0A} - \bar{x}_{1A})]\hat{\beta}_{0B} \tag{5.9}$$

Here, the same prices, namely the ones of the respective base category, $\hat{\beta}_{0B}$, are assumed. Thus, $E$ measures the amount of the change of the gap attributable to differences in observed characteristics. It is the component referred to as differences in quantities, i.e. the explained part. The unexplained component becomes the following:

$$\begin{aligned}
U &= \hat{\alpha}_1^{full} + \underbrace{\bar{x}_{1A}[(\hat{\beta}_{0B} - \hat{\beta}_{1B}) - (\hat{\beta}_{0A} - \hat{\beta}_{1A})]}_{u} \\
&= \hat{\alpha}_1^{full} + u
\end{aligned} \tag{5.10}$$

$U$ measures the change of differences in the intercepts, $\hat{\alpha}_1^{full}$, as well as the change over $Y$ of the differences in coefficients by $G$. Characteristics are hold fix at $\bar{x}_{1A}$. Additionally, we observe now two interaction terms, $I1$ and $I2$, accounting for the fact that differences in characteristics and parameters exist simultaneously between the four groups. The interaction effects are the following:

$$I1 = (\bar{x}_{1A} - \bar{x}_{1B})(\hat{\beta}_{1B} - \hat{\beta}_{0B}) \tag{5.11}$$

and

$$I2 = (\bar{x}_{1A} - \bar{x}_{0A})(\hat{\beta}_{0A} - \hat{\beta}_{0B}) \tag{5.12}$$

$I1$ accounts for differences in prices by $G$ given changes in the set of endowments across $Y$.[8] $I2$ catches changes in coefficients over $Y$ given that endowments between $G$ are different.[9]

Despite using the decomposition approach based on the OVB formula, we can compare differences in pay gaps by estimating a system of Seemingly Unrelated Equations (SURE). Using the SURE method allows errors to be correlated across equations and is more efficient. However, we prefer the more intuitive or more familiar interpretation of the method outlined above. Furthermore, the model based on the OVB formula catches otherwise unobserved interaction effects.

### 5.3.3 Robustness of the Method Proposed and Problems of the Standard Approach

The Oaxaca-Blinder decomposition suffers from various problems. In particular, the method is not unique and its components may be unstable when different controls are added to the Mincer-type wage equation. As the Oaxaca-Blinder decomposition is not unique, the choice of the non-discriminatory wage structure matters and the results may change according to the reference category chosen (Reimers, 1983; Cotton, 1988, Neumark, 1988, Oaxaca and Ransom, 1994; Fortin, 2008). Several solutions have been proposed in the literature to solve the so-called index-number problem. Suggestions in the literature consist in estimating a

[8]In the case of the GPG over time, $I1$ catches year-specific effects in endowments given gender-related differences in prices in the ending period. That is assuming that in the ending period differences in prices between men and women persist (compared to the starting period), it accounts for changing endowments of women over time.

[9]In the first case of the empirical application, $I2$ assumes different endowments between women and men in the starting period and asks how coefficients change over time given gender differences in quantities.

pooled wage structure (Neumark, 1988; Oaxaca and Ransom, 1994) or assigning different weights to the two groups (Reimers, 1983; Cotton, 1988). The intercept-shift approach including the group indicator and parameter restrictions, re-writes the decomposition in terms of advantages of men and disadvantages of women (Fortin, 2008). Thereby, the decomposition does no longer depend on the choice of the non-discriminatory wage structure. In the empirical application in Section 5.3.2, we take men and the ending period as base category or non-discriminatory wage structure.[10] Indeed, the standard case of the Oaxaca-Blinder decomposition assumes positive discrimination against women, i.e. it takes men as the non-discriminatory wage structure. For a recent application, see for example Mandel and Semyonov (2014). We can easily change the reference category by imposing different weights across groups (following Cotton, 1988; Reimers, 1983) and show in Appendix 5.A that the standard case of the GPG can be decomposed in the sense of the intercept-shift approach as proposed by Fortin (2008) based on the OVB formula. In the case of a detailed decomposition, the standard Oaxaca-Blinder decomposition varies with the choice of the left-out category of categorical variables included in the estimation. We show the invariance with respect to categorical variables of the decomposition approach based on the OVB formula in Appendix 5.B. The coefficients of the categorical variables are transformed making them invariant to the choice of the (omitted) base category (Gardeazabal and Ugidos, 2004; Fortin, 2008). Moreover, in Appendix 5.C, we show that the decomposition based on the intercept-shift approach holds also for our proposed decomposition of pay gaps between groups $G$ and $Y$. In Appendix 5.D, we show that the critique of Lee (2015) stating that the intercept-shift approach relies on second moments, while first moments should be considered, does not apply to our proposed decomposition approach with gender dummies along with parameter restrictions.[11] We derive the results in the appendices based on the GPG. However, the derived results are not only valid for the case of the GPG but can be applied to a variety of decomposition problems.

## 5.4 Empirical Implementation

In this Section, we consider the change of the GPG over time (case 1) as well as the PPWG between men and women (case 2). By applying our proposed approach, we are able to draw

[10]In the second empirical application, men in the public sector are the non-discriminatory wage structure.
[11]That is the model outlined in Appendix 5.C.

inference on the diverse contributors to the GPG over time.[12] The results from the standard model are also shown for the sake of comparison.

### 5.4.1 Data and Descriptive Statistics

We use the 2014 and 2005 cross-sectional files of the survey PLUS[13] from the Italian Institute for the Development of Vocational Training for Workers (ISFOL). The data was collected jointly with the Italian Ministry of Labor and Social Policy. Special characteristics of the survey are that it provides broad information on the intervieews' working profiles and motivation to work as well as on the demographic and family background of the participants. Data collection is conducted by Computer Assisted Telephone Interviewing (CATI) and the data set is based on subjective measures only.

In 2005, the original sample contains 38,940 observations. In the wave 2014, 54,961 individuals were interviewed. In our analysis, we focus on full-time employees aged 18-64 years. We include only individuals in the sample that work at least 36 hours per week and exclude self-employed workers from the analysis. The sample is further restricted to earnings from the main job only, i.e. from the job that yields the highest income. After dropping observations with missing data on other variables of interest, our sample contains 9,495 positive wage observations in 2005 and 8,423 in 2014. For the analysis of the evolution of the GPG over time, we pool together the two cross sections of 2005 and 2014. For the analysis of the PPWG between men and women, we use the latest release, i.e. the wave of 2014. In 2005, our sample contains 4,778 women (50.3%) and 4,717 men (49.7%). In the 2014-release, 3,828 (45.4%) individuals are female and 4,595 (54.6%) are male. In 2014, 1,799 women (52.8% of total public-sector employment) and 1,607 men (47.2% of total public-sector employment) are occupied in the public sector. Thus, slightly more women than men are employed in the public sector. The OLS estimates are based on the natural logarithm of net hourly wages as dependent variable. The data set includes also a variable for monthly gross earnings. However, 98.0% of all observations contain missing values.[14] Therefore, we prefer to use the monthly-based net income as dependent variable. Table 5.1 and 5.2 report means and standard deviations for some of the variables included in the analysis for the two cases under consideration, respectively. We use the same set of control variables in both

---

[12]In the second case, we draw inference on the components of the PPWG by gender.

[13]Participation, Labor, Unemployment Survey (PLUS)

[14]The survey contains also gross annual earnings. Unfortunately, gross annual earnings divided by the number of months in a calendar year (including a 13$^{\text{th}}$ month), differ by more than 800 Euros (per month) from the reported monthly gross income.

empirical applications of the standard and proposed estimation method. Detailed information on the variables used in the analysis can be found in Appendix 5.E.

**Descriptive Statistics Case 1**

Table 5.1 shows that women have on average higher educational attainment than men and that their human capital increased from 2005 to 2014 (*Schooling*). For men, the increase is less pronounced. Men still outperform women in terms of labor market characteristics (*Exper* and *Tenure*). However, while the average years of experience of women increased over the last decade, men's average years of experience decreased slightly. Nonetheless, the average level of labor market experience is still higher for men than for women in 2014. On average, men hold more often an unlimited contract in both years (*Contract_Type*). The proportion of married women and men reduced slightly over the last decade (*Married*). The share of individuals employed in Northern Italy decreased slightly for both men and women (*North*). In 2014, more females than males are employed in highly specialized occupations, while for the wave of 2005, the opposite holds (*Manager*).

Table 5.1 Descriptive Statistics Case 1

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Women | | | | Men | | | |
| | 2005 | | 2014 | | 2005 | | 2014 | |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. |
| Exper | 16.23 | 11.33 | 17.73 | 12.08 | 20.51 | 12.86 | 20.19 | 12.95 |
| Tenure | 10.42 | 9.822 | 13.52 | 11.46 | 14.10 | 11.70 | 15.41 | 12.34 |
| Schooling | 12.72 | 2.722 | 14.30 | 1.486 | 12.26 | 2.842 | 13.95 | 1.397 |
| Contract_Type | 0.838 | 0.369 | 0.862 | 0.345 | 0.879 | 0.327 | 0.884 | 0.321 |
| Married | 0.591 | 0.492 | 0.580 | 0.494 | 0.580 | 0.494 | 0.577 | 0.494 |
| Italian | 0.989 | 0.103 | 0.987 | 0.115 | 0.994 | 0.0768 | 0.993 | 0.0857 |
| North | 0.533 | 0.499 | 0.502 | 0.500 | 0.463 | 0.499 | 0.480 | 0.500 |
| Centre | 0.205 | 0.404 | 0.223 | 0.416 | 0.183 | 0.387 | 0.211 | 0.408 |
| Manager | 0.118 | 0.323 | 0.247 | 0.431 | 0.136 | 0.343 | 0.232 | 0.422 |
| Intermed_Prof | 0.617 | 0.486 | 0.609 | 0.488 | 0.465 | 0.499 | 0.499 | 0.500 |
| Observations | 4,778 | | 3,828 | | 4,717 | | 4,595 | |

**Descriptive Statistics Case 2**

Table 5.2 shows that the average level of educational attainment is higher in the public compared to the private sector. Women have on average higher educational attainment than men in both sectors. Female civil servants are even better educated than their female colleagues in the private sector. Similarly, men in the public sector have higher educational performance compared to their male peers in the private sector. Men outperform women in both sectors in terms of labor market presence and job tenure. About the equal amount of male and female employees is married, yet, the proportion of married employees is higher in the public sector. In the public sector, men and women are more often employed in highly specialized jobs. The proportion of highly specialized females in public employment is higher than that of males.

Table 5.2 Descriptive Statistics Case 2

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Women | | | | Men | | | |
| | Private Sector | | Public Sector | | Private Sector | | Public Sector | |
| Variables | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. | Mean | Std.Dev. |
| Exper | 14.09 | 10.65 | 21.83 | 12.29 | 17.69 | 12.57 | 24.84 | 12.35 |
| Tenure | 9.766 | 9.442 | 17.75 | 12.04 | 12.57 | 11.47 | 20.68 | 12.18 |
| Schooling | 14.13 | 1.454 | 14.48 | 1.500 | 13.79 | 1.320 | 14.26 | 1.481 |
| Contract_Type | 0.819 | 0.385 | 0.911 | 0.286 | 0.859 | 0.348 | 0.928 | 0.258 |
| Married | 0.471 | 0.499 | 0.703 | 0.457 | 0.495 | 0.500 | 0.730 | 0.444 |
| Italian | 0.978 | 0.147 | 0.997 | 0.0577 | 0.991 | 0.0964 | 0.996 | 0.0610 |
| North | 0.555 | 0.497 | 0.442 | 0.497 | 0.553 | 0.497 | 0.343 | 0.475 |
| Centre | 0.218 | 0.413 | 0.228 | 0.420 | 0.210 | 0.407 | 0.214 | 0.410 |
| Manager | 0.140 | 0.347 | 0.367 | 0.482 | 0.180 | 0.384 | 0.327 | 0.469 |
| Intermed_Prof | 0.646 | 0.478 | 0.569 | 0.495 | 0.498 | 0.500 | 0.502 | 0.500 |
| Observations | 2,029 | | 1,799 | | 2,988 | | 1,607 | |

## 5.4.2 Empirical Results

We first present the decomposition results from the standard Oaxaca-Blinder approach and discuss the conclusions drawn on the change of the wage gap in this framework. Next, we apply the method derived in Section 5.3 in order to directly estimate changes of the wage gaps and in order to draw inference on the diverse contributors to the change of the gap.

### The Gender Pay Gap over Time

A general finding in the literature is that the gap in pay by gender was reduced over time (Blau and Kahn, 2006; Goldin, 2014; Mandel and Semyonov, 2014). The part attributable to observed characteristics and therefore referred to as explained component increased, while the unexplained part, i.e. the component due to differences in returns to wage-related characteristics and differences in the intercepts, decreased.

Indeed, by applying the traditional approach to our data, we also find a reduction of the GPG in hourly wages over time; 12.4% in 2005 and 9.5% in 2014.[15] Table 5.3 shows that the gaps are highly statistically significant in either case. The composition of the gap also changed across the decade. In 2005, the explained component does not play a role in determining the GPG (as it is not statistically significant), while in 2014, the endowments part becomes highly statistically significant and contributes to a narrowing of the GPG (negative term). Differences in the unexplained component are statistically significant in both years. The component in 2014 decreased slightly (86.2% in 2005 versus 84.3% in 2014). A relatively small decrease in the unexplained component of the GPG in 2014 is in line with results of other scholars (e.g. Fortin, 2008; Mandel and Semyonov, 2014). In 2005, differences in endowments and coefficients that exist simultaneously between men and women have a statistically significant impact as well, what is no longer the case in 2014.

All in all, our data delivers results in line with the literature, when applying the standard estimation methodology. The GPG declined over the last decade, differences in endowments (in favor of women) have become statistically significant in 2014 and the part of the GPG due to differences in prices has declined.

[15]The estimated GPGs in this paper are larger than the pay gaps found by Eurostat (2017). Eurostat (2017) finds wage gaps amounting to 4.4% in 2006 (missing in 2005) and 6.1% in 2014 for Italy. These relatively larger gaps are due to our sample restriction of considering only employees working at least 36 hours per week.

Table 5.3 Standard Decomposition of the GPG in 2005 and 2014

| | (1) | (2) |
|---|---|---|
| Variables | 2005 | 2014 |
| *Differential* | | |
| $\overline{ln(w_M)}$ | 1.999*** | 2.134*** |
| | (0.006) | (0.007) |
| $\overline{ln(w_F)}$ | 1.875*** | 2.039*** |
| | (0.006) | (0.007) |
| Difference | 0.124*** | 0.095*** |
| | (0.008) | (0.009) |
| *Decomposition* | | |
| Endowments | 0.008 | -0.016*** |
| | (0.006) | (0.006) |
| Coefficients | 0.107*** | 0.107*** |
| | (0.008) | (0.009) |
| Interaction | 0.009* | 0.004 |
| | (0.006) | (0.006) |
| *%–Contribution* | | |
| Endowments | 6.5 | 12.6 |
| Coefficients | 86.2 | 84.3 |
| Interaction | 7.3 | 3.1 |
| Observations | 9,495 | 8,423 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

*Notes:* For the GPG in 2014, the %–contribution for the endowments effect is $\frac{|0.016|}{(|0.016|+0.107+0.004)} \times 100$.

Next, we directly estimate the change of the GPG between 2014 and 2005 and decompose that change in explained and unexplained components as well as interaction effects. Table 5.4, column (1), shows the base model of case 1. The coefficient estimate of *femyear* shows the change of the GPG from 2014 to 2005.[16] The difference between the GPG in 2014 and 2005 amounts to $-0.03$ log points and is statistically significant. Given the negative sign, the GPG has decreased over time. The magnitude as well as the sign of the change is also visible by

[16]That is *femyear* is the interaction term of the two group dummies *female* and *year*.

looking at the aggregate GPGs from the outcome of the standard estimation in Table 5.3. However, now we can also conclude that this reduction in the GPG is statistically significant. The full model is presented in column (2) of Table 5.4. We immediately see that the part of the price effect due to differences in the intercepts, $\hat{\alpha}_1^{full}$, is not statistically significant. Similarly, the effect of being a woman or in year 2005, all else equal, becomes statistically insignificant. The remaining coefficient estimates show the expected signs.[17]

Table 5.4 OLS Estimates of Log Hourly Wages – Case 1, Base and Full Specification

| | (1) | (2) |
|---|---|---|
| | Basic Specification | Full Specification |
| femyear | -0.028** | -0.051 |
| | (0.012) | (0.185) |
| female | -0.095*** | -0.148 |
| | (0.009) | (0.152) |
| year | -0.135*** | -0.010 |
| | (0.009) | (0.130) |
| *Groups of Covariates* | | |
| Labor Market Presence | No | Yes |
| Educational Attainment | No | Yes |
| Job Characteristics | No | Yes |
| Demographic and Family Background Characteristics | No | Yes |
| Industrial and Occupational Dummies | No | Yes |
| Interaction Terms | No | Yes |
| Observations | 17,918 | 17,918 |
| R-squared | 0.050 | 0.291 |

Robust standard errors in parentheses
$***p < 0.01, **p < 0.05, *p < 0.1$

Table 5.5 presents the results from our proposed decomposition. The results show that the change of the GPG is only explained by the quantity effect. The change of the GPG over time is explained by changes in observed characteristics between men and women (in favor of women) over time. We know from Table 5.1 that women's set of observable human capital and labor market characteristics (*Schooling*, *Exper*) is increasing over the last decade, while that of men is partly even decreasing (*Exper*) or remained lower than that of women (educational attainment). In fact, in educational matters, women have outpaced men (Goldin, 2006). The results from the standard method suggest that the coefficients part of the GPGs, i.e. the part due to differences in returns on observable characteristics, was a main contributor to the GPG in either year with decreasing importance in the ending

[17] The full regression output is shown in Table 5.F.1 in Appendix 5.F.

period. However, by estimating the difference of the GPG over time directly, we see that this so-called discriminatory part has not significantly changed over the last ten years in Italy. The decomposition shows that the only factor that contributes statistically significantly to the narrowing of the gap are better observable characteristics of women. Hence, the closing of the GPG is not explained by anti-discrimination laws, changes in attitudes towards women in the labor market or changes in the family structure and birth control (Fortin, 2015). The latter is, apart from the unexplained part ($U$), caught by the interaction effects accounting for simultaneous differences in endowments over time and changing prices between men and women ($I1$) as well as variation in the set of endowments by gender and changing prices over time ($I2$). The components account for the effects of changes in institutional settings or attributes towards women on prices (given differences in endowments). Yet, the effects are not statistically significant.

Table 5.5 Decomposition of the Change in the GPG over Time – Case 1

| | (1) |
|---|---|
| | Pooled Sample (2005 and 2014) |
| *Decomposition* | |
| E | -0.023*** |
| | (0.007) |
| I1 | 0.002 |
| | (0.013) |
| I2 | -0.006 |
| | (0.006) |
| u | 0.050 |
| | (0.179) |
| Total = E + I1 + I2 + u | 0.023 |
| | (0.185) |
| Observations | 17,918 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

### The Public-Private Sector Wage Gap between Men and Women

In the literature, a positive wage gap between the public and the private sector is found (Lucifora and Meurs, 2005; Melly, 2005; Christodfides and Michael, 2006; Arulampalam et al. 2007; Mandel and Semyonov, 2014). Table 5.6 shows that also in our data for Italy, we find differences in earnings by sector, with higher wage levels in the public sector. A general result is that women are better-off in the public compared to the private sector, while for men the public-sector premia is less important (e.g. Melly, 2005). We find different PPWGs by gender as well; 23.2% for women and 19.8% for men (see Table 5.6). Both gaps are found to be highly statistically significant. Also, the composition of the PPWGs differs by gender. For women, the PPWG is mainly due to the unexplained part (54.3%). In contrast, for men, the endowments effect is the main driver of the pay gap (59.9%). Interaction effects are rather small but more important for the wage gap in the female subsample (15.5% compared to 6.1% in the male subsample).

Table 5.6 Standard Decomposition of the PPWG for Women and Men in 2014

| | (1) | (2) |
|---|---|---|
| | Women | Men |
| *Differential* | | |
| $\overline{ln(w_{\text{Public\_Sector}})}$ | 2.162*** | 2.263*** |
| | (0.009) | (0.011) |
| $\overline{ln(w_{\text{Private\_Sector}})}$ | 1.930*** | 2.065*** |
| | (0.010) | (0.008) |
| Difference | 0.232*** | 0.198*** |
| | (0.013) | (0.013) |
| *Decomposition* | | |
| Endowments | 0.070*** | 0.118*** |
| | (0.015) | (0.015) |
| Coefficients | 0.126*** | 0.067*** |
| | (0.016) | (0.024) |
| Interaction | 0.036** | 0.012 |
| | (0.018) | (0.023) |
| *%–Contribution* | | |
| Endowments | 30.2 | 59.9 |
| Coefficients | 54.3 | 34.0 |
| Interaction | 15.5 | 6.1 |
| Observations | 3,828 | 4,595 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

The decomposition outcome of the PPWG between men and women using our proposed model is provided in Tables 5.7–5.8. The results from the base model suggest that there is a positive and statistically significant difference in the PPWG between men and women equal to $-0.03$ log points.[18] The dummy variable for working in the private sector (*private*) negative and significant, tells us that there is a wage loss for working in the private sector compared to public-sector employment. As expected, the coefficient on the $female$-dummy shows that being a women has a significant and negative impact on labor income. In the

[18]Indicated by the interaction of the dummies *female* and *private*; *fempriv*.

full model, the effect of private-sector employment as well as being female on wages turns statistically insignificant. Yet, the interaction term *fempriv*, is statistically significant and strongly negative ($-0.72$). Hence, $\hat{\alpha}_1^{full}$, i.e. the part of the price or unexplained effect due to differences in the starting points is statistically significant. This implies that there is a premia for simply working in the public sector and that this premia is higher for women than for men. Again, the remaining parameter estimates impact on wages as expected.[19]

By looking at the decomposition, we find that the difference in observable characteristics across sectors and gender, $E$, does play a statistically significant role in explaining the difference of the PPWG between men and women. In particular, the explained component drives the negative PPWG by gender as best-educated females are more often located in the public sector (Bordogna, 2012; Piazzalunga and Di Tommaso, 2015). The difference in the unexplained component, $u$, of the PPWG between men and women is significant as well and shows that the change works towards a positive PPWG between men and women. This implies that more egalitarian pay schemes in the public sector are ruled out by female discrimination in prices in both sectors. Moreover, we observe simultaneously differences in characteristics between women and men as well as difference in coefficients between the private and the public sector (for men; $I1$). Hence, more favorable endowments of men in the private sector compared to women in the private sector and higher pay schemes in the public sector narrow the (negative) PPWG between men and women. All in all, for case 2, the conclusions drawn from the standard estimation are confirmed; both quantity and price effects contribute to the difference in the PPWG between men and women. Yet, we gain the additional insight that the set-up or organization of the public sector does play a role as well. That is institutional norms of the public sector being relatively more gender-equal in combination with more discriminatory practices against women in the private sector lead to an increase of the significant difference in the PPWG between men and women in 2014 in Italy.

[19]The complete regression outcome of the full model is shown in Table 5.F.2 in Appendix 5.F.

Table 5.7 OLS Estimates of Log Hourly Wages – Case 2, Base and Full Specification

| | (1) | (2) |
|---|---|---|
| | Basic Specification | Full Specification |
| fempriv | -0.034* | -0.724** |
| | (0.019) | (0.289) |
| female | -0.101*** | 0.278 |
| | (0.014) | (0.196) |
| private | -0.198*** | 0.309 |
| | (0.013) | (0.205) |
| *Groups of Covariates* | | |
| Labor Market Presence | No | Yes |
| Educational Attainment | No | Yes |
| Job Characteristics | No | Yes |
| Demographic and Family Background Characteristics | No | Yes |
| Industrial and Occupational Dummies | No | Yes |
| Interaction Terms | No | Yes |
| Observations | 8,423 | 8,423 |
| R-squared | 0.069 | 0.236 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 5.8 Decomposition of the Change in the PPWG by Gender – Case 2

| | (1) |
|---|---|
| | Pooled Sample (Women and Men) |
| *Decomposition* | |
| E | -0.028*** |
| | (0.011) |
| I1 | 0.041* |
| | (0.022) |
| I2 | 0.002 |
| | (0.011) |
| u | 0.675* |
| | (0.357) |
| Total = E + I1 + I2 + u | 0.689* |
| | (0.360) |
| Observations | 8,423 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

## 5.5 Conclusion

Adding to the discussion of the convergence of the GPG over time and the persistence of a PPWG between men and women, we propose an alternative decomposition method allowing to draw inference on the difference of two wage gaps on aggregate as well as on its components. The model set-up bases on the OVB formula and the Gelbach decomposition. Despite additional insights on the composition of differences in gaps, the method can be made robust to the choice of the reference category (Reimers, 1983; Cotton, 1988, Neumark, 1988, Oaxaca and Ransom, 1994; Fortin, 2008) as well as to the indeterminacy problem (Lee, 2015). The method proposed can be applied to a variety of cases such as differences in the GPG and its drivers over time, across countries, sectors, occupations or unions. We empirically consider two cases; the change of the GPG over time as well as the PPWG between men and women in Italy.

The observed closing of the GPG over time is heavily discussed in the literature and the determination of the reasons of the narrowing is of huge interest, especially with regard to policy implications (Blau and Kahn, 2006; 2016; Goldin, 2014). Similarly, the PPWG that is found to differ for men and women is a topic of on-going research (Melly, 2005). Yet, up to know, in the standard estimation framework, direct inference on the difference of pay gaps and changes in their components could not be drawn. Conclusions were rather drawn by estimating the pay gaps separately in different subsamples and comparing the results ex post. In this way, it is not possible to test the significance of the change in the estimated pay gaps on aggregate or the components of the decomposition. Besides the estimation of the change of the GPG over time on aggregate as well as of the explained and unexplained component, our method also catches otherwise unobserved interaction effects across the respective groups of interest.

We find a significant convergence of the GPG over the last decade in Italy. The convergence of the GPG over time was found to be only explained by a reduction in differences in observable characteristics by gender. In contrast, by estimating the GPG separately for 2005 and 2014, i.e. following the standard approach in the literature, the relative decline in the contribution of the price component to the wage gap might have led to the conclusion that the implementation of anti-discrimination laws and changing attitudes towards women in the labor market have influenced the narrowing of the pay gap over time as well. Yet, these policies as well as changes in social norms seem to have been less effective than expected a priori. Thereby, we add to the literature on the convergence of the GPG over time for the case of Italy the finding that the closing of the pay differential by gender over the last decade was only due to the catching-up of women in terms of endowments. The results for the second case we have examined, i.e. the PPWG between men and women, point the attention to differences in the structure of the public and private sector, which are found to be important to explain the differential. Better educated females are more often employed in the public sector given more egalitarian pay schemes as well as job stability (Bordogna, 2012; Piazzalunga and Di Tommaso, 2015). In this case, the results derived from the standard approach concerning the explained and unexplained part are confirmed in the sense that both components contribute significantly to the change of the PPWG between men and women.

All in all, the analysis with the proposed decomposition method offers a better understanding of what has led to the narrowing of the GPG in the last ten years and what drives the difference in the PPWG between men and women. Most importantly, we can infer what drives the difference in the respective pay gaps in a statistically significant manner. The model proposed offers an intuitive approach to directly estimate changes in wage gaps between groups and can be applied to various problems.

# Appendices

The robustness of the decomposition is for simplicity shown for the case of the GPG. Deriving the robust model based on the GPG allows also for a better comparison of the method with the approaches in the literature (e.g. Fortin, 2008, uses the case of the GPG).[20] In Appendix 5.C, when considering differences of gaps, we derive the model for the GPG changing over time. Notably, the methods can be applied to various other decomposition problems.

## Appendix 5.A Solving the Index-Number Problem of Decompositions using the Intercept-Shift Approach

As is well known in the literature, the Oaxaca-Blinder decomposition is not unique. Therefore, the choice of the non-discriminatory wage structure (men or women) matters and leads to different results (Cotton, 1988; Oaxaca and Ransom, 1994). Several approaches have been proposed to circumvent this problem (Reimers, 1983; Cotton, 1988, Neumark, 1988, Oaxaca and Ransom, 1994; Fortin, 2008). We extend the method proposed by Gelbach (2016) in order to have a wage decomposition invariant to the reference category adopted. In particular, we adopt the decomposition proposed by Fortin (2008) that includes gender intercept shifts along with an identification restriction in the regression of females and males pooled together, when considering the standard case of the GPG for individual $i$:

$$ln(w_i) = \gamma_0 + \gamma_{0F} F_i + \gamma_{0M} M_i + X_i \gamma + \varepsilon_i$$

subject to:

$$\gamma_{0F} + \gamma_{0M} = 0$$

where $F_i$ is equal to one if the individual is female and zero otherwise and $M_i$ equals one if the individual is male and zero otherwise, i.e. $F_i = (1 - M_i)$. Correspondingly, the index $F$ identifies women and the index $M$ identifies men. For the pooled regression with male and

[20] The derived model is robust to the index-number problem and invariant with respect to categorical variables as well as robust to the indeterminacy problem.

female dummies, respectively, evaluated at the mean, we have:

$$\overline{ln(w_M)} = \hat{\gamma}_0 + \hat{\gamma}_{0M} M + \bar{X}_M \hat{\gamma}$$
$$\overline{ln(w_F)} = \hat{\gamma}_0 + \hat{\gamma}_{0F} F + \bar{X}_F \hat{\gamma}$$

The identification restriction imposes that the pooled wage equation truly represents a non-discriminatory wage structure, i.e. a wage structure, where the advantage of men is equal to the disadvantage of women:

$$\overline{ln(w_M)} - \overline{ln(w_F)} = (\bar{X}_M - \bar{X}_F)\hat{\gamma} + (\hat{\gamma}_{0M} - \hat{\gamma}_{0F})$$

The first component on the right-hand side, $(\bar{X}_M - \bar{X}_F)\hat{\gamma}$, is the explained part, while $\hat{\gamma}_{0M}$ and $\hat{\gamma}_{0F}$ are the *advantage of men* and the *disadvantage of women*, respectively. In particular, from the difference of the wage regression separately for men and women and the pooled wage regression with a gender dummy, we have:

$$\hat{\gamma}_{0M} = \bar{X}_M(\hat{\beta}_M - \hat{\gamma}) + (\hat{\beta}_{0M} - \hat{\gamma}_0) \quad \textit{advantage of men}$$
$$\hat{\gamma}_{0F} = \bar{X}_F(\hat{\beta}_F - \hat{\gamma}) + (\hat{\beta}_{0F} - \hat{\gamma}_0) \quad \textit{disadvantage of women}$$

where $\hat{\beta}_{0M}, \hat{\beta}_{0F}$ are the intercepts and $\hat{\beta}_M, \hat{\beta}_F$ are the estimated coefficients of wage equations estimated separately for men and women:

$$ln(w_{iM}) = \beta_{0M} + X_{iM}\beta_M + \varepsilon_{iM} \tag{5.A.1}$$
$$ln(w_{iF}) = \beta_{0F} + X_{iF}\beta_F + \varepsilon_{iF} \tag{5.A.2}$$

In order to adopt the above wage decomposition within the conditional decomposition framework proposed by Gelbach (2016), we estimate the following wage equation:

$$ln(w_i) = \gamma_0 + \gamma_{0F}F_i + \gamma_{0M}M_i + X_i\gamma + X_iF_i\gamma_{XF} + X_iM_i\gamma_{XM} + \nu_i \tag{5.A.3}$$

subject to:

$$\begin{aligned} \gamma_{0F} + \gamma_{0M} &= 0 \\ \gamma_{X_kF} + \gamma_{X_kM} &= 0 \quad \text{for } k = 1 \dots K \end{aligned}$$

where $\gamma_{X_kF}$ and $\gamma_{X_kM}$ are the parameters of the interaction term between the $k$th regressor $X_i$ and the dummy $F_i$ and $M_i$, respectively. The error term is represented by $v_i$. Then,

$$\begin{aligned}\overline{ln(w_M)} &= \hat{\gamma}_0 + \hat{\gamma}_{0M} + \bar{X}_M\hat{\gamma} + \bar{X}_M\hat{\gamma}_{XM} \\ \overline{ln(w_F)} &= \hat{\gamma}_0 + \hat{\gamma}_{0F} + \bar{X}_F\hat{\gamma} + \bar{X}_F\hat{\gamma}_{XF}\end{aligned}$$

Consequently, the GPG becomes:

$$\begin{aligned}\overline{ln(w_M)} - \overline{ln(w_F)} &= (\hat{\gamma}_{0M} - \hat{\gamma}_{0F}) + (\bar{X}_M - \bar{X}_F)\hat{\gamma} + \bar{X}_M\hat{\gamma}_{XM} - \bar{X}_F\hat{\gamma}_{XF} \\ &= -2\hat{\gamma}_{0F} + (\bar{X}_M - \bar{X}_F)\hat{\gamma} - (\bar{X}_M + \bar{X}_F)\hat{\gamma}_{XF} \qquad (5.A.4)\end{aligned}$$

First, we observe that it can be easily shown that there exists the following relationship between the parameter estimates of equations (5.A.1)-(5.A.2) and (5.A.3):

$$\begin{aligned}\hat{\gamma} + \hat{\gamma}_{XF} &= \hat{\beta}_F \\ \hat{\gamma}_0 + \hat{\gamma}_{0F} &= \hat{\beta}_{0F} \\ \hat{\gamma} - \hat{\gamma}_{XF} &= \hat{\beta}_M \\ \hat{\gamma}_0 - \hat{\gamma}_{0F} &= \hat{\beta}_{0M}\end{aligned}$$

Therefore, the GPG of (5.A.4) can be re-written in terms of the Fortin decomposition as:

$$\begin{aligned}\overline{ln(w_M)} - \overline{ln(w_F)} &= (\hat{\beta}_{0M} - \hat{\gamma}_0) - (\hat{\beta}_{0F} - \hat{\gamma}_0) + (\bar{X}_M - \bar{X}_F)\hat{\gamma} + \bar{X}_M(\hat{\beta}_M - \hat{\gamma}) - \bar{X}_F(\hat{\beta}_F - \hat{\gamma}) \\ &= (\bar{X}_M - \bar{X}_F)\hat{\gamma} + [\bar{X}_M(\hat{\beta}_M - \hat{\gamma}) + (\hat{\beta}_{0M} - \hat{\gamma}_0)] - [\bar{X}_F(\hat{\beta}_F - \hat{\gamma}) + (\hat{\beta}_{0F} - \hat{\gamma}_0)] \\ & \qquad (5.A.5)\end{aligned}$$

Second, the estimation can be recast in terms of the sequential decomposition of Gelbach by considering the following base model for individual $i$:

$$ln(w_i) = \gamma_0^{base} + (F_i - M_i)\gamma_{0F}^{base} + \varepsilon_i^{base} \qquad (5.A.6)$$

where the $1 \times 2$ vector of regressors $X_{i1}$ of the base specification contains for each observation $i$ a constant and the difference between the two dummy variables $F_i$ and $M_i$, $(F_i - M_i)$. The full model is defined as follows:

$$ln(w_i) = \gamma_0^{full} + (F_i - M_i)\gamma_{0F}^{full} + X_i\gamma + (F_i - M_i)X_i\gamma_{XF} + \varepsilon_i^{full} \qquad (5.A.7)$$

where the regressors $X_i$ as well as the interaction between $X_i$ and the difference between the two dummy variables $F_i$ and $M_i$ are contained in the $1 \times 2K$ vector $X_{i2}$. The regressors in $X_{i2}$ are the omitted variables. By the OVB formula the following relationship holds (for the model evaluated at the mean):

$$\begin{bmatrix} \hat{\gamma}_0^{base} \\ \hat{\gamma}_{0F}^{base} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_0^{full} \\ \hat{\gamma}_{0F}^{full} \end{bmatrix} + (X_1'X_1)^{-1}X_1'X_2 \begin{bmatrix} \hat{\gamma} \\ \hat{\gamma}_{XF} \end{bmatrix} \tag{5.A.8}$$

where $\left[\hat{\gamma}_0^{base}, \hat{\gamma}_{0F}^{base}\right]'$ is the $2 \times 1$ vector of coefficient estimates of $X_{i1}$ from the base model (5.A.6) evaluated at the mean; $\left[\hat{\gamma}_0^{full}, \hat{\gamma}_{0F}^{full}\right]'$ is the $2 \times 1$ vector containing the coefficient estimates of $X_{i1}$ from the full model (5.A.7) evaluated at the mean and $\left[\hat{\gamma}, \hat{\gamma}_{XF}\right]'$ is the vector of coefficients estimates of $X_{i2}$ from the full model (5.A.7) evaluated at the mean, i.e. $\hat{\gamma}^{full}$ with dimension $2K \times 1$. First observe that $\hat{\gamma}_{0F}^{base}$ is equal to $\frac{\overline{ln(w_F)} - \overline{ln(w_M)}}{2}$ and that $\hat{\gamma}_{0F}^{full}$ is equal to $\frac{\hat{\beta}_{0F} - \hat{\beta}_{0M}}{2}$. As in Section 5.3, we are interested in the second row of equation (5.A.8). Given the relationship in (5.A.8), we observe that:

$$\hat{\gamma}_{0F}^{base} = -\frac{\overline{ln(w_M)} - \overline{ln(w_F)}}{2} = -\frac{\Delta}{2} = \hat{\gamma}_{0F}^{full} + \eta \hat{\gamma}^{full} \tag{5.A.9}$$

where $\Delta$ is the GPG and $\eta = \left[\frac{(\bar{x}_F - \bar{x}_M)}{2}, \frac{(\bar{x}_F + \bar{x}_M)}{2}\right]$ contains the sample means of observable characteristics in $X_i$ obtained from the linear projection of $X_i$ and $(F_i - M_i)X_i$ with respect to $X_{i1}$ (at the mean). The row vector $\eta$ has dimension $1 \times 2K$. Moreover, we have $\hat{\gamma}_{0F}^{full} = \frac{\hat{\beta}_{0F} - \hat{\beta}_{0M}}{2} = \frac{(\hat{\beta}_{0F} - \hat{\gamma}_0^{full}) - (\hat{\beta}_{0M} - \hat{\gamma}_0^{full})}{2}$. Consequently, the GPG can be written as:

$$-2\gamma_{0F}^{base} = -2\gamma_{0F}^{full} + (\bar{x}_M - \bar{x}_F)\gamma - (\bar{x}_M - \bar{x}_F)\gamma_{XF} \tag{5.A.10}$$

what completes the proof of decomposition equivalence.

# Appendix 5.B Invariance Decomposition with respect to Categorical Variables

A second type of identification issue arises when dummy variables are considered in a detailed wage decomposition. Oaxaca and Ransom (1999) show that the assignment of the explained part of the GPG to specific variables is not invariant to the choice of reference groups. This problem can be easily solved by imposing the following parameter restrictions as proposed by Gardeazabal and Ugidos (2004), Yun (2005) and Fortin (2008):

$$\sum_{j=1}^{C_k} \gamma_{jk} = 0, \; k \in C \tag{5.B.1}$$

where $C$ denotes the set of categorical variables, and $C_k$ the number of categories for variable $k$. The neutral, i.e. non-sensitive to any left-out category, Oaxaca-Blinder decomposition follows. The zero-sum restriction (5.B.1) is applied to the wage equation, when female and male wages are estimated separately as well as to the pooled regression with gender dummies. The latter is additionally estimated with the identification restriction $\gamma_{0M} + \gamma_{0F} = 0$ on the gender parameters. Thereby, the intercepts, $\beta_{0M}$, $\beta_{0F}$ and $\gamma_0$, are no longer influenced by the choice of the reference category in the case of categorical variables.

The restriction (5.B.1) can also be applied to the method proposed in Section 5.3 leading to indicator variables that are invariant to the choice of the left-out category in the case of categorical variables.

## Appendix 5.C Estimating Differences of Gaps with the Intercept-Shift Approach

The extension of the decomposition described in Appendix 5.A to the case of the estimation of the difference of wage gaps follows straightforward. We consider, as in Section 5.3, the indicator variable $Y_i$ that takes values $\{0,1\}$. Again, when the indicator variable $Y_i$ is used as an index ($Y$), $Y_i = 0$ corresponds to $B$ and $Y_i = 1$ to $A$. Similarly, in order to circumvent confusion with the intercept (referred to as $\beta_0$ in coherence with Appendix 5.A), the gender index is not numerical here, but $G \in \{F,M\}$ with $F =$ female and $M =$ male replacing the numerical index $\{1,0\}$, respectively. The set of regressors considered in Section 5.3.2 are hence transformed as follows:

$$\begin{aligned} X_{i1} &= [1,(F_i - M_i)Y_i,(F_i - M_i),Y_i] \\ X_{i2} &= [X,(F_i - M_i)X_i,Y_iX_i,(F_i - M_i)Y_iX_i] \end{aligned}$$

for each individual $i$, with $X_{i1}$ having dimension $1 \times 4$ and $X_{i2}$ having dimension $1 \times 4K$. $X_{i1}$ contains the interaction of $(F_i - M_i)$ with $Y_i$; $(F_i - M_i)Y_i$. The second set of regressors, $X_{i2}$ contains the $1 \times K$ vector of characteristics $X_i$ as well as the interaction of $X_i$ with $(F_i - M_i)$

and $Y_i$; $(F_i - M_i)X_i, Y_iX_i$ and $(F_i - M_i)Y_iX_i$, respectively. The base model is then:

$$ln(w_i) = \gamma_0^{base} + (F_i - M_i)Y_i\gamma_{FY}^{base} + (F_i - M_i)\gamma_F^{base} + Y_i\gamma_Y^{base} + \varepsilon_i^{base} \tag{5.C.1}$$

while the full model is defined as follows:

$$\begin{aligned} ln(w_i) = {} & \gamma_0^{full} + (F_i - M_i)Y_i\gamma_{F_iY_i}^{full} + (F_i - M_i)\gamma_F^{full} + Y_i\gamma_Y^{full} \\ & + X_i\gamma + (F_i - M_i)X_i\gamma_{XF} + Y_iX_i\gamma_{XY} + (F_i - M_i)Y_iX_i\gamma_{XYF} + \varepsilon_i^{full} \end{aligned} \tag{5.C.2}$$

where $\gamma_0^{base}$ is the constant and $\gamma_{FY}^{base}, \gamma_F^{base}, \gamma_Y^{base}$ are the coefficients of the the base model (5.C.1), $\gamma_0^{full}, \gamma_{FY}^{full}, \gamma_F^{full}, \gamma_Y^{full}$ are the corresponding constant and coefficients of $X_{i1}$ from the full model (5.C.2). $\gamma, \gamma_{XF}, \gamma_{XY}, \gamma_{XYF}$ are the $K \times 1$ coefficient vectors of $X_{i2}$ from the full model (5.C.2). The second row of the linear projection of $X_{i2}$ with respect to $X_{i1}$ at the mean is contained in the following $1 \times 4K$ vector:

$$\zeta = \left[\frac{(\bar{x}_{0A} - \bar{x}_{1A}) - (\bar{x}_{0B} - \bar{x}_{1B})}{2}, \frac{(\bar{x}_{0A} + \bar{x}_{1A}) - (\bar{x}_{0B} + \bar{x}_{1B})}{2}, \frac{(\bar{x}_{1A} - \bar{x}_{0A})}{2}, \frac{(\bar{x}_{1A} + \bar{x}_{0A})}{2}\right]$$

Consider the equivalence between the following parameter estimates evaluated at the mean:

$$\begin{aligned} \hat{\gamma}_0^{full} - \hat{\gamma}_{FY}^{full} - \hat{\gamma}_F^{full} + \hat{\gamma}_Y^{full} &= \hat{\beta}_{0,MA} \\ \hat{\gamma}_0^{full} + \hat{\gamma}_{FY}^{full} + \hat{\gamma}_F^{full} + \gamma_Y^{full} &= \hat{\beta}_{0,FA} \\ \hat{\gamma}_0^{full} + \hat{\gamma}_F^{full} &= \hat{\beta}_{0,FB} \\ \hat{\gamma}_0^{full} - \hat{\gamma}_F^{full} &= \hat{\beta}_{0,MB} \\ \hat{\gamma} + \hat{\gamma}_{XF} + \hat{\gamma}_{XY} + \hat{\gamma}_{XYF} &= \hat{\beta}_{FA} \\ \hat{\gamma} - \hat{\gamma}_{XF} + \hat{\gamma}_{XY} - \hat{\gamma}_{XYF} &= \hat{\beta}_{MA} \\ \hat{\gamma} + \hat{\gamma}_{XF} &= \hat{\beta}_{FB} \\ \hat{\gamma} - \hat{\gamma}_{XF} &= \hat{\beta}_{MB} \end{aligned}$$

Observe that $\hat{\gamma}_{FY}^{base}$ is equal to $\frac{\Delta GPG}{2}$ and $\hat{\gamma}_{FY}^{full}$ is equal to $\frac{(\hat{\beta}_{0,MB} - \hat{\beta}_{0,FB}) - (\hat{\beta}_{0,MA} - \hat{\beta}_{0,FA})}{2}$. Given the fact that

$$\begin{aligned} \hat{\gamma}_{FY}^{base} &= \frac{\left(\overline{ln(w_{MB})} - \overline{ln(w_{FB})}\right) - \left(\overline{ln(w_{MA})} - \overline{ln(w_{FA})}\right)}{2} \\ &= \frac{\Delta GPG}{2} \end{aligned}$$

The relationship:

$$\hat{\gamma}_{FY}^{base} = \hat{\gamma}_{FY}^{full} + \zeta \hat{\gamma}^{full}$$

can be re-written in terms of the $\Delta GPG$ as:

$$\begin{aligned}
2\hat{\gamma}_{FY}^{base} &= \Delta GPG = \\
&= \underbrace{[(\hat{\beta}_{0,MB} - \hat{\beta}_{0,FB}) - (\hat{\beta}_{0,MA} - \hat{\beta}_{0,FA})]}_{\hat{\gamma}_{FY}^{full}} + \underbrace{(\Delta\bar{x}^{B} - \Delta\bar{x}^{A})\hat{\gamma}}_{\Lambda} \\
&+ \underbrace{(\sum \bar{x}^{A} - \sum \bar{x}^{B})\hat{\gamma}_{XF}}_{\Omega} - \underbrace{\Delta\bar{x}^{A}\hat{\gamma}_{XY}}_{\Theta} + \underbrace{\sum \bar{x}^{A}\hat{\gamma}_{XYF}}_{\Upsilon}
\end{aligned}$$

where $\Delta\bar{x}^{Y}$ is the difference between the average level of observed characteristics of men and women in a certain year, with $Y \in \{A, B\}$ and $\sum \bar{x}^{Y}$ represents the sum of observable labor market characteristics present for men and women in $Y$. Recall that the model can be re-written in terms of the OVB formula as follows:

$$\begin{aligned}
2\hat{\gamma}_{FY}^{base} &= \hat{\gamma}_{FY}^{full} + \hat{\delta}^{\Lambda} + \hat{\delta}^{\Omega} + \hat{\delta}^{\Theta} + \hat{\delta}^{\Upsilon} \\
\hat{P} + \hat{Q} &= \hat{\gamma}_{FY}^{full} + \hat{\delta}^{\Lambda} + \hat{\delta}^{\Omega} + \hat{\delta}^{\Theta} + \hat{\delta}^{\Upsilon}
\end{aligned}$$

with $P$ accounting for the price effect and $Q$ for the quantity effect. In particular,

$$\begin{aligned}
\hat{P} &= \hat{\gamma}_{FY}^{full} + \Upsilon \\
\hat{Q} &= \Omega + \underbrace{\Theta}_{Y\text{-specific term}} + \underbrace{\Lambda}_{\text{gender-specific term}}
\end{aligned}$$

$\hat{\gamma}_{FY}^{full}$ represents the change in the disadvantage of women over time. Thereby, accounting for the relative improvement (or deterioration) of women's position in the labor market. $\Lambda$ measures the amount of the pay difference attributable to differences in observable characteristics assuming the same prices over time and gender. $\Omega$ accounts for differences in human capital and other observable labor market characteristics in the economy over time. The underlying prices are the coefficient estimates obtained when considering only individuals with $F_i = 1$ given $X_i$. Equivalently, the prices could be expressed as the coefficient estimates obtained when considering only individuals with $F_i = 0$ given $X_i$ thanks to the constraint imposed: $\gamma_{XF} = -\gamma_{XM}$. $\Theta$ accounts for differences in endowments by gender holding the second indicator variable fixed, i.e. setting the index $Y = A$. The component $\Upsilon$ can be re-written as:

$$\begin{aligned}\Upsilon &= [\sum \bar{x}^A \hat{\gamma}_{XYF}] \\ &= [\bar{x}_{1A}\hat{\gamma}_{XYF} + \bar{x}_{0A}(-\hat{\gamma}_{XYM})] \\ &= \underbrace{\bar{x}_{FA}\hat{\gamma}_{XYF}}_{\text{disadvantage of women}} - \underbrace{\bar{x}_{MA}\hat{\gamma}_{XYM}}_{\text{advantage of men}}\end{aligned}$$

For the component $\Upsilon$, the underlying set of characteristics are the average male and female endowments observed in $Y = A$, respectively. The prices can be expressed in terms of men's advantage or women's disadvantage given average characteristics of $X_i \; \forall \; i$.

Again, the pooled wage equation including the gender parameters and the male and female earnings equations are estimated separately using additional constraints for each categorical variable, i.e. under the zero-sum constraint (5.B.1).

# Appendix 5.D Intercept-Shift Approach versus Pooled-Sample Approach

Lee (2015) shows that the intercept-shift approach proposed by Fortin (2008) presents two drawbacks. Firstly, the reference parameter for the Oaxaca-Blinder decomposition, i.e. the parameter that would prevail in a 'fair' world under no discrimination, relies on the variance difference among categories. Secondly, the reference intercept is arbitrary: the same Oaxaca-Blinder decomposition holds with vastly different reference intercepts.

However, it can be easily shown that our proposed decomposition does not suffer from any of these aspects. Our decomposition arises from a specification that allows different intercepts and slopes. In addition, the constraints imposed on the parameters that identify the counterfactual reference parameters are the parameters such that the advantage of men is equal to the disadvantage of women. In fact, in our model the slope that would prevail under *no discrimination*, $\gamma$, is the sample average of the group slopes; $\beta_{0M}$ and $\beta_{0F}$:

$$\gamma = 0.5\beta_{0M} + 0.5\beta_{0F}$$

i.e. it is equivalent to considering the weights proposed by Reimers (1983).[21] Moreover, the constraint:

$$\beta_{0F} - \gamma_{0F} = \beta_{0M} + \gamma_{0F}$$

[21] See also Lee (2015).

prevents the indeterminacy problem shown by Lee (2015). It turns out, that in our model, the intercept indeterminacy problem highlighted by Lee (2015) is ruled out by imposing the constraint that the advantage of men should be equal to the disadvantage of women.

# Appendix 5.E Definition of Variables

Table 5.E.1 Definition of Variables

| Variable Name | Definition |
|---|---|
| | Dependent Variables |
| Lhwage | Natural logarithm of net hourly wages |
| | Hourly wages in Euros, net of taxes and social security contributions |
| | Independent Variable |
| | Group Dummies and Interaction Terms |
| female | One if the individual is a woman, zero otherwise |
| year | One if year is 2005, zero otherwise |
| private | One if individual is employed in the private sector |
| femyear | Interactive effect of *year* and *female*, i.e. one if employee is observed in 2005 and is female, zero otherwise |
| fempriv | Interactive effect of *private* and *female*, i.e. one if employee is employed in the private sector and is female, zero otherwise |
| Inter_female_$X$ | Interactive effect of *female* and the set of regressors $X$; *Inter_female_Exper–Inter_female_Intermed_Prof* |
| Inter_year_$X$ | Interactive effect of *year* and the set of regressors $X$; *Inter_year_Exper–Inter_year_Intermed_Prof* |
| Inter_femyear_$X$ | Interactive effect of *femyear* and the set of regressors $X$; *Inter_femyear_Exper–Inter_femyear_Intermed_Prof* |
| Inter_private_$X$ | Interactive effect of *private* and the set of regressors $X$; *Inter_private_Exper–Inter_private_Intermed_Prof* |
| Inter_fempriv_$X$ | Interactive effect of *fempriv* and the set of regressors $X$; *Inter_fempriv_Exper–Inter_fempriv_Intermed_Prof* |
| | Labor Market Presence |
| Exper | Number of years of prior work experience |
| Exper2 | *Exper* squared |
| Tenure | Number of years worked for current employer |

Educational Attainment

| | |
|---|---|
| Schooling | Number of years of schooling completed |

Job Characteristics

| | |
|---|---|
| Work_Climate | Individual's level of satisfaction with the working climate at the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Stab | Individual's level of satisfaction with the stability of the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Time | Individual's level of satisfaction with the working time at the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Work_Task | Individual's level of satisfaction with the tasks at the individual's current job $\in (0,4)$, where 4 is the highest level of satisfaction and 0 the lowest |
| Contract_Type | One if the individual holds an unlimited contract, zero otherwise |

Demographic Background

| | |
|---|---|
| Italian | One if individual is Italian, zero otherwise |
| Homeowner | One if individual owns a house (including houses financed by bank loans), zero otherwise |
| North | One if the individual lives and works in the North of Italy, zero otherwise |
| Centre | One if the individual lives and works in the Centre of Italy, zero otherwise |

Family Background

| | |
|---|---|
| Married | One if individual is married, zero otherwise |
| Educ_Moth_Uni | One if mother's education is equal to *Laurea*, i.e. mother holds a university degree, zero otherwise |
| Educ_Fath_Uni | One if father's education is equal to *Laurea*, i.e. father holds a university degree, zero otherwise |

Industry and Occupations

| | |
|---|---|
| Sec_Ind | One if individual is engaged in the industrial sector, zero otherwise |
| Sec_Tour | One if individual is engaged in tourism, zero otherwise |
| Sec_Trans | One if individual is engaged in transport, zero otherwise |
| Sec_Comm | One if individual is engaged in communication, zero otherwise |
| Sec_Fina | One if individual is engaged in financial sector, zero otherwise |
| Sec_Serv | One if individual is engaged in firm services, zero otherwise |
| Sec_PA | One if individual is engaged in the public administration, zero otherwise |

| | |
|---|---|
| Sec_Heal | One if individual is engaged in health, zero otherwise |
| Sec_Prof | One if individual is engaged in science and other professional activities, zero otherwise |
| Manager | One if individual executes intellectual professions; scientific and highly specialized occupations, zero otherwise |
| Intermediate_Prof | One if individual executes intermediary positions in commercial, technical or administrative sectors, health services and technicians, zero otherwise |

# Appendix 5.F Regression Output from the Full Specification

Table 5.F.1 OLS Estimates of Log Hourly Wages – Case 1, Full Specification

| | (1) |
|---|---|
| Variables | |
| femyear | -0.051 |
| | (0.185) |
| female | -0.148 |
| | (0.152) |
| year | -0.010 |
| | (0.130) |
| Exper | 0.019*** |
| | (0.002) |
| Exper2 | -0.000*** |
| | (0.000) |
| Tenure | 0.004*** |
| | (0.001) |
| Schooling | 0.038*** |
| | (0.005) |
| Contract_Type | 0.080*** |
| | (0.023) |
| Work_Climate | 0.001 |
| | (0.008) |
| Work_Time | 0.009 |
| | (0.007) |
| Work_Task | -0.002 |
| | (0.008) |
| Work_Stab | -0.024*** |
| | (0.007) |
| North | 0.060*** |
| | (0.014) |
| Centre | 0.038** |
| | (0.015) |

| | |
|---|---|
| Italian | 0.004 |
| | (0.065) |
| Homeowner | -0.006 |
| | (0.018) |
| Married | 0.062*** |
| | (0.014) |
| Educ_Moth_Uni | -0.011 |
| | (0.033) |
| Educ_Fath_Uni | 0.069*** |
| | (0.027) |
| Manager | 0.136*** |
| | (0.020) |
| Intermed_Prof | 0.035*** |
| | (0.013) |
| Constant | 1.163*** |
| | (0.110) |
| Industrial and Occupational Dummies | Yes |
| Interaction Terms | Yes |
| Observations | 17,918 |
| R-squared | 0.291 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 5.F.2 OLS Estimates of Log Hourly Wages – Case 2, Full Specification

| | (1) |
|---|---|
| Variables | |
| fempriv | -0.724** |
| | (0.289) |
| female | 0.278 |
| | (0.196) |
| private | 0.309 |
| | (0.205) |
| Exper | 0.019*** |
| | (0.004) |
| Exper2 | -0.000*** |
| | (0.000) |
| Tenure | 0.002 |
| | (0.001) |
| Schooling | 0.055*** |
| | (0.007) |
| Contract_Type | 0.182*** |
| | (0.057) |
| Work_Climate | 0.014 |
| | (0.012) |
| Work_Time | -0.001 |
| | (0.013) |

| | |
|---|---|
| Work_Task | -0.004 |
| | (0.014) |
| Work_Stab | -0.017 |
| | (0.013) |
| North | 0.049** |
| | (0.023) |
| Centre | 0.072*** |
| | (0.023) |
| Italian | -0.177*** |
| | (0.063) |
| Homeowner | 0.050 |
| | (0.032) |
| Married | 0.031 |
| | (0.026) |
| Educ_Moth_Uni | 0.074 |
| | (0.058) |
| Educ_Fath_Uni | 0.043 |
| | (0.046) |
| Manager | 0.118*** |
| | (0.032) |
| Intermed_Prof | -0.015 |
| | (0.024) |
| Constant | 1.046*** |
| | (0.147) |
| Industrial and Occupational Dummies | Yes |
| Interaction Terms | Yes |
| Observations | 8,423 |
| R-squared | 0.236 |

Robust standard errors in parentheses
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

# References


Arulampalam, W., Booth, A. L., and Bryan, M. L. (2007). Is there a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial and Labor Relations Review*, 60(2), 163-186.

Bar, M., Kim, S., and Leukhina, O. (2015). Gender Wage Gap Accounting: The Role of Selection Bias. *Demography*, 52(5), 1729-1750.

Black, S. E., and Spitz-Oener, A. (2010). Explaining Women's Success: Technological Change and the Skill Content of Women's Work. *The Review of Economics and Statistics*, 92(1), 187-194.

Blau, F. D., and Kahn, L. M. (1992). The Gender Earnings Gap: Learning from International Comparisons. *The American Economic Review*, 82(2), 533-538.

Blau, F. D., and Kahn, L. M. (2003). Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1), 106-144.

Blau, F. D., and Kahn, L. M. (2006). The US Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blau, F. D., and Kahn, L. M. (2016). *The Gender Wage Gap: Extent, Trends, and Explanations* (No. W21913). National Bureau of Economic Research.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4), 436-455.

Bordogna, L. (2012). Employment Relations and Union Action in the Italian Public Services – Is There a Case of Distortion of Democracy. *Comparative Labor Law and Policy Journa*l, 34(2), 507-530.

Christofides, L. N., and Michael, M. (2013). Exploring the Public-Private Sector Wage Gap in European Countries. *IZA Journal of European Labor Studies*, 2(1), 1-53.

Cotter, D. A., Hermsen, J. M., and Vanneman, R. (2004). *Gender Inequality at Work*. Russell Sage Foundation.

Cotton, J. (1988). On the Decomposition of Wage Differentials. *The Review of Economics and Statistics*, 70(2), 236-243.

England, P. (2006). Toward Gender Equality: Progress and Bottlenecks. In Blau F., Brinton M., and Grusky D. (Eds.), *The Declining Significance of Gender?* (pp. 245-264). Russell Sage Foundation.

Eurostat (2017). *Gender Pay Gap in Unadjusted Form*. http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (accessed 27-05-2017).

Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *Journal of Human Resources*, 43(4), 884-918.

Fortin, N. M. (2015). Gender Role Attitudes and Women's Labor Market Participation: Opting-out AIDS and the Persistent Appeal of Housewifery. *Annals of Economics and Statistics*, (117/118), 379-401.

Fortin, N. M., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. *Handbook of Labor Economics*, 4, 1-102.

Gardeazabal, J., and Ugidos, A. (2004). More on Identification in Detailed Wage Decompositions. *The Review of Economics and Statistics*, 86(4), 1034-1036.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much?. *Journal of Labor Economics*, 34(2), 509-543.

Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. *American Economic Review*, 96(2), 1-21.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4), 1091-1119.

Gornick, J. C., and Jacobs, J. A. (1998). Gender, the Welfare State, and Public Employment: A Comparative Study of Seven Industrialized Countries. *American Sociological Review*, 63(5), 688-710.

Grimshaw, D. (2000). Public Sector Employment, Wage Inequality and the Gender Pay Gap in the UK. *International Review of Applied Economics*, 14(4), 427-448.

Jann, B. (2008). The Blinder-Oaxaca Decomposition for Linear Regression Models. *The Stata Journal*, 8(4), 453-479.

Lee, M. J. (2015). Reference Parameters in Blinder-Oaxaca Decomposition: Pooled-Sample versus Intercept-Shift Approaches. *Journal of Economic Inequality*, 13(1), 69-82.

Lucifora, C., and Meurs, D. (2006). The Public Sector Pay Gap In France, Great Britain and Italy. *Review of Income and Wealth*, 52(1), 43-59.

Mandel, H., and Semyonov, M. (2014). Gender Pay Gap and Employment Sector: Sources of Earnings Disparities in the United States, 1970-2010. *Demography*, 51(5), 1597-1618.

Melly, B. (2005). Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression. *Empirical Economics*, 30(2), 505-520.

Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *Journal of Human Resources*, 23(3), 279-295.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

Oaxaca R. and Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61(1), 5-21.

Oaxaca, R. L., and Ransom, M. R. (1999). Identification in Detailed Wage Decompositions. *The Review of Economics and Statistics*, 81(1), 154-157.

Piazzalunga, D., and Di Tommaso, M. L. (2015). *The Increase of the Gender Wage Gap in Italy during the 2008-2012 Economic Crisis*. IZA Discussion Papers, No. 9931.

Reimers, C. W. (1983). Labor Market Discrimination Against Hispanic and Black Men. *The Review of Economics and Statistics*, 65(4), 570-579.

Yun, M. S. (2005). A Simple Solution to the Identification Problem in Detailed Sage Decompositions. *Economic Inquiry*, 43(4), 766-772.

# Chapter 6

# Conclusion

Gender differences in pay are a pervasive feature of modern labor markets. Despite the convergence of the wage gap over time and the implementation of equal-pay legislation, gender-related wage gaps continue to persist all around the world. This thesis examines gender-specific selection into wage work as well as its effects on the Gender Pay Gap (GPG). It contributes to the empirical literature on gender differences in pay by providing additional insights on the drivers of the GPG and by examining the pay gap in different environments (public-contest recruitment, overeducation). In particular, a double selection approach is applied in order to obtain consistent estimates of the wage gap and of its components by public-contest selection or overeducation. From the methodological perspective, this work provides two new approaches to estimate the GPG. First, the Recentered-Influence-Function OLS (RIF-OLS) model based on Unconditional Quantile Regressions (UQRs) is extended in order to account for sample selection issues across the wage distribution. The method proposed has several advantages compared to the standard approach such as being relatively easy to implement and interpret. In particular, UQRs allow for the unconditional mean interpretation that is important in (detailed) decomposition models and especially relevant for policy makers. Second, it provides an estimation approach for the GPG over time. The proposed estimation procedure offers a comprehensive and path-independent decomposition. In particular, direct inference on the change of the GPG as well as of its components over time can be drawn. The proposed model can be applied to various other decomposition problems of changes in mean group differences.

After providing an overview of the subject on gender differences in pay as well as its relevance, the GPG is analyzed separately for public-contest recruited employees and non public-contest selected employees in Chapter 2. In the former case, the GPG disappears, while it remains positive and significant in the latter. For the sample of young individuals, we find a reversal of the GPG in case of public-contest recruitment. The reversal among young public-contest selected workers is entirely due to the explained part. In contrast, the GPG among individuals that are not hired by public contest is mainly driven by the unexplained part of the wage gap. In particular, the result is valid for a time period of ten years using both the panel dimension of the data set as well as the single cross sections. The result still holds when controlling for double selection into employment and public-contest recruitment. The estimation strategy yields consistent estimates of the gender-specific wage equations, and thus of the components of the GPG. The disappearance of the GPG among public-contest selected employees, however, depends on the institutional environment it is conducted in. Yet, the result is not only driven by the public sector as we find among civil servants that are not selected by public contest a significant and positive GPG.

Chapter 3 presents the GPG for overeducated and properly educated workers. Overeducation may signal the incapacity of the labor market to absorb higher levels of education. As policy makers want to address this potential inefficiency, information on the contributors to pay disparities attributed to overeducation are important. The wage gap between men and women is found to be significantly higher among overeducated workers. The decomposition outcome is then adjusted for both labor force participation and the decision to accept a job offer for that the individual is actually overeducated. After the correction, the discriminatory part of the wage gap among overeducated workers vanishes. Higher wage gaps between men and women among overeducated workers are mainly explained by less favorable sets of unobservable characteristics of overeducated women (relative to overeducated men). So far, these two research fields, the GPG on the one side and overeducation on the other, have not been integrated jointly in the literature. Moreover, studies looking at gender differences in pay in the graduate labor market do not find that overeducation significantly triggers the wage gap. Contrary, by applying a bivariate probit model, we find that unobservables, due to the overeducation choice, explain a major part of the pay gap among mismatched workers.

In Chapter 4, a model is proposed in order to correct gender-related selection at specific quantiles as well as inequality measures across the wage distribution. The approach is based on the RIF-OLS procedure providing an intuitive and easy way to interpret and to estimate quantile-specific wage equations. In particular, it allows for the unconditional mean interpretation, an important feature when decomposing wage gaps. Moreover, the method is interesting for policy evaluation as it provides estimation results that are applicable to the entire distribution. In contrast, the main part of the quantile literature focuses on Conditional Quantile Regressions (CQRs) that allow only for conclusions on the conditional wage distribution (and not on the unconditional). In fact, CQRs provide results that are only relevant for subgroups of the target population. The extended model uses semiparametric estimators for selection correction and hence does not require any distributional assumptions of the error terms. In contrast, in the standard selection-correction model, i.e. the Heckman-two-step procedure, parametric estimators for selection correction, relying on normally distributed errors are assumed. The results suggest that employment selection significantly differs among men and women all along the wage distribution. This underlines the importance of considering quantile-specific selection when it comes to policy measures. Indeed, policy makers may be particularly interested in addressing wage inequality at diverse or extreme points of the earnings distribution. The results of Chapter 4 find that the adjusted GPG is underestimated at the top but overestimated at the median and bottom of the earnings distribution in the base model, i.e. without sample selection correction.

Finally, in Chapter 5, an alternative estimation approach based on the Omitted Variable Bias (OVB) formula is proposed allowing to directly estimate the change in the GPG over time as well as changes in its components over time. The model, contrary to the standard procedure in the literature, allows to conduct statistical inference of the change in mean group differences. The proposed method can be applied to a variety of decomposition problems. We show that the estimation approach can be easily made robust to the index-number problem of the standard Oaxaca-Blinder decomposition as well as to the indeterminacy problem of the intercept-shift approach. By using the proposed model, it is shown that the convergence of the GPG, during the last decade in Italy, is entirely due to the catching-up of women in terms of observable characteristics. As a second empirical application, the Public-Private Sector Wage Gap (PPWG) between men and women is estimated. The difference in the PPWG for men and women is driven by both endowments and coefficients. Additionally, interaction effects significantly impact on the difference of the sector-related wage gap between men and women.

All in all, the findings show that it is important to consider generally unobserved characteristics in the estimation of the GPG at the mean as well as at different points of the wage distribution. In particular, the effect of employment selection on earnings varies along the wage distribution. Public-contest and overeducation selection significantly impact on the level of wages. Given public-contest recruitment, the GPG disappears on average and the GPG among overeducated workers is mainly explained by differences in (generally) unobservable personal characteristics. The unexplained component is an important driver of the GPG and did not significantly decline over the last decade in Italy.