ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Overeducation and the Gender Pay Gap in Italy. A Double Selectivity Approach

Carolina Castagnetti
(Università di Pavia)

Luisa Rosti
(Università di Pavia)

Marina Toepfer
(University of Hohenheim)

 # 144 (10-17)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

October 2017

Overeducation and the Gender Pay Gap in Italy. A Double Selectivity Approach

Carolina Castagnetti[1], Luisa Rosti[2] and Marina Toepfer[3]

October 19, 2017

Abstract

We use a large Italian data set (ISFOL-PLUS 2005-2014) to estimate the gender pay gap (GPG) among overeducated workers. We show that overeducation is an important driver of the GPG. This result holds when controlling for sample selection and endogeneity problems, too. Neglecting selectivity issues may lead to the conclusion that discrimination is the most important driver of the GPG. Yet, when accounting for self-selection and endogeneity bias overeducation is found to merely reflect unobserved differences in personal characteristics such as innate ability. The selection coefficients for both the participation and the overeducation decision allow explaining almost the entire GPG.

Keywords
overeducation; gender pay gap; double selection; Italy; discrimination; wages

JEL Classifications: I26, J16, J31, J71

[1] Università di Pavia, Dipartimento di Scienze Economiche e Aziendali, via San Felice 5, 27100 Pavia
[2] Università di Pavia, Dipartimento di Scienze Economiche e Aziendali, via San Felice 5, 27100 Pavia. Corresponding Author. E-mail address: lrosti@eco.unipv.it
[3] University of Hohenheim, Institute of Economics, Chair of Labour Economics and Service Economics

1. Introduction

This paper contributes to the existing literature by integrating insights from two usually separate research fields: overeducation and the Gender Pay Gap (GPG). As in the literature the main sources of the GPG are near the same proposed to explain gender differences in overeducation choices (that is, explanations focused on the preference-based choices made by women or on job-related constraints faced by women) we investigate whether overeducation may be an important driver of the GPG.

Workers in occupations that require less schooling than they actually have are labeled *overeducated* (Sicherman, 1991; Hartog, 2000; Sloane, 2014). Despite the broad literature on the incidence of overeducation by gender and on the relationship between overeducation and earnings on the one hand[1], and the even wider literature on the GPG remaining resilient despite more than thirty years of equal-pay legislation on the other[2], we find very few studies focusing on the relationship between overeducation and the GPG. Moreover, these few studies find, overall, that overeducation does not matter for explaining the GPG. For example, Boll and Leppin (2013) in Germany and Li and Miller (2012) in Australia find that overeducation does not substantially contribute to the GPG among graduates.

However, previous studies also criticize empirical strategies that do not take into account the need to control either for the *participation bias*, that is, non-random self-selection into employment (Battu et al., 1999) or for the *endogeneity bias*, that is, non-random self-selection into overeducation (Dolton and Silles, 2008). In order to consistently estimate the gender-specific wage equations and the components of the GPG, we follow these suggestions by applying a bivariate selectivity model. The model accounts simultaneously for both the participation bias and the endogeneity bias (as in Tunali, 1986; Sorensen, 1989; Cutillo and Di Pietro, 2006).

The results suggest that overeducation is an important driver of the GPG in Italy. The GPG is significantly higher among overeducated workers than among properly educated workers because women's unobservable characteristics driving female employment into overeducation also drag down female wages more than men's unobservable characteristics drag down male wages.

Unobserved heterogeneity consists of differences in individual productivity such as innate ability, school quality and on-the-job training, motivation and effort requirement, as well as commitment to paid work. Provided that any estimated impacts of overeducation on wages are free from heterogeneity bias, all the unexplained component of the GPG vanishes in our data, and the whole difference is explained by endowments (a small part) and the selection variable into overeducation (the big part). Controlling for both participation bias and endogeneity bias, we find that women possess better observed characteristics than men and receive the same reward as men for these characteristics in the overeducated sample.

As our results show that the discriminatory component of the GPG disappears among the overeducated workers but remains significant among the properly educated ones, we further investigate why overeducation can fight gender discrimination in pay whereas a proper match fails.

A possible explanation is that by their higher educational attainment overeducated women signal their actual although low productivity to employers (Spence, 1973) and overcome statistical discrimination. Statistical discrimination occurs when employers use average characteristics of groups to predict individual worker productivity (Arrow, 1973).

Our data show 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 send by overeducated women: They possess valuable, though, unobservable characteristics, and are available to work. Among the properly educated workers the signaling effect is less clear, because education also features human capital skills required for the job.

We draw the conclusion that overeducation is, first, a first-best matching for individuals (both men and women) compensating with more education for their lower productive characteristics. Second, it may be a signaling device for women spending their useless-for-the-job diploma to inform employers on their valuable though unobservable productive characteristics and fight gender wage discrimination.

Our results may be 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 investments their lower unobservable differences in productivity, there is no waste of resources. This implies that the need for greater investments in higher education is no more limited.

It is worth noting that overeducation is a serious problem in Italy, where the predicted probability of being overeducated for both men and women is very high, independently from the educational level or the estimation technique applied (Flisi et al., 2014). Yet, in Italy, the share of individuals with tertiary education is among the lowest of all EU member states. 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[3].

The paper is organized as follows. In the next two Sections, we define the phenomenon of overeducation as well as of the GPG and present the corresponding method of assessment. Section 4 describes the background literature. Section 5 outlines the problem of double selectivity. In Section 6, we outline the data set used as well as sample restrictions imposed. In Section 7, we present our estimation results. In Section 8, we discuss our results. Finally, we conclude.

2. Overeducation Definition and Method of Assessment

According to the literature overeducation is a pervasive feature of modern labor markets; 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 per cent to 42 per cent. On average, 26 per cent of all worker in the United States and 22 per cent 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 33.4 per cent (35.2 per cent in the male sample and 31.8 per cent in the female sample)[4].

As stated, the literature usually considers workers as overeducated when they have completed more years of education than the current job requires (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[5].

Following previous studies (Mc Guinness, 2006; Hartog, 2000), several methods of overeducation assessment can be identified in empirical studies. These indicators can be classified into three groups: objective, subjective, and statistical.

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 be 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 what is 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 who 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).

In the survey ISFOL PLUS, overeducation is self-assessed by the workers and it is recorded according to a positive or negative reply to the following direct question: "Is your level of education necessary for your current job?".

We must point out that the survey does not request the interviewees to specify either if they refer to a substantial or formal necessity of their education degree or how much the skills acquired during their educational career are actually used.

However, on the one hand these caveats are typical of every survey analyzing overeducation by means of subjective answers, and on the other hand it must be stressed that ISFOL PLUS survey provides a very good opportunity to assess the phenomenon, as the abundance of information obtained and the large sample size allow to study wage levels and individual features by controlling for many explanatory variables.

3. Definition and Assessment of the GPG

The GPG represents the difference between the average gross hourly earnings of men and women expressed as a percentage of average gross 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, 2016a).

As already mentioned, for the economy as a whole, in 2014, the unadjusted GPG was 16.1 per cent in the European Union (EU-28) and 6.5 per cent in Italy[6]. It is important to point out that a small GPG in gross hourly wage does not imply a thin overall income inequality between women and men within the economy. When considering the gross annual income instead of the hourly wage, the differential increases significantly due to the 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 also differences in labor market participation and employment contribute substantially to the difference in average earnings of women versus men.

To give a complete picture of the gender earnings gap, Eurostat (2016b) has developed a new synthetic indicator called *gender overall earnings gap*. The indicator measures the impact of three combined factors (hourly earnings, paid hours and employment rate) on the average earnings of all men of working age compared to women, and it is estimated to be 41.1 per cent in Europe and 44.3 per cent in Italy in 2010[7]. At EU level, the *gender overall earnings gap* was driven mostly by the GPG (contribution of 37 per cent) and the gender employment gap (contribution of 35 per cent), with minor contribution of the gender gap in paid hours (28 per cent). In Italy, the gender gap in employment rates was the main contributor to the total earnings gap (65 per cent), followed by the gender gap in paid hours (26 per cent) and by the GPG (9 per cent). Although the GPG in hourly wages is only a relatively small part of the overall income inequality by gender in Italy, it is precisely the analysis of that small difference which brings out discrimination from the data, and we know from Becker (1985) that even small amounts of discrimination against women can cause huge differences in wages.

4. Background Literature

The aim of this paper is to study the relationship between overeducation and the GPG. The GPG may arise from differences in personal and job characteristics of working men and women, or may be the result of disparity in wages that persists when male and female workers have similar personal and job characteristics. This residual gap cannot be justified on grounds of productivity, than revealing the presence of gender discrimination in pay.

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. As our main statement in this paper is that overeducation is an important driver of the GPG, we summarize, first, 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 are near the same proposed in the literature to explain gender differences in overeducation choice. We then explore the findings about the risk of overeducation by gender, summarizing the results of studies that isolate the gender effect in the overeducation risk by controlling for a large set of related variables. Finally, we present both the literature pictures and our own results about the relationship between overeducation and gender discrimination in pay.

4.1 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 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, 2004; Ridgeway and Correll, 2004; England, 2010).

Economists also argue that women earn less than men because of the division of labor within the family, which results in productivity differences between men and women through its effect on human capital accumulation. Becker (1965) emphasizes the importance of household production in economic theory, suggesting 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 (Goldin, 2014), and that women with children earn less than other women (Waldfogel, 1997).

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 individuals' productivity (Mincer and Polachek, 1974). It is worth noting that while in the past men typically had better access to university-level institutions, nowadays female graduates exceed the number of male graduates, and on average female students outperform male students in academic achievements in most OECD countries (OECD 2009). Unfortunately, despite their progress in higher education, women continue to choose traditionally female majors (England and Li, 2006). Empirical studies have found that the choice of college major explains to a greater extent the GPG than differences in levels of education (Bobbitt-Zeher, 2007).

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. If female dominated occupations have some benefits making easier to combine work and family life, these benefit may result in lower wages (Solberg and Laughlin, 1995).

The last explanation on the demand side of the labor market, for gender income differences, is discrimination against women, that is, employers' gender-biased decisions on the allocation of individuals across and within occupations. The data 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 (Arrow, 1972; Stiglitz, 1973) occurs when employers make hiring and promoting decisions based not on an individual's personal characteristics, but on the average productivity of the individual's gender. It is worth noting that the assessment of productivity in the workplace is strongly influenced by stereotypes that may affect our judgments of others, creating workplace discrimination (Eagly and Steffen, 1984). Both gender stereotyping and wage discrimination have been well documented in empirical research (Blau et al. 2010). Among these, Authors (2013) show that stereotyping is clearly related to gender wage discrimination.

4.2 Theories on overeducation

We find in the literature 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). In the human capital perspective (Mincer, 1958; Becker, 1964), overeducation is a mechanism for labor market adjustment when there is an excess supply of high-skilled workers, and it exists as 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[8]. The human capital theory has received some confirmation in studies estimating the negative impact of work experience on overeducation risk (e.g. Alba-Ramirez, 1993; Büchel and Battu, 2003; Boll et al., 2016).

In the job signaling model (Arrow, 1973; Spence, 1973; Stiglitz, 1975) education is used as a screening device to identify higher ability workers. Firms are assumed to have imperfect information about the productivity of workers, and in response to this gap, 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. Kedir et al. (2012) show that signaling effects are relevant in their empirical analysis. Estimating a model with individual and employment status fixed effects, they test whether the positive returns to overeducation pervasively found by the previous literature are just a consequence of signaling activities. They really find that overeducation does not increase worker's productivity. Any positive returns to overeducation are merely due to the signaling effect. Additional 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).

A different approach (Bauer, 2002; Chevalier, 2003) argues that overeducation may be only apparent, as a consequence of measurement errors due to unobserved heterogeneity. Even if 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 that the overeducated may substitute surplus education 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, or education quality, or unmeasured skills, or worker preferences (e.g. preference for family-friendly work schedules). Chevalier (2003) names this measurement error *apparent overeducation*, to be distinguished from *genuine overeducation*. In any case, in terms of productivity, the theory postulates that if overeducation is only *apparent*, all workers are correctly matched.

As stated by Leuven and Oosterbeek (2011), 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. Other empirical evidences of the key role of unobserved heterogeneity in explaining the overeducation phenomenon are Leuven and Oosterbeek, 2011; McGuiness and Bennet, 2007; Bauer 2002). In line with these findings, our approach explicitly accounts for unobserved heterogeneity and our results support this body of literature.

Other explanations for overeducation have also been suggested. For example, the career mobility theory (Sicherman and Galor, 1990), the theory of job competition (Thurow, 1975), the assignment theory (Sattinger, 1993), and the theory of differential overqualification (Frank, 1978). (See Boll et al., 2016, for a survey).

4.3 The overeducation risk by gender

In the literature, the risk to be overeducated may be related either to individual characteristics on the one hand and to job related outline on the other. However, the specific linkage of these factors to the overeducation risk is often weak in empirical research (Boll et al., 2016).

We consider individual characteristics first. A substantial determinant of being overeducated may be a lower level of individual ability. Chevalier and Lindsey (2009) find a negative correlation between a measure of unobserved ability and the probability to be overeducated. Also Dolton and Vignoles (2000) and Green et al. (2002) find the same result.

Among individual's characteristics the role of gender differences has received a large amount of attention in the literature. As in many countries the share of overeducated workers among women is higher than among men, econometric studies attempt to isolate the gender effect by controlling for a large set of related variables in order to discover discrimination, if any.

A slight majority of these studies finds that the effect of gender on overeducation risk is insignificant (Groot and Maassen van den Brink, 2003; Green and McIntosh, 2007; Capsada-Munsech, 2015). Conversely, Alba-Ramirez (1993), Groot (1996) and European Commission (2012) obtained the result that male employees face a slightly higher overeducation risk, an effect which is however in all cases only weakly significant. In contrast, Ortiz and Kucel (2008) estimate that female workers are at significantly higher risk. Robst (2007) finds that males are more likely to be overeducated due to career-related reasons, while females are more likely to be mismatched due to family-related reasons (although he acknowledges the possibility of reporting bias due to social norms).

These contrasting results could at least in some part be explained by differences in the measurement methods applied. McGoldrick and Robst (1996) point at this possibility by comparing estimation outcomes based on an objective measure of overeducation (statistical distribution) with a subjective one (questionnaires). They find that women are significantly more likely to experience overeducation than men under the subjective measure, but significantly less likely under the statistical measure.

Moreover, as Büchel and van Ham (2003) document, these mixed results may be due to the selection process concerning female labor market participation. On the one hand, a high reservation wage can induce a female job applicant to turn down low-pay offers with low qualification requirements, thereby reducing the frequency of overeducation. On the other hand, some comparatively less demanding jobs especially in administration allow for more time flexibility than most high-level leadership positions. This fact raises the attractiveness of these jobs for women, bringing an opposite effect of children on overeducation risk. Without controlling for self-selection, Büchel and van Ham (2003) find for female workers, a significant positive impact of the number of children on the risk of overeducation. Significance however disappears in the Heckman self-selection specification, indicating that the most relevant effect of children already influences the decision to enter the labor market.

Finally, the risk to be overeducated may be related to both employment sector and job characteristics.

The existing literature suggests that the risk of overeducation is lower in the public than in the private sector (Wolbers 2003). Barone and Ortiz (2011) find that countries with a large share of employment in the public sector display lower overeducation incidence, despite the prominent expansion of higher education. Strictly related to public-sector employment are the contract length and the job tenure. People with fixed term contracts are more likely to work in positions for which they are overeducated than people with permanent ones. Green and McIntosh (2007) as well as Ortiz (2010) find some evidence for a significantly lower overeducation risk among workers in permanent positions. Moreover, Büchel and van Ham (2003); Büchel and Battu, (2003); Groot and Maassen van den Brink, (2003); Ortiz, (2010) and European Commission, (2012) find a significant negative effect of job tenure on overeducation risk.

5. Decomposing the GPG Accounting for Double Selection

The GPG may arise from differences in personal and job characteristics of working men and women, or may be the result of disparity in wages that persists when male and female workers have similar personal and job characteristics. This residual gap cannot be justified on grounds of productivity, than revealing the presence of gender discrimination.

As we are interested in studying the relationship between overeducation and the GPG, we focus on the adjusted measures of differences in hourly wages that persist even when employed women and men are similar with regard to personal and job characteristics.

The standard approach in decomposing the GPG is the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca , 1973; see, e.g. Weichselbaumer and Winter-Ebmer, 2005). 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 the 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 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 (overeducated individuals and properly educated individuals), we estimate a Mincer-type wage equation separately for men and women (Appendix A.1). Then, we describe the decomposition method applied (Appendix A.2).

We first check the adequacy of our data to explain differences in wages other than the GPG and verify the hypothesis that the same type of wage decomposition can capture most of the differences in productivity that explain the pay gap between groups other than gender (Overeducated vs. Properly Educated Individuals; Public vs. Private Sector; Full-time vs. Part-time Employment, Recruitment by Public Contest vs. Recruitment without Public Contest).

The outcome of paid work either as a properly educated worker or as an overeducated worker is only observed for a non-random sample. Therefore, the coefficients obtained from Ordinay 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 participation decision may depend on some positive factors such as individual ability, commitment to paid work, 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 selectivity bias that stems from not considering the participation decision may be particularly relevant in Italy given low female participation in the Italian labor market (see De la Rica et al., 2008; 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 a job for which they are overeducated or to accept only jobs offers in accordance with their educational level. In this case, sorting into the over- or properly educated sample may be a result of differences in unobservable characteristics between the individuals. Failure to account for 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, i.e. there would be an overinvestment in educational attainment and a waste of resources. However, this may not be the case when overeducation mainly reflects unobserved differences in productive characteristics that are unobservable in the data.

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). 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 two-stage selection model to include multiple decisions (Dubin and McFadden, 1984; Tunali, 1986; Sorensen, 1989; Schmertmann, 1994). 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 A.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, which are then included in the wage regression and present the decomposition expression when accounting for double selection into the sample (the Oaxaca-Blinder model with double selection).

6. Data and Sample Restriction

We use the complete release of the survey ISFOL 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 with the Italian Ministry of Labor and Social Policy that was started in 2005. The survey was released up to now for the following years 2005, 2006, 2008, 2010, 2011 and 2014. The project aims particularly at creating a data set for the study of wage inequality by gender. Hence, it delivers broad information on the personal working profiles and individual motivation to work as well as on the cultural and territorial background of the participants (Centra and Cutillo, 2009). ISFOL PLUS covers the whole population with focus on the working population. The data was collected by means of Computer Assisted Telephone Interviewing (CATI). One of the main characteristics of the national survey is that only answers with direct responses were considered, that is no proxies were used. 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 lose individuals. Thus, the sample composition changes. As the transition in and out of overeducation is very low (approximately one percent), we base the analysis on a pooled regression model and include dummies for the different releases or years 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. In the original sample, there are 120,353 individuals with at least 13 years of schooling or high school diploma.

We also exclude students, pensioners and disabled individuals from the sample because their job choices are limited (Beblo et al. 2003). Furthermore, we exclude "involuntarily" unemployed individuals from the sample. We consider as "involuntarily" unemployed those individuals self-reporting their unemployment status and answering YES to the question: *Would you be immediately available to work?* The aim of this restriction is to form a homogenous sample of individuals voluntarily out of the labor force and employed individuals, respectively (Beblo 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 per cent) and 19,452 are male (45.1 per cent). In the data 6,775 men (47.5 per cent) and 7,481 women (52.5 per cent) are working in jobs that require less schooling than they actually have (i.e. are overeducated). Thus, more than one third of 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.

Table 1 reports means and standard deviations for some of the explanatory variables used in the analysis. On average, overeducated workers (both males and females) have lower schooling, less experience and job tenure. Moreover, overeducated employees are younger and less often married or parents than properly matched employees. A full list of the variables used in the analysis along with their definitions and coding is provided in Appendix B, Table B1.

[Table 1 about here]

7. Estimation Results

In this Section we present our estimation results. First of all, we estimate the effects of overeducation on wages and calculate the incidence of overeducation for both men and women. We show that the difference in the GPGs between properly and overeducated individuals is significant and that the adjusted as well as unadjusted GPG is higher among overeducated workers. Then, we discuss the results from the model with double sample selectivity.

7.1 The effect of overeducation on wages and the GPG

Table 2 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 per cent[9]. The data also show that the GPG is much higher among overeducated workers (9.6 per cent) compared to properly educated ones (3.5 per cent)[10].

[Table 2 about here]

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, column (1), shows that the coefficient estimate of *overfem* [11] is negative and statistically significant. The coefficient estimate is the difference of the GPG between properly and overeducated individuals; $-\left(\Delta^{\mathrm{GPG}_{Over}} - \Delta^{\mathrm{GPG}_{Proper}}\right) = \Delta^{\mathrm{GPG}_{Proper}} - \Delta^{\mathrm{GPG}_{Over}}$. 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[12]. 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 as well as wave or year dummies. Personal characteristics include family status, nationality, regional controls and the educational background of the parents. Table 4 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[13]. The estimated coefficient for *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 per cent[14]. The coeffcient of the variable *female* being negative and significant confirms the usual results in the literature: being a woman reduces earnings. Here, the female wage penalty amounts to 7.3 per cent. This penalty is higher in the sample of overeducated individuals (9.6 per cent) and lower in the sample of properly educated individuals (6.9 per cent). The coefficient for the interaction term *overfem*, negative and significant, shows that women receive from being overeducated a wage penalty of 2.2 per cent. 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.

[Table 3 about here]

[Table 4 about here]

7.2 Likelihood to be overeducated

The risk of overeducation may differ for men and women[15]. 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 5, Panel A, shows that in the full sample men are actually more likely to be overeducated. Panel B of Table 5 confirms this result: being a woman significantly reduces the probability to be overeducated. This finding is in line with results from the literature (e.g. Groot, 1996; Cutillo and Di Pietro, 2006; European Commission, 2012) [16]. However, as stated, the results in the literature concerning the overeducation risk by gender are ambiguous. Different estimation techniques as well as different measures for overeducation (subjective or objective) may contribute to this ambiguity (McGoldrick and Robst, 1996). In our sample the incidence of overeducation is higher for men than for women. Nonetheless, female employees perceive a more pronounced overeducation penalty on earnings than men (see Table 3 or Table 2).

[Table 5 about here]

7.3 In search of discrimination

In Section 5.1, 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 6 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' ones. The unexplained or coeffcients 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 overeducated and properly educated individuals, but is higher among overeducated workers, it suggests that gender wage discrimination may be more important among overeducated workers. The coefficients component among overeducated employees amounts to 83.4 per cent compared to 66.7 per cent 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. 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 7 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 per cent of the difference in pay between overeducated 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 coeffcients effect in the GPG as well as in the GPG by overeducation is not data-driven.

[Table 6 about here]

[Table 7 about here]

7.4 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 and endogeneity problems. The estimated correlation between the error terms of the two binary choice equations considered, ρ, 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, ρ 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 8) [17]. 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 8 shows that relocating significantly 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 9 the values of the four selection variables we consider in this study, for both men and women: λ_{PA}^{Over}, λ_{PA}^{Proper} (participation choice) and λ_{OV}^{Over}, λ_{OV}^{Proper} (overeducation choice), where *Over* identifies the overeducated sample and *Proper* the properly educated sample[18]. The coefficient estimate of λ_{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 λ_{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 matching 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. As expected, the coefficient estimates for the overeducation choice result positive for overeducated workers and negative for their properly educated counterparts. The coefficient estimate of λ_{OV}^{Over} is significant and positive for both women and men, while the estimated coefficient of λ_{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 in line with their educational background. Overall, our data show that individuals who 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 estimated coefficients from the earnings equations that have been corrected for participation and endogeneity bias, respectively. We provide in Table 10 the results of the Oaxaca-Blinder decomposition with double selection. 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: $\delta^m_{M,PA}\lambda^m_{M,PA} - \delta^m_{F,PA}\lambda^m_{F,PA}$, with $m = Over, Proper$. Analogously, the overeducation component of the GPG is equals to; $\delta^m_{M,OV}\lambda^m_{M,OV} - \delta^m_{F,OV}\lambda^m_{F,OV}$. 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 individuals, the overeducation decision exerts a strong positive impact on wages for both men and women (as shown in Table 9). However, the corresponding set of unobservables, λ^{Over}_{OV} , is more favorable for men, i.e. $\lambda^{Over}_{M,OV} > \lambda^{Over}_{F,OV}$. Consequently, the overeducation component is a net driver of the GPG among overeducated workers. On the contrary, the overeducation component is statistically significant but negative for the properly educated sample.

The set of unobservables, λ^{Proper}_{OV}, is more favorable for women than for men. 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 significantly 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 ones 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 a higher educational attainment their lower level of ability, motivation or educational quality. Their set of unobservables is lower than that of women in the properly educated sample 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. On the contrary, 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 8 about here]

[Table 9 about here]

[Table 10 about here]

8. Discussion

The literature shows that the risk of overeducation may differ by gender, and in many countries the share of overeducated workers among women is higher than among men. Hence, we first of all test the incidence of overeducation by gender in our data. We find that women are less likely to be overeducated than men, as in Alba-Ramirez (1993), Groot (1996), Cutillo and Di Pietro (2006) and European Commission (2012). 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 measure, as in our data, than under the objective measure (McGoldrick and Robst, 1996). Conversely, a possible explanation for the lower female probability to be overeducated is the higher concentration of best educated females in the public sector. As the literature shows that the risk of overeducation is lower in the public than in the private sector (Wolbers, 2003; Barone and Ortiz, 2011), the higher concentration of female employment in the public sector may help to explain women's lower risk to be overeducated. Another possible explanation consistent with our results may be as follows: We show in our analysis that overeducated workers possess lower unobservable characteristics compared to properly educated individuals. Moreover, women voluntarily out of work possess lower unobservable characteristics compared to working women, too. As we find that less productive men self-select mainly into overeducation, while less productive women split up either into overeducation or out of the labor force (e.g. housewives not available to work), we receive, as a consequence, that men are more likely to be overeducated than women. In line with these results, 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. On the other hand, some less demanding jobs, especially in administration, allow for more time flexibility than most high-level positions. The attractiveness of these jobs is higher for women, in so increasing their overeducation risk.

Second, we show that the effect of overeducation on earnings differs for men and women. By estimating a Mincerian wage equation we find that the wage reduction for being a woman amounts to 7.3 per cent. This penalty is higher in the sample of overeducated individuals (9.6 per cent) and lower in the sample of properly educated (6.9 per cent). These results suggest that overeducation may be an important driver of the GPG in Italy.

Our data show that the GPG is higher among overeducated workers compared to properly educated (9.6 per cent vs. 3.5 per cent). Therefore, we investigate why overeducation leads to higher disparity in pay between women and men. By applying the Oaxaca-Blinder methodology to study the drivers of the GPG we find that women possess better observable characteristics than men, but get lower reward for these characteristics in both the over- as well as the properly educated sample. In our data, the explained component of the GPG halves among overeducated workers compared to properly educated (16 per cent vs. 33 per cent), and the unexplained component exceeds eighty per cent in the overeducated sample (83 per cent vs. 67 per cent). Hence, we inquire why overeducation leads to an increase of gender discrimination in pay.

We know from the literature (e.g. Leuven and Oosterbeek 2011) that most of the difference in earnings between overeducated and properly educated workers are caused by a failure to control for unobserved heterogeneity. Chevalier (2003) argues that not all workers with a given educational qualification are perfect substitutes due to unobserved heterogeneity. Empirical evidences of the key role of differences in personal characteristics such as innate ability, school quality and on the job training, motivation and commitment to paid work, in explaining the overeducation phenomenon are in Bauer (2002), Chevalier (2003), Cutillo and Di Pietro (2006), McGuiness and Bennet (2007), Pecoraro (2016).

In order to consistently estimate the gender-specific wage equations, we apply a bivariate selectivity model to simultaneously account for both the participation bias and the endogeneity bias (as in Tunali 1986; Sorensen 1989; Cutillo and Di Pietro 2006). Our results show that the GPG is mainly explained by the overeducation componente. We present the estimates of four selection variables: two for the participation choice (male and female sample), and two for the overeducation choice (male and female sample). The coefficients for the participation choice in the overeducated sample, positive and significant for both men and women indicate that individuals actually out of the labor force possess lower unobserved characteristics, and would earn less than their overeducated peers, even if they were selected into overeducation. Also the coefficient estimates of the overeducation choice result positive and significant for both men and women indicating that overeducated workers possess lower unobserved characteristics, and would earn less than their properly educated peers, even if they were in a job matching their level of education. Conversely, as expected, the coefficient estimates of the overeducation decision for properly educated workers are negative and significant for both men and women.

While previous studies find that overeducation does not matter for explaining the GPG, we find that the overeducation choice is an important driver of the GPG. Moreover, we find significant differences in the GPG between over and properly educated individuals.

Boll and Leppin (2013) estimate the incidence of overeducation among German graduates following the realized matches approach and accounting for omitted variable bias. They find that overeducation induces severe wage losses compared to properly matched graduates (as in our data). They also find that the losses are even more pronounced for women (as in our data). Though overeducation does not contribute to the observed gender wage gap in their analysis, their results show severe selection effects, particularly with regard to women, due to unobserved heterogeneity. That is, female graduates are to a higher extent than their male counterparts subject to selection processes which themselves are driven by unobserved personal traits. Also in our data, when accounting for participation bias and endogeneity bias, the unexplained component of the GPG vanishes, and overeducation is found to merely reflect unobserved differences in personal characteristics.

Li and Miller (2012) use data from the Graduate Destination Surveys of alumni from Australian universities, and use either the job analysis approach or the realized matches approach to assess the educational mismatch. Their results show that the incidence of overeducation is higher for males than it is for females (as in our data). They also find that overeducation effects account for only a negligible portion of the GPG. However, the Oaxaca-Blinder- decomposition reveals sorting effects for males and females in their data, too. Moreover, it is worth noting that the results of Li and Miller (2012) may be affected by a *young age* effect that decreases the female wage disadvantage. Indeed, they consider only individuals four months after the completion of a qualification, when graduates are young, while the GPG is larger in the older age groups (Eurostat, 2016b).

9. Conclusion

In this paper we explicitly consider the effect of the overeducation choice on the GPG in Italy. The case of Italy may be of particular interest for the study of overeducation, given that on the one hand, a large share of individuals is overeducated, while on the other hand, the share of individuals with tertiary education is among the lowest in the European Union. This is important for policy issues. If overeducation indicates the incapacity of the Italian labor market to absorb all the newly graduates, there is an overinvestment in education and a waste of resources. Yet, this may not be the case if overeducated workers possess low productive characteristics and try to compensate their lower level of productivity by more investment in education in order to increase their employment probability. Our results show that this is the case, in Italy. Overeducation simply compensates for lower productivity levels, there is no waste of human capital, the need for greater investments in higher education is not limited, and the share of individuals with tertiary education may grow.

This conclusion is even more important for women, as their share among graduates is high and growing in Italy, and the wage penalty for overeducation is higher in the female sample.

Overall, controlling for unobserved heterogeneity, we find that women possess better observed characteristics than men and get the same reward as men for these characteristics in the overeducated sample. That is, controlling for unobserved characteristics removes all the unexplained component of the GPG. The part usually attributed to discrimination, i.e. the coefficients effect, vanishes, and the whole difference is explained by the endowments (a small part) and the overeducation component (the big part).

As in the overeducated sample, also in the properly educated sample women possess better observed characteristics than men. However, in contrast with the overeducated sample, the coefficients effect remains significant even when controlling for unobserved heterogeneity in the properly educated sample. Our results also show that, contrariwise to women self-selecting into overeducation, properly educated women possess a better set of unobservable characteristics compared to their male peers. These better unobservable characteristics reduce the unexplained part of the GPG by a non negligible fraction (from 67 per cent to 54 per cent), but a substantial gender discrimination remains among properly educated workers. Hence, we inquire why overeducation can fight gender discrimination in pay whereas a proper match fails.

A possible explanation is that by their higher educational attainment overeducated women signal their actual, though comparably lower productivity levels to employers and overcome statistical discrimination. Their wages are lower than men's because their either observable or unobservable characteristics are lower than men's. However, they suffer no gender discrimination as they receive the same reward as men for their endowments due to their signaling activity.

In the statistical discrimination framework, employers use average characteristics of groups to predict individual worker productivity. In this context, the education level can be seen as a proxy for unobserved positive individual characteristics, such as productivity (Arrow 1972). As Livanos and Núñez (2012) argue, discrimination arises from an adverse selection problem where the hidden information is women's commitment to professional career. In that case, education may act as a signal to employers, that is as a proof of future commitment of female workers to their careers. If the educational level is an effective signal of workers' commitment, then gender discrimination towards females among overeducated females should be lower.

For example, in absence of any signal, the employer could assess the impact of low commitment on productivity by observing females' choice between career and family tasks. In that case, the employer might impose the same wage penalization to all female workers, regardless of whether they are actually committed to their career or not. As this information remains private, employers only can rely on observable signals, such as the education level, in order to differentiate their wage offers. As higher education requires a certain level of investment, dedication and effort, employers consider women with higher levels of education to be less likely to abandon their jobs in the case of marriage or child-birth (Livanos and Núñez, 2012). This should be the case in our data, while we expect that among the properly educated workers the signaling effect is less clear, because education also features human capital skills required for the job.

Our data show that overeducated men and women possess worse unobservable characteristics than individuals in the properly educated sample. Moreover, overeducated working women possess worse unobservable characteristics than overeducated men. However, overeducated women (not men) are better than out of employment individuals. The latter may be the signal send by overeducated women: They possess some valuable though unobservable characteristics, and are available to work.

We draw the conclusion that overeducation is, first, the first best matching for individuals (both men and women) compensating with more education their lower levels of productive characteristics. Second, it may be a signaling device for women spending their useless-for-the-job diploma to inform employers on their valuable though unobservable productive characteristics, in so fighting gender wage discrimination.

These findings support further investment in education as a tool of eliminating discrimination in the labor market. Moreover, as females are less overeducated than males despite their larger representation in higher education, there should not be concern that expanding higher education will disadvantage females.

It is worth noting, however, that to the extent that education above the required level does not increase productivity enough to compensate for their cost, public investment leading to overeducation will be inefficient. This is the reason why, as suggested by Kedir et al. (2012), a severe limitation of the overeducation literature, and a possible limit of our study, too, is the potentially confounding effects of educational signals, that is, we do not know whether the returns to overeducation are the consequence of increases in productivity or rewards to educational signals. Moreover, further research could include in the analysis information associated to education but not available in our data, such as the field of study, the student's performance and the postgraduate credentials. Such information will permit a more precise control of the heterogeneity among overeducated workers, and confirm whether or not there is gender discrimination in pay.

References

Alba-Ramírez 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*. Geneva, International Labour Office.

Arrow, K. J. (1972). Models of Job Discrimination, In: A. H. Pascal, (ed), *Racial Discrimination in Economic Life*, D.C. Heath, Lexington.

Arrow, K. J. (1973). The Theory of Discrimination, in O. Aschenfelter and A. Rees (eds) *Discrimination in Labor Markets,* Princeton, Princeton University Press.

[Authors] 2013

Barone, C. and Ortiz, L. (2011). Overeducaction among University Graduates: A Comparative Analysis of its Incidence and the Importance of Higher Education Differentiation, *Higher Education*, 61, 325-327.

Battu, H., Belfield C. R., and Sloane P. J. (1999). Overeducation Among Graduates: A Cohort View, *Education Economics*, 7(1), 21-38.

Bauer, T. (2002). Educational mismatch and wages: a panel analysis, *Economics of Education Review* 21(3), 221-229.

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

Becker, G. S. (1964). *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education.* Chicago, University of Chicago Press.

Becker, G.S. (1965). A Theory of the Allocation of Time, *The Economic Journal*, 75(299), 493-517.

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

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

Blau, F. D., Ferber, M. A. and Winkler A. E. (2010). *The economics of women, men and work*, Upper Saddle River, NJ: Prentice Hall.

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

Bobbitt-Zeher, D. (2007). The gender income gap and the role of education, *Sociology of Education* 80 (1), 1-22.

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

Boll, C., Leppin J. S., Rossen A., and Wolf A. (2016). Magnitude and Impact Factors of the Gender Pay Gap in EU Countries, Tech. Rep., European Commission-Directorate-General for Justice, European Union.

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

Buchel F. and van Ham M. (2003). Overeducation, regional labor markets, and spatial flexibility, *Journal of Urban Economics*, 53(3), 482-493.

Capsada-Munsech, Q. (2015). The Role of Social Origin and Field of Study on Graduates' Overeducation: The Case of Italy, *Higher Education*, 69(5), 779-807.

Caroleo, F. E. and Pastore, F. (2016). Overeducation: A disease of the school-to-work transition systems, in G. Coppola and N. O'Higgins (eds) *Youth and the Crisis: Unemployment, education and health in Europe*, London, Routledge.

Ceccarelli, C. and Cutillo, A. (2015). Composizione occupazionale e differenziale salariale di genere, *Rivista Italiana di Economia Demografia e Statistica* LXIX(1), 31-38.

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

Chevalier, A. (2003). Measuring over-education, *Economica* 70(3), 509-531.

Correll, S. J. (2004). Constraints into preferences: Gender, status and emerging career aspirations, *American Sociological Review* 69:93-113.

Croce, G. and Ghignoni, E. (2015). Educational mismatch and spatial flexibility in Italian local labour markets, *Education Economics*, 23(1), 25-46.

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.

De la Rica, S., Dolado, J. J., and Llorens, V. (2008). Ceilings or Floors?: Gender Wage Gaps by Education in Spain, *Journal of Population Economics* 21(3), 751-776.

Dolton P.J., and Silles, M.A. (2008). The effects of over-education on earnings in the graduate labor market, *Economics of Education Review* 27(2), 125-139

Dolton, P. and Vignoles, A. (2000). The incidence and effects of overeducation in the UK graduate labour market, *Economics of Education Review* 19(2), 179-198.

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

Eagly, A. H., and Steffen V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles, *Journal of Personality and Social Psychology* 46(4), 735-54.

England, P. (2010). The gender revolution: Uneven and stalled. *Gender & Society* 24(2), 149-66.

England, P., and Li, S. (2006). Desegregation stalled: The changing gender composition of college majors, 1971-2002. *Gender & Society* 20(5), 657-77.

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

Eurostat (2016a) - http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_statistics Accessed 27 May 2016.

Eurostat (2016b) - http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics Accessed 27 May 2016

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

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

Freeman, R.B. (1976). *The Overeducated American*. New York, Academic Press.

Green, F., McIntosh, S. and Vignoles, A. (2002). The utilization of education and skills: evidence from Great Britain, *The Manchester School* 70(6), 792–811.

Green, F. and McIntosh S. (2007). Is there a Genuine Under-Utilization of Skills amongst the Over-Qualified? *Applied Economics* 39(4), 427-439.

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.

Groot, W and Maassen van den Brink, H (2003), The dynamics of skill mismatches in the Dutch labour market, in F. Büchel, A. Mertens and A. de Grip (eds.) *Overeducation in Europe*. Edward Elgar Publishing.

Hakim, C. (2000). *Work-Lifestyle Choices in the 21st Century: Preference Theory*, Oxford, Oxford University Press.

Hartog, J. (2000). Overeducation and earnings. Where are we, where should we go?, *Economics of Education Review* 19(2), 131-147.

Kedir A., Kyrizi A. and Martinez-Mora F. (2012) Signalling and Productivity Effects of Overeducation: Is It Really a Waste of Resources? Working Paper No. 12/19, 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.

Leuven, E. and Oosterbeek, H. (2011). Overeducation and mismatch in the labor market, IZA Discussion paper series, No. 5523.

Li, I. W., and Miller, P. W. (2012). Gender Discrimination in the Australian Graduate Labour Market, IZA Discussion Paper 6595.

Livanos, I. and I. Nunez (2012) The effect of higher education on the gender wage gap, *International Journal of Education Economics and Development*, 3(1), 33-47.

McGoldrick, K. M. 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 Jessica Bennett J. (2007). Overeducation in the graduate labour market: A quantile regression approach, Economics of Education Review, 2007, vol. 26, issue 5, 521-531.

McGuinness, S. and Sloane, P. J. (2010). Educational/skill mismatch nel mercato del lavoro dei laureate: un'analisi comparata, in AlmaLaurea, *XII Rapporto sulla condizione occupazionale dei laureati. Investimenti in capitale umano nel future di Italia ed Europa*, Bologna, Il Mulino.

Meng, C., and 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), S76-S108.

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

OECD. 2009. Education at a glance. OECD Indicators. OECD Publishing, Paris. http://dx.doi.org/10.1787/eag-2009-en

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.

Ortiz, L. (2010). Not the right job, but a secure one. *Work, Employment & Society*, 24(1) 47-64.

Ortiz, L. and Kucel A. (2008). Do Fields of Study matter for Over-Education? The Cases of Spain and Germany, *International Journal of Comparative Sociology*, 49(4-5), 305-327.

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., and Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on cultural beliefs in social relations, *Gender & Society* 18(4), 510-31.

Robst, J. (2007). Education and job match: The relatedness of college major and work, *Economics of Education Review*, 26(4), 397-407.

Sattinger, M. (1993): Assignment models of the distribution of earnings, *Journal of Economic Literature* 31(2), 851-880.

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

Sicherman, N. (1991). Overeducation in the Labor Market, *Journal of Labor Economics* 9(2), 101-22.

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

Sloane, P. J. (2014). Overeducation, skill mismatches, and labor market outcomes for college graduates. IZA World of Labor 88.

Solberg, E. and Laughlin, T. (1995). The gender pay gap, fringe benefits, and occupational crowding, *Industrial and 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, A.M. (1973). Job Market Signaling, *Quarterly Journal of Economics*, 87(3), 355-374.

Stiglitz, J. E. (1973). Approaches to the Economics of Discrimination, *American Economic Review Papers and Proceedings*, 63(2), 287-95.

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(B), 235-282.

Thurow, L. C. (1975). *Generating Inequality: Mechanisms of distribution in the U. S. economy*, New York, Macmillan Interactive Publishing.

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

Weichselbaumer, D. and Winter-Ebmer, R. (2005). A Meta-Analysis of the International Gender Wage Gap, *Journal of Economic Surveys* 19(3), 479-511.

Wolbers, M. H. J. (2003). Job mismatches and their labour-market effects among school-leavers in Europe, *European Sociological Review*, 19(3), 249-266.

Notes

1. See Boll et al. (2016) for a survey.

2. See Authors (2013) for a survey.

3. As in Cutillo and Di Pietro (2006).

4. Following the literature (e.g. Leuven and Oosterbeek 2011) we exclude from the sample individuals with less than high school diploma because their likelihood to be overeducated is very low. Other studies on overeducation in Italy find that about 30.0 per cent of university graduates are overeducated. For example; Cutillo and Di Pietro (2006); Croce and Ghignoni (2015).

5. This distinction has relevant policy implications in Italy, where a high percentage of the population claims to be overqualified but not simultaneously overskilled. This "*somehow questions the ability of the education system to provide the necessary skills for the jobs currently available in the labor marke*t" (Flisi et al. 2014, p. 5)

6. In the years 2004-2012, the GPG calculated on the net hourly wage for dependent employment varies from 4.8 per cent to 7.9 per cent, while the gross monthly wage ranges between 22.4 per cent and 25.8 per cent (Ceccarelli and Cutillo, 2015).

7. In the same year, the raw GPG was 16.2 per cent in Europe and 7.2 per cent in Italy (Eurostat, 2016b).

8. Freeman (1976) was one of the first economists to express concern about the potential problem of overinvestment. He found that in a situation of excess supply, graduates might increasingly be forced to accept non-graduate jobs.

9. This value is slightly lower than that estimated by Eurostat (2016b) in the period 2005-2014 (5.6 per cent). This is because we keep also the self-employed, while Eurostat considers only employees in enterprises with more than 10 employees.

10. Also, Cutillo and Di Pietro (2006) find a lower pay gap for properly educated workers relative to overeducated workers in Italy.

11. *Overfem* is the interaction of the dummies female and over. The dummy over takes the value 1 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.

12. The coefficients of *female* in column (2) and (3) of Table 3 represent the negative of the GPGs, i.e. $\beta_i^{female} = -(\Delta^{\mathrm{GPG_i}})$, where $\mathrm{i} = \mathrm{Over}$, Proper and $\Delta^{\mathrm{GPG_i}} = \overline{\ln(W_M)} - \overline{\ln(W_F)}$, for the respective subsample.

13. The full regression output is shown in Table C1 in Appendix C. Table C2 in Appendix C shows the regression output by gender and over- or proper education.

14. Cutillo and Di Pietro (2006) find a wage penalty of 4.4 per cent associated with overeducation in a sample of university graduates; McGuinness and Sloane (2010) find a wage penalty of 4 per cent for young university graduates. Caroleo and Pastore (2016) find a wage penalty of 12 per cent for university graduates five years after graduation.

15. In Section 5.4, 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.

16. For example, Cutillo and Di Pietro (2006) also find a negative and statistically significant effect of women on the overeducation probability.

17. Table D1, Appendix D, shows the full regression output for the bivariate probit of the participation and overeducation selection equations.

18. We present in Table D2, Appendix D, the full regression output with the selection correction terms. In the following, for notational simplicity; $\bar{\lambda} = \lambda$ and $\hat{\delta} = \delta$.

Tables

Table 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	9.367	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

Table 2: Log of Net Hourly Wages in Euro and Raw GPG in per cent

	(1) Full Sample	(2) Overeducated Sample	(3) 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 per cent	4.7	9.6	3.5

Table 3: OLS Estimates of Log Hourly Wages with Dummies *female, overeducation* and *overfem*

Variables	(1) Full Sample	(2) Overeducated Sample	(3) 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 4: OLS Estimates of Log Hourly Wages

Variables	(1) Full Sample	(2) Overeducated Sample	(3) 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

Table 5: 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	1.000
P-value	
H1 : Difference < 0	0.000
P-value	

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

Table 5: Risk of Overeducation, by Gender
(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	-0.394***	-0.222***
	(0.018)	(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

Table 6: Decomposition of the GPG by Sample

	(1) Overeducated Sample	(2) 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<0.1

Table 6: Decomposition of the GPG versus Other Pay Gaps

	(1) Men vs. Women	(2) Properly Educated vs. Overeducated Individuals	(3) Public vs. Private Sector	(4) Full time vs. Part-time	(5) Recruitment by Public Contest vs. Recruitment without Public Contest
Differential					
$\overline{\ln(W_{Group\ 0})}$	2.135***	2.194***	2.334***	2.130***	2.373***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003501)
$\overline{\ln(W_{Group\ 1})}$	2.088***	1.938***	1.949***	2.018***	1.971***
	(0.003)	(0.004)	(0.003)	(0.006)	(0.002936)
Difference	0.047***	0.256***	0.384***	0.111***	0.402***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.004569)
Decomposition					
Endowments	-0.044***	0.200***	0.315***	0.171***	0.286***
	(0.005)	(0.004)	(0.009)	(0.004)	(0.010879)
Coefficients	0.091***	0.056***	0.070***	-0.059***	0.116***
	(0.006)	(0.005)	(0.010)	(0.006)	(0.011515)
Coefficients in % (Absolute Value)					
Observations	43,178	43,178	43,178	43,178	43,178

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

Table 8: Bivariate Probit Results: Instruments and Correlation Coefficient for the Participation and Overeducation Decision

	(1) Female Sample	(2)	(3) Male Sample	(4)
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)
ρ	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 9: Selection Variables, Definition and Values

Panel A		
Overeducated Sample	(1) Female Sample	(2) Male Sample
$\lambda_{PA}^{\text{Over}}$ measures the selection bias from the *participation decision* for overeducated individuals	0.094** (0.044)	1.318** (0.546)
$\lambda_{OV}^{\text{Over}}$ measures the selection bias from the *overeducation decision* for overeducated individuals	0.498*** (0.135)	0.478*** (0.120)
Observations	7,481	6,775
Panel B		
Properly Educated Sample	(1) Female Sample	(2) Male Sample
$\lambda_{PA}^{\text{Proper}}$ measures the selection bias from the *participation decision* for properly educated individuals	0.049** (0.022)	0.047 (0.157)
$\lambda_{OV}^{\text{Proper}}$ 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 10: Oaxaca-Blinder Decomposition of Gender Log Hourly Wage Gap adjusted for Double Selection by Overeducation

	(1) Overeducated Sample	(2) Properly Educated Sample
Difference	0.096***	0.035***
	(0.006)	(0.008)
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

Appendices

Appendix A

Table A1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Full Sample		Overeducated Sample		Properly Educated Sample		Full Sample Female Sample		Overeducated Sample Female Sample		Properly Educated Sample Female Sample		Full Sample Male Sample		Overeducated Sample Male Sample		Properly Educated Sample Male Sample	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Log of Hourly Wage	2.075	0.511	1.921	0.463	2.176	0.516	2.060	0.521	1.876	0.480	2.166	0.513	2.093	0.500	1.964	0.441	2.188	0.519
Schooling	13.06	2.644	11.83	2.952	13.86	2.064	13.35	2.536	12.14	2.967	14.05	1.934	12.74	2.723	11.54	2.907	13.62	2.191
Maximum_D_Mark	0.0452	0.208	0.0172	0.130	0.0634	0.244	0.0563	0.231	0.0246	0.155	0.0746	0.263	0.0327	0.178	0.00999	0.0994	0.0495	0.217
Eng Skill	0.346	0.476	0.287	0.452	0.385	0.487	0.360	0.480	0.311	0.463	0.387	0.487	0.332	0.471	0.263	0.440	0.382	0.486
Exper	18.36	12.73	17.12	12.81	19.17	12.62	16.95	12.04	15.20	11.71	17.97	12.11	19.94	13.29	18.97	13.52	20.65	13.07
Tenure	12.96	11.65	10.59	10.68	14.50	12.00	11.69	10.95	8.796	9.222	13.36	11.50	14.39	12.24	12.32	11.65	15.91	12.44
Extra Hours	0.492	0.500	0.484	0.500	0.498	0.500	0.462	0.499	0.455	0.498	0.466	0.499	0.527	0.499	0.512	0.500	0.537	0.499
Married	0.558	0.497	0.507	0.500	0.591	0.492	0.576	0.494	0.527	0.499	0.604	0.489	0.538	0.499	0.487	0.500	0.576	0.494
Kids	0.574	0.495	0.535	0.499	0.599	0.490	0.599	0.490	0.564	0.496	0.619	0.486	0.545	0.498	0.506	0.500	0.573	0.495
Kids 3	0.0775	0.267	0.0754	0.264	0.0789	0.270	0.103	0.304	0.102	0.303	0.103	0.304	0.0491	0.216	0.0496	0.217	0.0488	0.215
Age	39.98	12.73	37.64	12.38	41.50	12.73	39.05	11.90	36.42	11.23	40.56	12.02	41.03	13.53	38.81	13.29	42.67	13.47
Observations	74,540		29,329		45,211		39,420		14,407		25,013		35,120		14,922		20,198	

Appendix B

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 regression and present the decomposition expression when it is accounted for double selection (the Oaxaca-Blinder model with Double Selection) into the sample.

B.1. Wage equation

Consider the following model of wage determination:

$$\ln(\mathrm{W}_i) = X_i'\beta + \gamma S_i \quad + \varepsilon_i \tag{1}$$

with $i = 1, \dots, N$ and where $\ln(\mathrm{W}_i)$ is the logarithm of net hourly wages, β 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[4] and γ is the corresponding coefficient**.** The error term is described by ε_i.
In order to analyze the effect of overeducation on wages, the wage model (1) is evaluated at the mean by OLS, separately for men and women:

$$\overline{\ln(W_G)} = \bar{X}_G'\hat{\beta}_G \; + \gamma_G S_{\ G} \tag{2}$$

with $G = M, F;\ G = M$ identifies the male sample and $G = \ F$ identifies the female sample. $\overline{\ln(W_G)}$ is the logarithm of net hourly wages evaluated at the mean, $\hat{\beta}_G$ 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.

[4]- i.e. $S_i = 1$ if the individual is overeducated and zero otherwise.

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

$$\overline{\ln(W_G)}^{\mathrm{m}} = \bar{X}_G^{\mathrm{m}'} \hat{\beta}_G^{\mathrm{m}} \tag{3}$$

with $\mathrm{m} = \mathrm{Over}, \mathrm{Proper}$; $\mathrm{m} = \mathrm{Over}$ is for overeducated individuals and $\mathrm{m} = \mathrm{Proper}$ is for properly educated individuals.

B.2. The Oaxaca-Blinder Model

Starting from equation (3) and using the implicit assumptions in Oaxaca (1973) and Blinder (1973) we decompose the wage differential in three distinct parts; endowments, coefficients and interaction components[5] :

$$\begin{aligned} \overline{\ln(W_M)}^{\mathrm{m}} - \overline{\ln(W_F)}^{\mathrm{m}} &= \bar{X}_M^{\mathrm{m}'} \hat{\beta}_M^{\mathrm{m}} - \bar{X}_F^{\mathrm{m}'} \hat{\beta}_F^{\mathrm{m}} \\ &= (\bar{X}_M^{\mathrm{m}'} - \bar{X}_F^{\mathrm{m}'})\, \hat{\beta}_F^{\mathrm{m}} + \bar{X}_F^{\mathrm{m}'} (\hat{\beta}_M^{\mathrm{m}} - \hat{\beta}_F^{\mathrm{m}}) \\ &\quad + (\bar{X}_M^{\mathrm{m}'} - \bar{X}_F^{\mathrm{m}'})\, (\hat{\beta}_M^{\mathrm{m}} - \hat{\beta}_F^{\mathrm{m}}) \end{aligned} \tag{4}$$

where $\overline{\ln(W_G)}^{m}$ is again the logarithmic net wage evaluated at the mean for the respective subsample, $G = M, F$ and $\mathrm{m} = \mathrm{Over}, \mathrm{Proper}$, with $\bar{X}_G^{\mathrm{m}}$ and $\hat{\beta}_G^{\mathrm{m}}$ being $(\mathrm{K} \times 1)$ 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*" 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

[5] - 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 for the three-fold model to the endowment component. Similarly, the use of the pay structure for the low earning group in the simple decomposition is equivalent to adding the interaction term for the three-way model to the discrimination component (Li and Miller 2012).

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, i.e. differences in remuneration across gender, this may indicate the presence of gender discrimination.

The last term in equation (4) is the "*interaction*" effect that takes the simultaneous existence of differences in endowments and coefficients by gender into account. A negative interaction effect means that characteristics for which women have a lead over men pay off less for women than for men given higher male remuneration (Winsborough and Dickinson 1971).

B.3. Selection Rules

Endogeneity arises from correlation of S_i with the error term ε_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 (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 or not and one for the decision to accept a wage offer for which the individual is overeducated or not. The selection rules are described by the following relations:

Participation Selection: $$Y^*_{i\,Parti} = Z'_i\gamma + \mathrm{u}_{i\,Parti} \quad (5)$$

Overeducation Selection: $$Y^*_{i\,Over} = Q'_i\alpha + \mathrm{u}_{i\,Over} \quad (6)$$

where $\mathrm{Y}^*_{i\,Parti}$ represents the unobserved indexes of utility that individual i uses to make the decision to work or not and $\mathrm{Y}^*_{i\,Over}$ represents the unobserved indexes of utility that individual i uses to make the decision to be overeducated or not; with Z_i and Q_i being $(\mathrm{K}_z \times 1)$ and $(\mathrm{K}_Q \times 1)$ vectors of explanatory variables, respectively, and u_i are assumed to be $N(0,1)$ with $cov(\mathrm{u}_{Parti}, \mathrm{u}_{Over}) = \rho$.

Each equation describing the selection rules should include at least one variable that influences the decision to participate in the labor market but is uncorrelated with wages, and at least one variable that influences the overeducation decision but is uncorrelated with wages. Moreover, these two instrumental variables must be mutually independent. We use a dummy for "participation of the partner" as an instrument to identify the participation decision. If one partner is inactive, the probability that the other works is higher, in order to ensure a family income. The same holds for single households; singles have higher probability to participate in the labor market in order to ensure their income. The dummy for "participation of the partner" is one if the partner of the individual works and zero if the partner does not work the individual has no partner. This is in line with the finding in the literature that the decision to work is strongly correlated with spousal income (Devereux, 2004; Bar *et al.*, 2015). Additionally, we use the "Age5064" indicating whether the individual is aged between 50 and 64 years or not as a proxy for the last stage of the career. Both variables are excluded from the earnings and overeducation equation as they are assumed to affect the individual propensity to participate in the labour market only.

For identification of the overeducation decision, we use variables measuring the "level of satisfaction" with the job as instruments. The probability of accepting a job as overeducated increases if the work is satisfactory and provides career opportunities. Yet, it is exogenous to the wage level. In particular, the level of satisfaction with the retribution, development and work prospective at the current job are included in the overeducation selection equation only. Moreover, we add the dummy identifying whether the individual holds a citizenship other than Italian to the overeducation selection rule only as foreigners might have higher propensity to accept wage offers as overeducated.

After having estimated the bivariate probit for the participation and overeducation selection equations, we interpret the estimated correlation between the error terms of these two binary equations, ρ.

The probabilities of observing a positive labor income as overeducated or properly educated worker are given below:

$$\Pr\left(Y^*_{Parti} > 0, Y^*_{Over} > 0\right) = \Pr\left(u_{Parti} > Z'\gamma,\ u_{Over} > -Q'\alpha\right) = G(Z'\gamma, Q'\alpha, \rho) \tag{7}$$

$$\Pr(Y^*_{Parti} > 0, Y^*_{Over} \leq 0) = \Pr(u_{Parti} > Z'\gamma,\ u_{Over} \leq -Q'\alpha) = G(Z'\gamma, -Q'\alpha, -\rho) \tag{8}$$

where G(.) is the standard bivariate normal distribution and ρ is the correlation coefficient between the two selection rules. 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 the overeducated sample, m = Over:

$$\lambda^{\text{Over}}_{Parti} = \frac{f(Z'\gamma)F\left[\frac{Q'\alpha - \rho Z'\gamma}{\sqrt{(1-\rho^2)}}\right]}{G(Z'\gamma, Q'\alpha, \rho)} \tag{9}$$

$$\lambda^{\text{Over}}_{Over} = \frac{f(Q'\alpha)F\left[\frac{Z'\gamma - \rho Q'\alpha}{\sqrt{(1-\rho^2)}}\right]}{G(Z'\gamma, Q'\alpha, \rho)} \tag{10}$$

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

$$\lambda^{\text{Proper}}_{Parti} = \frac{f(Z'\gamma)F\left[-\frac{Q'\alpha - \rho Z'\gamma}{\sqrt{(1-\rho^2)}}\right]}{G(Z'\gamma, -Q'\alpha, -\rho)} \tag{11}$$

$$\lambda^{\text{Proper}}_{Over} = \frac{-f(Q'\alpha)F\left[\frac{Z'\gamma - \rho Q'\alpha}{\sqrt{(1-\rho^2)}}\right]}{G(Z'\gamma, -Q'\alpha, -\rho)} \tag{12}$$

where f (.) is the standard normal density, while F(.) is the standard normal distribution and ρ is the correlation coefficient between the two selection rules.

Adding the selection terms λ_{Parti} and λ_{Over} to the earnings equations in (3) allows us to consistently estimate the earnings for the overeducated and properly educated subsamples, respectively (Lee 1983; Tunali 1986) and yields the following augmented model of wage determination:

$$\overline{\ln(W_G)}^{\mathrm{m}} = \bar{X}_G^{\mathrm{m}\prime}\hat{\beta}_G{}^{\mathrm{m}} + \hat{\delta}_{\mathrm{G},Parti}{}^{\mathrm{m}}\bar{\lambda}_{\mathrm{G},Parti}{}^{\mathrm{m}} + \hat{\delta}_{\mathrm{G},Over}{}^{\mathrm{m}}\bar{\lambda}_{\mathrm{G},Over}{}^{\mathrm{m}} \tag{13}$$

where $\mathrm{m} = \mathrm{Over}, \mathrm{Proper}$ and $G = M, F$.

B.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 in the sample, the decomposition expression (4) becomes the following:

$$\begin{aligned}
&\overline{\ln(W_M)}^{\mathrm{m}} - \overline{\ln(W_F)}^{\mathrm{m}} = \bar{X}_M^{\mathrm{m}\prime}\hat{\beta}_M^{\mathrm{m}} - \bar{X}_F^{\mathrm{m}\prime}\hat{\beta}_F^{\mathrm{m}} \\
&= (\bar{X}_M^{\mathrm{m}\prime} - \bar{X}_F^{\mathrm{m}\prime})\,\hat{\beta}_F^{\mathrm{m}} + \bar{X}_F^{\mathrm{m}\prime}(\hat{\beta}_M^{\mathrm{m}} - \hat{\beta}_F^{\mathrm{m}}) + \\
&(\bar{X}_M^{\mathrm{m}\prime} - \bar{X}_F^{\mathrm{m}\prime})\,(\hat{\beta}_M^{\mathrm{m}} - \hat{\beta}_F^{\mathrm{m}}) + \\
&(\hat{\delta}_{M,Parti}{}^{\mathrm{m}}\bar{\lambda}_{M,Parti}{}^{\mathrm{m}} - \hat{\delta}_{F,Parti}{}^{\mathrm{m}}\bar{\lambda}_{F,Parti}{}^{\mathrm{m}}) + \\
&(\hat{\delta}_{M,Over}{}^{\mathrm{m}}\bar{\lambda}_{M,Over}{}^{\mathrm{m}} - \hat{\delta}_{F,Over}{}^{\mathrm{m}}\bar{\lambda}_{F,Over}{}^{\mathrm{m}})
\end{aligned} \tag{15}$$

Despite, the endowments, coefficients and interaction 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 leading to the decision to participate in the labor market or not and to accept a job offer for which he or she is overeducated or not.

Appendix C

Table C1: OLS Estimates of Log of Net Hourly Wages

Variables	(1) Full Sample	(2) Female Sample	(3) Male Sample	(4) Full Sample	(4) Overeducated Sample	(5) Properly Educated Sample
overeducation	-0.024***	-0.053***	-0.041***			
	(0.004808)	(0.005391)	(0.004995)			
oversex	-0.048***					
	(0.006569)					
female	-0.040***			-0.060***	-0.091***	-0.043***
	(0.004271)			(0.003389)	(0.005344)	(0.004391)
Schooling	0.040***	0.042***	0.036***	0.041***	0.025***	0.055***
	(0.000854)	(0.001300)	(0.001128)	(0.000848)	(0.001179)	(0.001249)
Maximum_D_Mark	0.058***	0.048***	0.067***	0.058***	0.039**	0.043***
	(0.008696)	(0.010291)	(0.016003)	(0.008714)	(0.019711)	(0.009655)
Eng_Skill	0.014***	0.015***	0.015***	0.014***	0.029***	0.009*
	(0.003782)	(0.005115)	(0.005620)	(0.003785)	(0.006349)	(0.004683)
Exper	0.017***	0.016***	0.020***	0.017***	0.012***	0.020***
	(0.000646)	(0.000899)	(0.000938)	(0.000646)	(0.000979)	(0.000858)
Exper2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000014)	(0.000019)	(0.000020)	(0.000014)	(0.000021)	(0.000018)
Tenure	0.003***	0.003***	0.003***	0.004***	0.003***	0.003***
	(0.000239)	(0.000339)	(0.000339)	(0.000239)	(0.000356)	(0.000324)
Kids	0.004	0.003	0.009	0.005	-0.003	0.011*
	(0.004460)	(0.006057)	(0.006657)	(0.004467)	(0.006908)	(0.005784)
Kids_3	0.033***	0.036***	0.022**	0.034***	0.029***	0.034***
	(0.006384)	(0.008076)	(0.010449)	(0.006394)	(0.010446)	(0.008045)
Married	0.042***	0.042***	0.039***	0.042***	0.047***	0.042***
	(0.004528)	(0.005822)	(0.007489)	(0.004539)	(0.007083)	(0.005834)
Intermed_Prof	0.023***	0.036***	0.014**	0.039***	0.057***	0.010
	(0.004275)	(0.006899)	(0.005445)	(0.004180)	(0.005826)	(0.006406)
Manager	0.182***	0.192***	0.173***	0.202***	0.159***	0.147***
	(0.006225)	(0.009502)	(0.008448)	(0.006109)	(0.011971)	(0.008147)
Public_Sector	0.121***	0.150***	0.089***	0.130***	0.059***	0.138***
	(0.003887)	(0.005467)	(0.005521)	(0.003852)	(0.006289)	(0.004884)
Home_Time	0.008***	0.007***	0.008***	0.007***	0.005***	0.009***
	(0.000398)	(0.000516)	(0.000637)	(0.000398)	(0.000573)	(0.000548)
North	0.056***	0.045***	0.067***	0.056***	0.091***	0.039***
	(0.003901)	(0.005678)	(0.005383)	(0.003907)	(0.006253)	(0.004968)
Centre	0.024***	0.013*	0.036***	0.023***	0.055***	0.008
	(0.004659)	(0.006739)	(0.006429)	(0.004666)	(0.007442)	(0.005931)
Homeowner	0.048***	0.052***	0.043***	0.049***	0.053***	0.047***
	(0.004185)	(0.006006)	(0.005809)	(0.004189)	(0.006061)	(0.005729)
Extra_Hours	-0.014***	-0.022***	-0.008*	-0.013***	-0.006	-0.019***
	(0.003219)	(0.004477)	(0.004610)	(0.003222)	(0.005051)	(0.004139)
Constant	1.082***	1.006***	1.110***	1.043***	1.281***	0.852***
	(0.013784)	(0.020835)	(0.018359)	(0.013138)	(0.018789)	(0.019048)
Observations	74,540	39,420	35,120	74,540	29,329	45,211
R-squared	0.276	0.272	0.283	0.274	0.135	0.293

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

Appendix D

Derivation of the GPG in Percent:

$$
\begin{aligned}
\frac{W_M - W_F}{W_M} &= \frac{W_M}{W_M} - \frac{W_F}{W_M} \\
&= \ln(1) - [\ln(W_F) - \ln(W_M)] \\
&= 0 + \ln(W_M) - \ln(W_F) \\
&= \ln(W_M) - \ln(W_F) \\
&= 0.033 \\
&\Rrightarrow 0.033 * 100 = 3.3\%
\end{aligned}
$$

Appendix E

Table E1. Variable Definitions

Variable Name	**Definition**
Dependent Variables	
Participation Decision (*P*articipation)[6]	One if the respective individual decided to participate in the labor market, zero otherwise
Overeducation Decision (*Overeducation*)	One if the respective individual is overeducated, zero otherwise
Net Hourly Wages	Hourly wages in Euros and net of taxes and social security contributions
Log Hourly Wage ($ln(W)$)	The natural log of net hourly earnings. Wages are in Euros and net of taxes and social security contributions
Independent Variables	
Female	One if the respective individual is a woman, zero otherwise
Oversex	Interactive effect of *Overeducation* and *Female*, i.e. one if individual is overeducated in current job and is female, zero otherwise
Not Overeducated (*Proper*)	One if individual's education level is a necessary pre-request to perform the job, zero otherwise
Actual Work Experience (*Exper*)	Actual work experience in years
Actual Work Experience Squared (*Exper2*)	Actual work experience squared
Tenure	Number of years worked for current employer
Education (*Schooling*)	Number of years of schooling completed
Maximum Degree Mark (*Maximum_D_Mark)*	One if individual achieved the maximum degree mark, zero otherwise
English Skills (*Eng_Skill*)	One if individual is able to communicate in English, zero otherwise
Public Firm (*Public_Sector*)	One if firm is a publicly owned firm, zero otherwise

[6] - In parentheses name of variable if different from variable label.

Extra Hours (*Extra_Hours*)	One if number of hours worked exceed hours stabilized in contract, zero otherwise
Manager	Intellectual professions; scientific, and highly specialized occupations
Intermediate Professions (*Intermed_Prof*)	Intermediary positions in commercial, technical or administrative sectors, health services, technicians.
Age	Age of respective individual (in years)
Age5064	One if age is between 50 and sixty-four years, zero otherwise
Married	One if married, zero otherwise
Children (*Kids*)	One if individual has at least one child, zero otherwise
Youngest Child Younger than Four Years (*Kids_3*)	One if age of youngest child is less or equal to three years, zero otherwise
Metropolitan Area (*City*)	One if the respective individual is located in a metropolitan area, zero otherwise
Northern Region (*North*)	One if individual lives and works in the North of Italy, zero otherwise
Central Region (*Centre*)	One if individual lives and works in the Centre of Italy, zero otherwise
Retribution of Work (*Retribution_Work*)	Level of satisfaction with the retribution in the current job, ∈ (0-4), where *low* = 4 and *high* = 0
Development of Work (*Develop_Work*)	Level of satisfaction with development of the current job, ∈ (0-4), where *low* = 4 and *high* = 0
Working Prospectives (*Prosp_Work*)	Level of satisfaction with the work perspectives in the current job, ∈ (0-4), where *low* = 4 and *high* = 0
Homeowner	One if individual owns a house, zero otherwise
Home_Time	Number of Hours spent at home during a workday on average
Partner Works (*Partner_Works*)	One if partner is employed, zero otherwise
Not Italian (*No_Italian*)	One if individual is not Italian, zero otherwise
ρ	Correlation between the error term of the two binary selection equations (participation and overeducation decision)

Selection Variables	
$\lambda_{Parti}^{\text{Over}}$	Measures the selection bias from the *participation decision* for those selected into overeducation
$\lambda_{Over}^{\text{Over}}$	Measures the selection bias from the *overeducation decision* for those that self-selected themselves in *overeducation*
$\lambda_{Parti}^{\text{Proper}}$	Measures the selection bias from the *participation decision* for those not selected into *overeducation*
$\lambda_{Over}^{\text{Proper}}$	Measures the selection bias from the *overeducation decision* for the properly educated individuals

Appendix F

The coefficient of the dummy variable "o*vereducation*" is estimated to be -0.024. Hence, overeducation lowers wages by 2.37%.

$$\ln\left(W_G^{\text{Over}}\right) - \ln\left(W_G^{\text{Proper}}\right) = \beta^{overeducated}$$

$$\ln\left(\frac{W_G^{\text{Over}}}{W_G^{\text{Proper}}}\right) = \beta^{overeducated}$$

$$\frac{W_G^{\text{Over}}}{W_G^{\text{Proper}}} = e^{\beta^{overeducated}}$$

$$\frac{W_G^{\text{Over}} - W_G^{\text{Proper}}}{W_G^{\text{Proper}}} = e^{\beta^{overeducated}} - 1$$

with *G=M, F.*

Appendix G

Decomposition of Pay Gaps between *Group 0* and *Group 1:*

$$\begin{aligned}\overline{\ln(W_{Group\ 0})} - \overline{\ln(W_{Group\ 1})} &= \bar{X}'_{Group\ 0}\hat{\beta}_{Group\ 0} - \bar{X}'_{Group\ 1}\hat{\beta}_{Group\ 1} \\ &= (\bar{X}'_{Group\ 0} - \bar{X}'_{Group\ 1})\,\hat{\beta}_{Group\ 1} \\ &+ \bar{X}'_{Group\ 1}(\hat{\beta}_{Group\ 0} - \hat{\beta}_{Group\ 1}) \\ &+ (\bar{X}'_{Group\ 0} - \bar{X}'_{Group\ 1})\,(\hat{\beta}_{Group\ 0} - \hat{\beta}_{Group\ 1})\end{aligned}$$

Appendix H

Table H1: OLS Estimates of Log Net Hourly Wages with Selection Terms

Variables	(1) Overeducated Sample Female Sample	(2) Male Sample	(3) Properly Educated Sample Female Sample	(4) Male Sample
Schooling	0.016***	0.014***	0.078***	0.070***
	(0.005165)	(0.004799)	(0.004094)	(0.005040)
Maximum_D_Mark	0.020	0.023	0.037***	0.058***
	(0.022622)	(0.040814)	(0.011494)	(0.017559)
Eng_Skill	0.031***	0.008	0.010	0.011
	(0.009541)	(0.009804)	(0.006305)	(0.007898)
Exper	0.011***	0.012***	0.019***	0.024***
	(0.001581)	(0.001324)	(0.001193)	(0.001338)
Exper2	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000033)	(0.000029)	(0.000025)	(0.000027)
Tenure	0.003***	0.004***	0.003***	0.002***
	(0.000581)	(0.000451)	(0.000421)	(0.000500)
Kids	-0.009	0.008	0.015**	0.019**
	(0.010139)	(0.009842)	(0.007514)	(0.009344)
Kids_3	0.023	0.058***	0.035***	0.024
	(0.014401)	(0.017333)	(0.010069)	(0.016030)
Married	0.038***	0.066***	0.038***	0.064***
	(0.012345)	(0.015929)	(0.008598)	(0.014327)
Intermed_Prof	0.060***	0.047***	0.059***	-0.005
	(0.008911)	(0.007721)	(0.011606)	(0.007771)
Manager	0.175***	0.133***	0.187***	0.137***
	(0.018684)	(0.015543)	(0.013450)	(0.010927)
Public_Sector	0.069***	0.047***	0.172***	0.097***
	(0.009493)	(0.008339)	(0.006600)	(0.007242)
Home_Time	0.004***	0.004***	0.010***	0.009***
	(0.000814)	(0.000940)	(0.000738)	(0.000952)
North	0.083***	0.114***	0.022**	0.056***
	(0.014646)	(0.010059)	(0.009936)	(0.008663)
Centre	0.047***	0.074***	-0.005	0.022**
	(0.013499)	(0.010389)	(0.009420)	(0.009092)
Homeowner	0.056***	0.050***	0.056***	0.039***
	(0.009128)	(0.008108)	(0.007969)	(0.008279)
Extra_Hours	-0.012	-0.004	-0.027***	-0.009
	(0.007577)	(0.006704)	(0.005507)	(0.006260)
$\lambda_{F,Parti}^{Over}$	-0.036			
	(0.036939)			
$\lambda_{F,Over}^{Over}$	0.044			
	(0.027046)			
$\lambda_{M,Parti}^{Over}$		0.102***		
		(0.031905)		
$\lambda_{M,Over}^{Over}$		0.142***		
		(0.033907)		
$\lambda_{F,Parti}^{Proper}$			-0.020	
			(0.025131)	
$\lambda_{F,Over}^{Proper}$			-0.170***	
			(0.026876)	
$\lambda_{M,Parti}^{Proper}$				0.047
				(0.032434)
$\lambda_{M,Over}^{Proper}$				-0.146***
				(0.034270)
Constant	1.345***	1.212***	0.355***	0.478***
	(0.096603)	(0.066709)	(0.091164)	(0.106025)
Observations	14,407	14,922	25,013	20,198
R-squared	0.097	0.165	0.291	0.303

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

Table H2: Bivariate Probit Results Participation and Overeducation Selection Equations

Variables	(1) Female Sample Overeducation	(2) Participation	(4) Male Sample Overeducation	(5) Participation
Kids	-0.026	-0.028**	-0.065***	0.089***
	(0.017358)	(0.011063)	(0.018236)	(0.012181)
Kids_3	-0.025	0.180***	0.013	0.555***
	(0.023532)	(0.016366)	(0.033681)	(0.029155)
Age	-0.022***	0.009***	-0.012***	-0.010***
	(0.000708)	(0.000519)	(0.000790)	(0.000529)
Schooling	-0.199***	0.126***	-0.197***	0.107***
	(0.004424)	(0.001468)	(0.003857)	(0.001536)
Eng_Skill	-0.018	-0.090***	-0.041***	-0.184***
	(0.013948)	(0.009107)	(0.015283)	(0.009981)
City	-0.001	0.054***	-0.043***	0.109***
	(0.016318)	(0.010319)	(0.016615)	(0.011245)
Married	-0.035*	-0.027**	-0.115***	0.306***
	(0.018359)	(0.013266)	(0.025830)	(0.015379)
No_Italian	-0.811***		-0.536***	
	(0.052918)		(0.077668)	
Maximum_D_Mark	-0.107***		-0.220***	
	(0.030259)		(0.040807)	
Retribution_Work	0.100***		0.043***	
	(0.007262)		(0.007514)	
Develop_Work	-0.179***		-0.133***	
	(0.007166)		(0.007351)	
Prosp_Work	-0.083***		-0.069***	
	(0.006196)		(0.006295)	
Age5064		-0.042***		-0.019
		(0.014851)		(0.014139)
Partner_Works		0.490***		0.428***
		(0.011651)		(0.012569)
North		0.443***		0.212***
		(0.008818)		(0.009465)
Centre		0.277***		0.144***
		(0.010881)		(0.011801)
Constant	4.386***	-2.452***	4.003***	-1.264***
	(0.114498)	(0.024806)	(0.109801)	(0.024508)
Observations	113,836	113,836	94,577	94,577
ρ	0.0513	0.0513	-0.144***	-0.144***

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

Table H3: Detailed Coefficient Effect Two-fold[7] and Three-fold Decomposition with Selection, Overeducated Sample

Variables	(1) Two-fold Decomposition	(2) Three-fold Decomposition
Schooling	-0.028	-0.029
	(0.081328)	(0.085597)
Maximum_D_Mark	0.00003	0.00007
	(0.000466)	(0.001147)
Eng_Skill	-0.061*	-0.007*
	(0.003598)	(0.004262)
Exper	0.034	0.027
	(0.039119)	(0.031336)
Exper2	-0.017	-0.012
	(0.023561)	(0.015978)
Tenure	0.010	0.007
	(0.009069)	(0.006473)
Kids	0.009	0.009
	(0.007154)	(0.0079752)
Kids_3	0.002	0.004
	(0.001119)	(0.02301)
Married	0.014	0.015
	(0.009814)	(0.010626)
Intermed_Prof	-0.004	-0.006
	(0.003814)	(0.004941)
Manager	-0.003*	-0.003*
	(0.001628)	(0.001502)
Public_Sector	-0.004*	0.004*
	(0.002503)	(0.002496)
Home_Time	0.004	0.005
	(0.010333)	(0.011294)
North	0.014*	0.016*
	(0.008187)	(0.009117)
Centre	0.006*	0.006*
	(0.003402)	(0.003664)
Homeowner	-0.005	-0.005
	(0.009000)	(0.009013)
Extra_Hours	0.004	0.004
	(0.005178)	(0.004602)
Constant	-0.134	-0.134
	(0.117398)	(0.117398)
Total	-0.104*	-0.106**
	(0.055269)	(0.053455)
Observations	29,329	29,329

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

[7] - The two-fold decomposition uses men as the reference category for the coefficients effect; $\bar{X}_M{}'(\hat{\beta}_M - \hat{\beta}_F)$, while the three-fold decomposition uses women as the reference category; $\bar{X}_F{}'(\hat{\beta}_M - \hat{\beta}_F)$.