DISCUSSION PAPER SERIES

IZA DP No. 7109

# Everything You Always Wanted to Know about Sex Discrimination

Ana Rute Cardoso
Paulo Guimaraes
Pedro Portugal

December 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Everything You Always Wanted to Know about Sex Discrimination

**Ana Rute Cardoso**
*Institute for Economic Analysis (CSIC),*
*Barcelona GSE and IZA*

**Paulo Guimaraes**
*Universidade do Porto,*
*University of South Carolina and IZA*

**Pedro Portugal**
*Bank of Portugal,*
*Universidade Nova de Lisboa and IZA*

Discussion Paper No. 7109
December 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 7109
December 2012

# ABSTRACT

## Everything You Always Wanted to Know about Sex Discrimination*

Earlier literature on the gender pay gap has taught us that occupations matter and so do firms. However, the role of the firm has received little scrutiny; occupations have most often been coded in a rather aggregate way, lumping together different jobs; and the use of samples of workers prevents any reliable determination of either the extent of segregation or the relative importance of access to firms versus occupations. Our contribution is twofold. We provide a clear measure of the impact of the allocation of workers to firms and to job titles shaping the gender pay gap. We also provide a methodological contribution that combines the estimation of sets of high-dimensional fixed effects and Gelbach's (2009) unambiguous decomposition of the conditional gap. We find that one fifth of the gender pay gap results from segregation of workers across firms and one fifth from job segregation. We also show that the widely documented glass ceiling effect operates mainly through worker allocation to firms rather than occupations.

JEL Classification: J31, J16, J24, J71

Keywords: gender wage gap, high-dimensional fixed effects, segregation

Corresponding author:

Ana Rute Cardoso
IAE-CSIC
Campus UAB
08193 Bellaterra, Barcelona
Spain
E-mail: anarute.cardoso@iae.csic.es

* We thank the Portuguese Ministry of Employment, Statistics Department, for access to the data and Lucena Vieira for very competent research assistance. Cardoso acknowledges the support of the Spanish Ministry of the Economy and Competitiveness (grant ECO2012-38460).

## 1. Introduction

Firms, alongside occupations, play a key role shaping wage differences across gender. Indeed, most models of labor market discrimination, whether based on the tastes of employers, co-workers or consumers, on employers' judgment of statistical evidence (Becker, 1957), or on unconscious mental associations (Bertrand et al., 2010), grant the firm the crucial decision making power. Moreover, firms are heterogeneous along several observable and unobservable dimensions other than the potential distaste for a particular group of workers. Therefore, in a market where there are *good* and *bad* firms or simply firms with different wage policies, segregation of workers across firms is likely to lead to a wage gap that will persist over time. There has been little exploration of this line of research, despite early work by Groshen (1991) on the topic. The relevance of worker allocation to firms as a determinant of gender wage differentials was confirmed by Carrington and Troske (1998), Bertrand and Hallock (2001), Bayard et al. (2003), Meng and Meurs (2004), and Woodcock (2008). Nevertheless, these analyses were based on data sets with important limitations: either they covered a restricted set of firms, occupations or cities, or, when covering broader populations, they used an algorithm to match workers and firms that was prone to introducing biases in the analysis; still other studies used samples of workers, thus precluding the proper measurement of gender segregation across units. Alongside this evidence, there is work documenting discriminatory practices in firms' hiring processes that lead to gender segregation across firms (Goldin and Rouse, 2000).

A second important dimension of wage formation is job title heterogeneity, which may influence wage rates for a variety of reasons. First, it is well known that tasks that involve risks of fatal or otherwise serious accidents are better paid than safe tasks. One should therefore expect significant compensating differentials attached to occupations such as bullfighters (included in our sample). Second, jobs that need to be executed under difficult or stressful conditions are also expected to be more highly remunerated than jobs performed in pleasant environments. For example, one should observe higher wages for individuals working on offshore oil platforms or in mines. Third, the complexity of some tasks may require heavy doses of specific training and/or unusually skilled workers. This would be a reason why, for example, brain surgeons and prima ballerinas earn higher wages. Fourth, some occupations are known to be chronically "overcrowded" whereas others are

thought to be in excess demand. For decades, it has been argued that there is an oversupply of teachers and an undersupply of nurses. Finally, the kind of technology in use may also foster unionization of the workplace and favor rent seeking. Production activities that imply the concentration of a large number of workers in a single plant (say, in auto or ship- building industries) facilitate industrial action, and thus improved wage conditions. To properly incorporate these and other such wage determinants one needs a very detailed accounting of the kind of jobs being undertaken by workers.

Empirical analysis of the role of occupational allocation has called attention to the "glass ceiling" effect —a rising gender pay gap as we move up the wage distribution, reflecting the difficulty faced by women in accessing the higher ranks of the occupational ladder, as a result of discriminatory promotion practices (Ransom and Oaxaca, 2005), women's lower reliance on informal networks to seek promotions (Cannings and Montmarquette, 1991), the fact that women receive fewer outside job offers (Blackaby et al., 2005), or males' distaste for having a female boss (Baldwin et al., 2001). Consistent with this evidence, Fortin and Huberman (2002) report on the decline in occupational segregation throughout the 20th century, paralleled by persistent segregation within occupations along hierarchical lines. Bertrand and Hallock (2001) found that differential access to the very top occupations could account for approximately half of the large gender pay gap among top corporate jobs in the US. In general, occupational allocation has been shown to be an important determinant of wage differences across gender (Amuedo-Dorantes and la Rica, 2006; Kunze, 2005; Groshen, 1991). As an exception, Manning and Swaffield (2008) found little role for the occupation explaining the build-up of the gender pay gap in the UK, from entry into the labor market, when hardly any gap exists, until 10 years later.

The line of literature on the role of occupations has also suffered from severe limitations. First of all, it has most often relied on a rather aggregate definition of occupations. This is a major drawback when we seek to ensure we are comparing men and women who perform the same actual tasks inside the firm. In this regard, the criticism by Altonji and Blank is particularly incisive: "Since occupational categories and occupational characteristics are often crudely measured, this raises the issue of whether important unobserved differences in the types of jobs women and men perform remain. This issue is hard to resolve without firm-level data." (Altonji and Blank, 1999, pg. 3221). Another line of analysis uses data representative of the whole economy, but reporting only a sample of workers within each firm. Given that there is no

stratification of the sample by gender, measures of gender segregation across firms are bound to be flawed —let alone segregation across jobs within the firm—as acknowledged by Amuedo-Dorantes and la Rica (2006) or Bayard et al. (2003). Finally, to overcome the serious measurement error that may result from using sample data and the share of women in an unit as the measure of gender segregation, different authors have used as regressors a set of dummy variables for the job inside the firm. One particular problem here is that dummies for job inside the firm absorb firm and occupation and we are left without the possibility of quantifying the separate impact of gender segregation across firms and jobs. More generally, the meta-analysis of the gender wage gap by Weichselbaumer and Winter-Ebmer (2005) has shown that contrasts in results seem to be driven mostly by the scope of the data analyzed – whether a narrow set of firms and occupations or a representative sample of the labor force.

Fortunately, in our dataset we have access to an unusually rich set of information that enables us to identify the collective agreement that regulates the employment contract applicable to each worker and, within each collective agreement, the detailed occupational category of each worker. We can thus identify the detailed job performed by the worker. The reason why this information is collected reflects the specificities of the Portuguese wage setting system (largely conformable to continental European practice). Each year, around 300 different collective agreements are negotiated. The collective agreement defines wage floors for each particular job title (so-called professional category or *categoria profissional*). On average, each collective agreement defines the wage floor for around 100 job titles. Overall, in a given year, one can thus classify each worker along some 30,000 collective agreement/job title combinations. The main reason why this survey was created was, indeed, to allow the officials from the Ministry of Employment to verify whether the employers were complying with the wage floors established by the collective agreement for the job-title of the worker.

In this study, we are confident that by incorporating job title fixed effects in the wage regression, as well as employer fixed effects, we can make good progress providing refined estimates of the gender pay gap filtered from the effects of job title heterogeneity and firm heterogeneity. Indeed, this remarkable longitudinal matched employer-employee dataset covers the whole manufacturing and service non-public sector in Portugal, with information on the firm, each of its workers, and their detailed job. The longitudinal character of our data allows us to use sets of high-dimension fixed effects,

covering the population of units.

Another strength of our work is its methodological contribution. The method used combines the estimation of a model with three high-dimensional fixed effects (Guimaraes and Portugal, 2010) and Gelbach's (2009) decomposition, which yields an unambiguous allocation of the gender pay gap into its components. This methodology can be used to extract information on the sources of the gender pay gap and it can be easily replicated for other environments and topics, in particular using detailed linked employer-employee data that are becoming increasingly available.

Section 2 presents indicators of female labor force integration in Portugal. Section 3 describes the data, while methods are discussed in Section 4. Section 5 provides the key results on the determinants of the gender pay gap and comments on the trends. Section 6 comments on our method as opposed to the traditional methodology used to search for the sources of the gender wage gap. Section 7 concludes.

## 2. Female labor force participation in Portugal

The pattern of female labor force participation over the life-cycle in Portugal is remarkably similar to that of males, unlike what prevails in several other countries, especially Southern European. In particular, around childbirth and during the child rearing years, women in Portugal do not show a decline in their labor force participation (Figure 1). Also, the degree of labor force commitment, evaluated as hours of work or share of workers on part-time, is rather similar across gender (Figure 2), with a very low incidence of part-time for either gender. Moreover, the observable attributes of males and females in the labor market are similar or favor women, namely in the case of education (see Figure 3, which reports a better educational endowment for young and prime-age women than men).

Therefore, three factors documented as candidate explanations for the gender pay gap, namely different education/training, career interruptions, and shorter hours of work (see in particular the detailed study by Bertrand et al. (2010) on MBA graduates) lose strength as sources of the gender pay gap in an economy-wide setting like Portugal.

## 3. Data: The *Quadros de Pessoal* Survey

The dataset employed in this study is quite unusual. The *Quadros de Pessoal* ("Personnel Records") is, by construction, a longitudinal matched

employer-employee-job title dataset. QP is an annual mandatory employment survey collected by the Portuguese Ministry of Employment that covers all establishments with at least one wage earner. Due to the mandatory nature of the survey, problems commonly associated with panel data sets, such as panel attrition, are considerably attenuated.

Data are available on each establishment (location, economic activity, and employment), the firm with which it is affiliated (location, economic activity, employment, sales, and legal framework), and each and every one of its workers (gender, age, education, skill, occupation, tenure, earnings, and the work schedule). The information on earnings is very detailed, precise, and complete. It includes the base wage (gross pay for normal hours of work), regular benefits, irregular benefits, and overtime pay. Information on standard and overtime hours of work is also available. Because the information on earnings is reported by the employer, it is likely to be subject to less measurement error than worker-provided earnings data. The fact that the information contained in the QP survey needs, by law, to be available in a public space at the establishment further reinforces our trust in the information.

A notable feature of the QP is that it collects information regarding the collective agreement that rules the wage dimension of the match between the employer and the employee. Furthermore, within each collective agreement, it identifies the particular job-title that the worker holds.

The dataset is longitudinal in nature. Each firm/establishment entering the database is assigned a unique identifying number and the Ministry implements several checks to ensure that a unit that has previously reported to the database is not assigned a different identification number. Similarly, the dataset provides codes for each collective agreement and, within each collective agreement, for the job-title of the worker. Workers also have a unique identification number that is obtained from their social security number. This allows us to follow them over the years and to match workers with their firms, the collective agreement and the corresponding job-title that they hold in each year.

A number of restrictions were imposed on the raw dataset. First, we limited our analysis to full-time workers in mainland Portugal, between 1996 and 2008. Second, we excluded workers from the Agriculture and Fishery sectors. Third, individuals younger than 18 years old and older than 65 years were, also, excised. Fourth, we dropped from the analysis workers whose monthly wages were below 80 percent of the mandatory minimum wage, which corresponds to the lowest admissible wage for apprentices. Finally, we

dropped around 1 percent of the total number of observations that did not belong to the largest connected group (see below). Our final sample included 28,212,770 observations.

The dependent variable used in our estimating equation is a measure of hourly labor earnings and is constructed as the ratio of the sum of base wages, regular benefits (including seniority payments), and overtime pay over the sum of normal hours of work and overtime hours.

## 4. Firm, job, and worker effects in wage regressions

In this section we discuss methodological aspects related to the estimation of the model and the decomposition approach of Gelbach (2009). The methodology applied in our paper expands that initially developed by Abowd et al. (1999), who presented a statistical framework permitting worker and firm fixed effects to be estimated simultaneously in wage regressions. However, as noted earlier, and as elaborated upon below, we include a third fixed effect for the job title and use a different algorithm to obtain an exact solution for the estimation problem.

The linear wage equation to be estimated has the form:

$$\ln w_{ifjt} = \mathbf{X}_{ift}\boldsymbol{\beta} + \theta_i + \phi_f + \lambda_j + \varepsilon_{ifjt} \quad , \tag{1}$$

which is related, in the statistical literature, with the "three-factor analysis of covariance." In this equation, $\ln w_{ifjt}$ is the natural logarithm of the real hourly wage of individual $i$ ($i = 1, ..., N$) working at firm $f$ ($f = 1, ..., F$) and holding a job title $j$ ($j = 1, ..., J$) at year $t$ ($t = 1, ..., T_i$). There are $T_i$ observations for each individual $i$ and a total of $N^*$ observations. $\mathbf{X}_{ift}$ is a row-vector of $k$ observed (measured) time-varying characteristics of individual $i$ (quadratic terms on age and on seniority within the firm). $\theta_i$ is the person or worker fixed effect (capturing observed and unobserved individual time-invariant heterogeneity), $\phi_f$ is the firm fixed effect (capturing observed and unobserved firm time-invariant heterogeneity), and $\lambda_j$ is the job title fixed effect (capturing observed and unobserved job title time-invariant heterogeneity). Growth in real wages is captured by a set of year dummies. According to this equation, there are five components that explain the wage variability:

1. the observed time-varying characteristics of workers ($X_{ift}\beta$);
2. the workers' heterogeneity or worker fixed effects ($\theta_i$);

3. the firms' heterogeneity or firm fixed effects ($\phi_f$);
4. the job titles' heterogeneity or job title fixed effects ($\lambda_j$); and,
5. an error term component ($\varepsilon_{ifjt}$), assumed to follow the conventional assumptions.

In matrix notation, the stacked system has the form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\phi} + \mathbf{L}\boldsymbol{\lambda} + \boldsymbol{\varepsilon} \quad . \tag{2}$$

In this equation $\mathbf{Y}$ is a $(N^* \times 1)$ vector of real hourly wage (in logs), $\mathbf{X}$ is a $(N^* \times k)$ matrix with $k$ observed time-varying characteristics of individuals, $\mathbf{D}$ is a $(N^* \times N)$ design matrix for the worker effects, $\mathbf{F}$ is a $(N^* \times F)$ design matrix for the firm effects, $\mathbf{L}$ is a $(N^* \times J)$ design matrix for the job title effects, $\boldsymbol{\theta}$ is a $(N \times 1)$ vector of worker effects, $\boldsymbol{\phi}$ is a $(F \times 1)$ vector of firm effects, $\boldsymbol{\lambda}$ is a $(J \times 1)$ vector of job title effects, and $\boldsymbol{\varepsilon}$ is a $(N^* \times 1)$ vector of disturbances (we assume that conditional on $\mathbf{X}$, $\mathbf{D}$, $\mathbf{F}$, and $\mathbf{L}$ mobility is exogenous, in order to make the design matrices orthogonal to the vector of disturbances).

Equations (1) and (2) can be interpreted as the conditional expectation of real hourly wages given the observable characteristics of workers, the date of observation, and the identity of individuals, employing firms, and job titles. The total number of parameters to be estimated is therefore $k + N + J + F$. However, it will not be possible to identify all worker, firm, and job title fixed effects, and a number of $G$ restrictions will have to be imposed allowing us to estimate only $k + N + J + F - G$ parameters.

Abowd et al. (2002) show that in a model with two high-dimensional fixed effects (firm and worker fixed effects) the number of restrictions that needs to be imposed equals the number of "mobility groups," that is, the number of groups of connected firms and individuals. Moreover, estimates of the fixed effects are comparable only within the same "mobility group".

With three fixed effects the situation is more complex. While it is possible to identify the number of restrictions that need to be imposed on the parameters, there is no clear parallel to the "mobility groups" found in the context of the model with two fixed effects. However, using an algorithm proposed by Weeks and Williams (1964), one can identify a subset of the data in which all fixed effects are connected. If we restrict analysis to this subset of the data, we are assured that the estimates of the fixed effects are comparable and, in this case, the number of restrictions, $G$, equals 2.

The full least squares solution to estimate the parameters in (1) solves the following set of normal equations:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'D} & \mathbf{X'F} & \mathbf{X'L} \\ \mathbf{D'X} & \mathbf{D'D} & \mathbf{D'F} & \mathbf{D'L} \\ \mathbf{F'X} & \mathbf{F'D} & \mathbf{F'F} & \mathbf{F'L} \\ \mathbf{L'X} & \mathbf{L'D} & \mathbf{L'F} & \mathbf{L'L} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \\ \boldsymbol{\phi} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{X'Y} \\ \mathbf{D'Y} \\ \mathbf{F'Y} \\ \mathbf{L'Y} \end{bmatrix}$$

However, the high-dimensionality of $\mathbf{D}$, $\mathbf{F}$ and $\mathbf{L}$ prevents the application of the conventional least squares formula. To estimate all parameters (worker, firm and job title fixed effects and the coefficients of all observed time-varying worker and firm characteristics) would require the inversion of a huge matrix. This is impossible to achieve using standard software routines and present-day computers. Accordingly, alternative approaches are required to estimate the full model parameters.

Abowd et al. (1999) proposed an approximate statistical solution that corresponds to using conditional estimation methods that provide estimators that are as similar as possible to full least squares, but computationally tractable. In a subsequent paper, Abowd et al. (2002) developed an algorithm that permits an exact solution of the least squares estimation of equations such as (1), for the case with two high-dimensional fixed effects.[1] In the present treatment, we followed an alternative methodology that provides the exact solution for the linear regression with three high-dimensional fixed effects. This procedure was developed by Guimaraes and Portugal (2010) for the estimation of linear regression models with two or more high-dimensional fixed effects. In brief, this methodology is based on a partitioned algorithm strategy and follows an iterative procedure that leads to the exact solution of the least squares problem. While computationally intensive given its iterative nature, the approach imposes minimum memory requirements. For a detailed description of this methodology and how it can be implemented to estimate equation (1) see Guimaraes and Portugal (2010).

We then apply Gelbach's exact decomposition, which is based on the formula for omitted variable bias. This decomposition allows for an unequivocal quantification of the portion of the gender pay gap due to each of the variables of interest. The idea is quite simple. Consider the basic regression of

[1] The user-written command *a2reg* programmed by Amine Ouazad is the Stata implementation of this algorithm.

(log) wages on a set of explanatory variables that includes the variable of interest (gender dummy), a quadratic term on experience, a quadratic on tenure, and year dummies:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{\varepsilon} \quad . \tag{3}$$

Following Gelbach (2009) we know from the omitted variables bias formula that the difference between the coefficients of the basic specification shown above and those of the full specification given in (1) can be expressed as

$$\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{\beta}} = \mathbf{P_X}\mathbf{D}\widehat{\boldsymbol{\theta}} + \mathbf{P_X}\mathbf{F}\widehat{\boldsymbol{\phi}} + \mathbf{P_X}\mathbf{L}\widehat{\boldsymbol{\lambda}} \quad , \tag{4}$$

where $\mathbf{P_X} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Note that $\mathbf{D}\widehat{\boldsymbol{\theta}}$, $\mathbf{F}\widehat{\boldsymbol{\phi}}$ and $\mathbf{L}\widehat{\boldsymbol{\lambda}}$ are column vectors containing the estimates of the fixed effects for the worker, firm and job titles, respectively. This means that the terms on the right-hand side of (4) are the coefficients of regressions of the estimates of the fixed effects on the explanatory variables included in the basic regression in (3). Thus, we can rewrite the above equation more succinctly as

$$\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\delta}}_{\boldsymbol{\theta}} + \widehat{\boldsymbol{\delta}}_{\boldsymbol{\phi}} + \widehat{\boldsymbol{\delta}}_{\boldsymbol{\lambda}} \quad . \tag{5}$$

Centering our interest on the coefficient for gender, we can now identify the conditional contribution of each fixed effect to the gender gap. Since the coefficient for gender in the full specification is zero, we have

$$\widehat{b}_G = \widehat{\delta}_{\boldsymbol{\theta}G} + \widehat{\delta}_{\boldsymbol{\phi}G} + \widehat{\delta}_{\boldsymbol{\lambda}G} \quad , \tag{6}$$

yielding an unambiguous decomposition of the wage gap that gives us the conditional contribution of workers, firms, and job titles to the gender pay gap. The interpretation of these coefficients is clear. For example, the coefficient $\widehat{\delta}_{\boldsymbol{\phi}G}$ can be interpreted as the log point reduction in the gender pay gap that would occur if males and females were equally distributed across firms, conditional on all other variables included in the full model. The two other coefficients may be interpreted in a similar fashion.

## 5. The sources of the gender pay gap

The raw gender pay gap in Portugal throughout the period under analysis was 24 log points. Conditional on the workers' age, to proxy for labor market experience, and tenure, the gap reduces to 23 log points, reflecting women's

very slight disadvantage in those observable attributes (column 1 in Table 2 in appendix reports the regression results on this basic specification given by equation 3). This average gap conceals an impressive contrast across the male and female wage distributions (see Figure 4). The female wage distribution is considerably more concentrated around a lower wage level —see the much higher peak and its shift to the left.

A relevant question that follows is therefore: How are males and females allocated in the labor market, so as to generate this outcome? How far can the access to different firms and to very detailed jobs go in explaining the contrast across wage distributions? Figure 5 reports the estimates from our model of the high-dimensional fixed effects for the firm, the job, and the worker obtained from estimation of equation (1). A novel view on the gender pay gap is obtained, showing exactly where in the distribution of firm and job quality the pay gap is determined. Females are not only employed in firms and jobs of lower average quality than males, but they are considerably more concentrated in such units, as indicated by the more peaked distributions and their shift to the left with respect to males' distribution. The concentration of female employment in lower paying units is particularly striking for jobs.

The following questions arise as a result: How much exactly does the allocation into firms and jobs matter for the gender pay gap? What is left of the gap once we account for those allocation mechanisms? Following our estimation of the sets of high-dimension fixed effects reported in Figure 5, we use Gelbach's (2009) methodology to provide an unambiguous answer to these questions. Table 1 reports the results of the decomposition. In the first row of Table 1 we follow equation (6) and implement the decomposition using as a reference the full model that includes firm, job, and worker fixed effects. We can see that the allocation of males and females to firms of different quality accounts for 4.3 log points of the gender pay gap (or 19% of the conditional gap). Interestingly, the allocation to jobs accounts for almost the same share of the total conditional gender wage gap as that due to firms. The remaining 63% of the gender gap persists within jobs and firms for workers of the same age and firm seniority (an issue to which we will return in the next section). When lumping together the job and the firm, we lose track of their separate contributions to the gender pay gap, and find that they jointly account for 45% of the conditional gap (second row of Table 1). This means that the allocation of job-titles within firms is responsible for a non-negligible part (1.7 log points) of the conditional gender gap.

The empirical distribution of the firm fixed-effects evinces the fundamen-

tal importance of heterogeneity of wage policies across firms. In Portugal, permanent firm heterogeneity accounts for more than one third of the total wage variation (Torres et al., 2012). This means that prospective employees, facing a highly dispersed distribution of wage offers, may benefit greatly if they intensively invest in job search to locate the firms with more generous wage policies. It is clear from Figure 5 and Table 1 that male workers are more successful than female workers at being hired and retaining jobs in high-wage firms.

Why are women penalized by the allocation across firms? One possibility would be that female workers are less efficient job searchers. A number of mechanisms may lie behind this outcome. First, women may search less intensively than men because they may have more productive alternatives for the allocation of their time (e.g., domestic production). Second, women may limit the set of job opportunities to jobs with particular characteristics (e.g., flexibility of work schedules). Third, women may shape their search strategy to the labor supply decisions of their husbands, as implied by first generation labor supply models. For example, those women may severely restrict the geographical boundaries of their job search. Fourth, women may have significant disadvantages, compared with men, exploiting their social networks to locate suitable job offers. Furthermore, there is some evidence that female dominated social networks tend to favor the placement of women in female dominated jobs and occupations (Mencken and Winfield, 2000). Fifth, women may underestimate the relevant distribution of wage offers, either because they expect to be discriminated against by some firms or because they misrepresent their true value to the firm and bargain wages less aggressively than men. A telling indication that women have lower expectations regarding wages is given by evidence that they report lower reservation wages than men for similar observed characteristics (Addison et al., 2004).

From the employer perspective it appears that high-wage firms hire (or retain) female workers less frequently than male workers. Conventional explanations based on the preferences of employers, customers, or co-workers may be raised. Alternatively, expectations of career interruptions by female workers may partially justify this asymmetry in hiring decisions. Nevertheless, it is not clear why such explanations should apply more intensively to firms that practice more generous wage policies. Indeed, not only are women sorted more frequently into low-wage firms, but also the wage penalty increases with the size of the firm fixed effect: the gender gap increases from -0.3 percent at the 10th percentile, to -4.9 percent at the median, to -11.4

percent at the 90th percentile (Table 3 in appendix). From this evidence, it is quite clear that the glass ceiling applies much more to firms than to occupations or job-titles (but see below).

Job-title heterogeneity also plays a significant role in wage variation. The notion of job-title summarizes the skill requirements and the hierarchical position held by the worker. Job-title segregation is simply a more accurate, refined, and disaggregated version of the so-called occupational segregation. Because the definition of the job-title is an outcome of the negotiation between trade unions and employer associations, it also reflects the bargaining power of the workers. The evidence provided in Figure 5 and Table 1 shows that female workers suffer a significant wage penalty associated with the process of sorting into different job-titles. Why are women allocated to lower-paying job-titles? What is behind job-title segregation? The presence of barriers to entry into high-paying job-titles, driven either by the hiring decisions of the employers or by the requirements of the job, is certainly one of the mechanisms at work. Job promotion decisions biased against female workers may also be at play. Overcrowding of job-titles highly preferred by female workers may drive their corresponding wages down. On the other hand, the access to some high-wage job-titles that are controlled by closed shop trade unions are frequently male-dominated (e.g. longshoremen). The allocation into job-titles is responsible for around one fifth of the total gender gap.

There is, however, no indication that the wage gap between men and women increases significantly along the job-title fixed effect dimension. When we compare the distribution of job-title fixed-effects among males and females, we obtain a -1.9 percent difference for the 10th percentile, -3.9 percent at the median, and just -2.4 percent for the 90th percentile. As hinted at above, there is no indication of a glass ceiling along the job-title dimension, once we take into account the heterogeneity of the firm's wage policies. This finding is consistent with recent work by Manning and Swaffield (2008), who concluded that occupational allocation in the UK has little role in explaining the widening of the gender pay gap during the first ten years after entry into the labor market.

The period under analysis comprises two decades and we are interested in checking for any possible changes in the sources of the gender pay gap. We have therefore performed a decomposition of the gap on a year by year basis. In Figure 6, we see that the overall gender pay gap —sum of the three components —widened during the first decade and narrowed during

the subsequent decade. Over the whole period, the shrinking of the gap was mostly due to the convergence in the worker fixed effect component, from a gender gap of -13.8 log points in 1986 to -11.8 log points by 2008. In the meantime, the importance of segregation across firms also declined, though only slightly, from a gap of -4.7 to -3.3 log points. The reverse was observed for segregation across jobs, reflected in a contribution to the overall conditional gap that increased from -3.2 to -4.6 log points. We therefore find a decline over time in the "discrimination" component of the gender gap and a slightly more equal access to firm types; nevertheless, access to job types would, by itself, have contributed to widening the gender pay gap.

## 6. Our proposal and previous methods

The most commonly used procedure to search for the sources of the gender pay gap is its decomposition into differences in labor market attributes (experience, seniority within the firm, schooling, etc.) and differences in the returns to those attributes. The latter would be the unexplained part of the gap or "discrimination" component. This approach has proved insightful at times and in countries where there was indeed a gap in the attributes that males and females brought to the labor market. However, even then this approach provided fewer clues regarding the mechanisms that could lead to different returns on male and female attributes. Today, in countries where there is a notable convergence in labor market attributes across gender, the method is no longer as insightful —if males and females had exactly the same attributes, all of the burden of the wage gap would fall on the "unexplained" component or differences in returns.

Symptomatic of this evolution in the labor market is the contrast in the focus of the chapters on gender in the Handbooks of Labor Economics a decade apart. The chapter by Altonji and Blank (1999) reported on the crucial importance attached in empirical analyses to gender differences in observable attributes, especially human capital endowments, as determinants of the gender wage gap; instead, Bertrand (2011) surveys evidence on psychological differences and social factors, such as risk preferences, competitiveness and bargaining attitudes, that may lead women and men to make contrasting occupational choices and to progress at a different pace on the job ladder. In line with this convergence of male and female observable attributes, the meta-analyses of research on the gender wage gap in the US by Jarrell and Stanley (2004) and Stanley and Jarrell (1998) provide a particularly inter-

esting core result —whereas the original study showed that, depending on whether the research design corrected for selection into employment, results would change (with lack of correction for selection leading to an overestimation of the female wage penalty), by the mid-2000s the meta-analysis by the same authors using the same methodology showed instead that correction for selection no longer influenced the magnitude of the estimated gender pay gap. The work by Olivetti and Petrongolo (2008) shows that in countries with higher female employment participation, the correction for selection into the labor market tends to have a smaller impact on the gender pay gap than in countries where female participation is low.[2]

The convergence in male and female observable attributes is particularly obvious in Portugal, where even their selection into the labor market at every stage of the life-cycle seems to obey the same rules. This presses for alternative tools of analysis that can provide guidance on the sources of the gender gap. We do so, by scrutinizing the role that the allocation of workers to firms —the ultimate decision makers —and to jobs —a very fine partition of occupations for purposes of wage bargaining —may have. We therefore highlight the decisions on labor market allocation that influence the gender pay gap.

In an economy where males and females show very similar labor force participation throughout their life-cycle and worker attributes are very similar, such that concerns over selection are minimized, we would not expect effort, motivation, or worker quality in general to differ across gender within very homogeneous units defined by detailed job inside the firm. However, the third panel of Figure 5 and Table 1 report a different situation. Within the same firm and very narrowly defined job cells, the wage distributions for males and females with the same age and firm seniority show remarkable contrast. We may therefore be getting closer to capturing true discrimination.

## 7. Conclusion

By combining the estimation of wage regressions with sets of high dimensional fixed effects and Gelbach's (2009) decomposition method, we provide

[2]In their analysis based on survey data, Portugal shows an intermediate gender employment gap, below the Anglo-Saxon and Scandinavian countries, close to Central Europe, and considerably above the Southern European countries of Greece, Spain, and Italy (Olivetti and Petrongolo, 2008, pg. 623, Figure 1).

an unambiguous answer to the questions: How much does the allocation of workers into firms and jobs matter for the gender pay gap? What is left of the gap once we account for those allocation mechanisms? Roughly one fifth of the conditional gender wage gap that subsists for workers with the same general labor market experience and the same seniority within the firm is due to their allocation to firms of different quality; a like share is due to their allocation to jobs of different quality. To the extent that males and females with the same age and tenure, doing the same narrowly defined job for the same firm would be expected to present similar ability, we would allocate the remaining three fifths of the gender wage gap to "discrimination".

We also show that the widely documented glass ceiling effect, according to which the gender pay gap widens at the top of the wage distribution, operates mostly through worker allocation to firms, with lower access of females to higher paying firms, rather than through worker allocation to jobs.

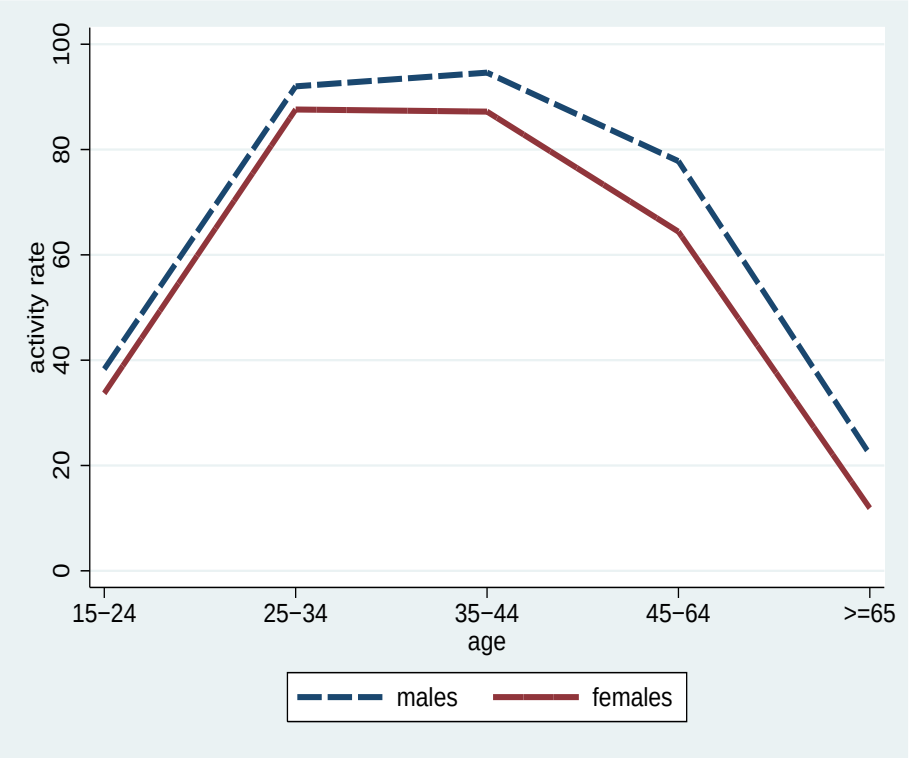
## Tables and Graphs

Table 1: Conditional Decomposition of the Gender Wage Gap

| Worker FE | Firm FE | Job FE | Job & Firm FE |
|---|---|---|---|
| 0.1444 | 0.0431 | 0.0425 | |
| (0.0001) | (0.0001) | (0.00004) | |
| 0.1270 | | | 0.1029 |
| (0.0001) | | | (0.0001) |

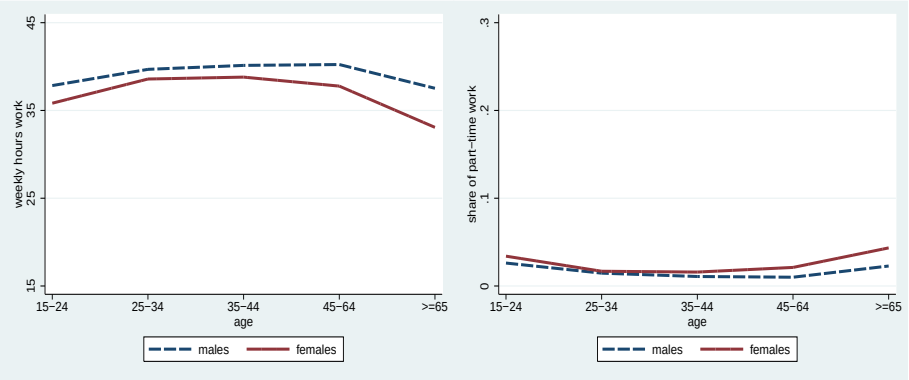
Note: Decompositions based on Gelbach (2009). The basic model includes as regressors a quadratic term on age, a quadratic term on tenure, and year dummies, apart from the gender dummy variable. Each full model further includes the sets of fixed effects specified. The number of fixed effects are: 576,459 for the firm; 107,785 for the job; 4,138,799 for the worker; and 4,593,383 for firm-job matches.

Figure 1: Activity Rate, 2010



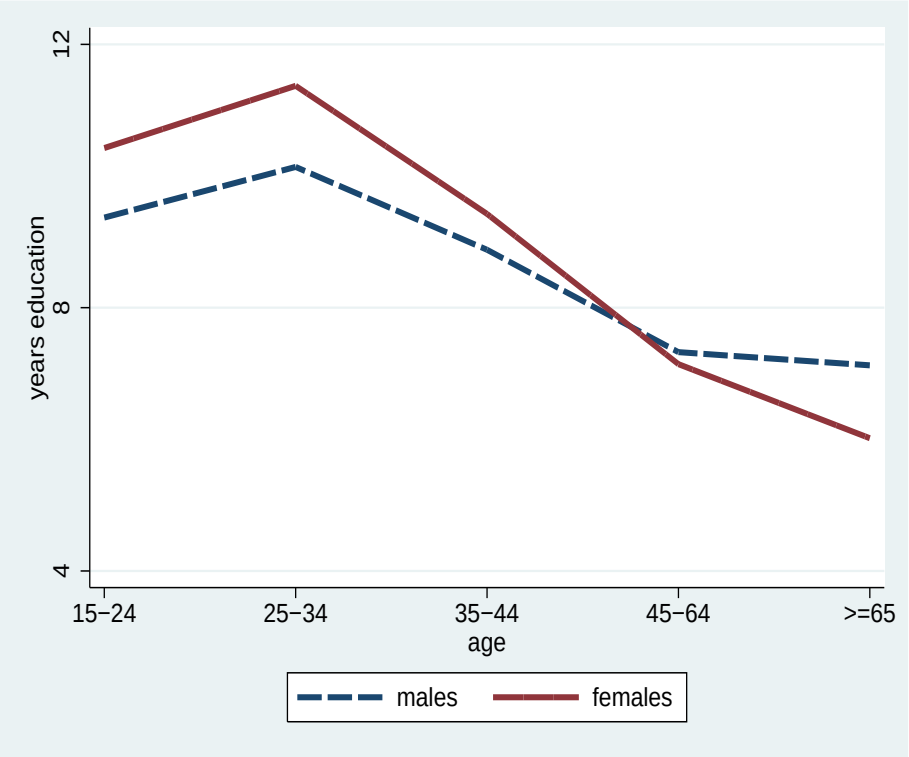

Source: INE, Inquérito ao Emprego, http://www.ine.pt, April 3, 2012.

Figure 2: Hours of Work, 2009



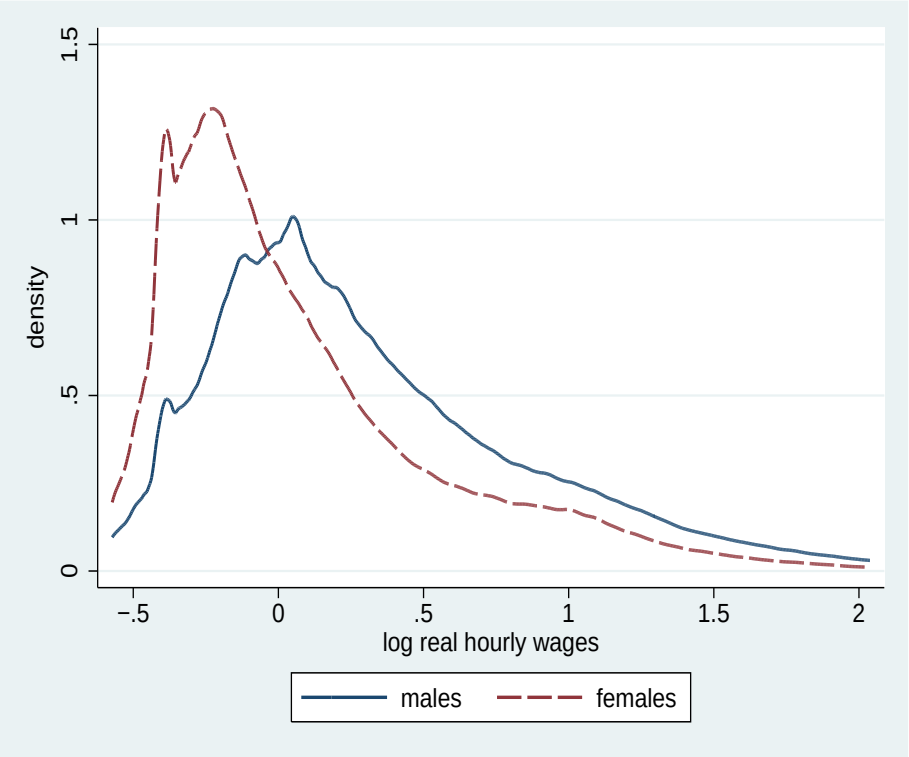

Source: MTSS, *Quadros de Pessoal* 2009.

Figure 3: Years of Schooling, 2009



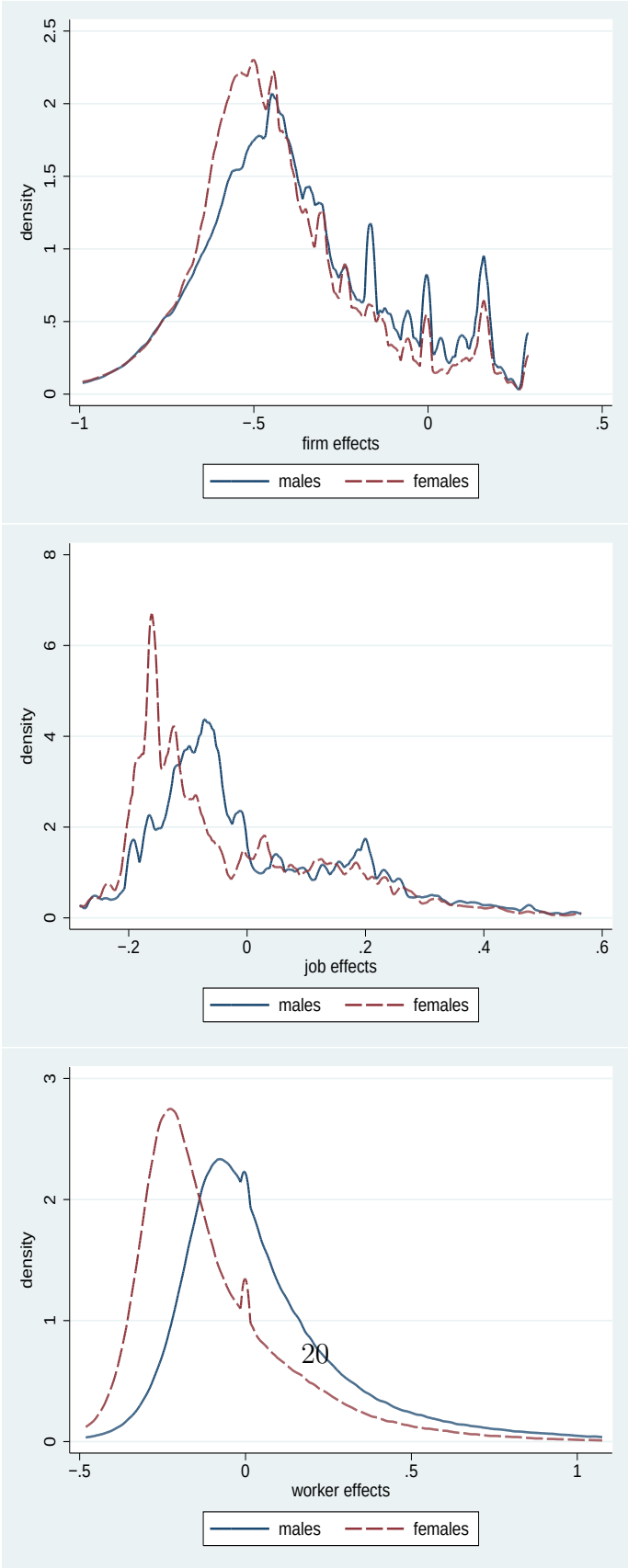

Source: MTSS, *Quadros de Pessoal* 2009.

Figure 4: Kernel Density of Raw Wages, by Gender




Note: The plot excludes wages below percentile 1 of the female distribution and above percentile 99 of the male distribution. Both kernels rely on a common bandwidth.

Figure 5: Kernel Densities of Fixed Effects (Job, Firm, Worker), by Gender




Note: All plots exclude wages below percentile 1 of the female distribution and above percentile 99 of the male distribution. Each graph uses a common bandwidth.

Figure 6: Conditional Decomposition of the Gender Wage Gap, Separately by Year

Note: See note to Table 1.

## Appendix: Additional Tables

Table 2: Regressions (Log) Real Hourly Wages

| | (1) | (2) |
|---|---|---|
| female | -.230 | |
| | (.005) | |
| age | .053 | .020 |
| | (.001) | (.00006) |
| age sq. | -.0006 | -.0002 |
| | (.00002) | (7.88e-07) |
| tenure | .002 | .0006 |
| | (.0001) | (4.41e-06) |
| tenure sq. | -1.32e-06 | -1.23e-06 |
| | (1.61e-07) | (1.11e-08) |
| year fixed effects | yes | yes |
| firm, job, and worker fixed effects | | yes |
| Obs. | 28,212,770 | 28,212,770 |
| $R^2$ | .189 | .915 |

Note: Standard-errors clustered at the firm level.

Table 3: Descriptive Statistics on the Fixed Effects

| | Percentile | | |
|---|---|---|---|
| | 10 | median | 90 |
| Firm fixed effect | | | |
| males | -0.698 | -0.421 | 0.031 |
| females | -0.700 | -0.470 | -0.083 |
| gap x 100 | -0.3 | -4.9 | -11.4 |
| Job fixed effect | | | |
| males | -0.169 | -0.047 | 0.257 |
| females | -0.188 | -0.086 | 0.233 |
| gap x 100 | -1.9 | -3.9 | -2.4 |

Note: Firm and job fixed effects estimated according to equation (1).

## References


Abowd, J. M., Creecy, R. H., Kramarz, F., 2002. Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.
URL `http://ideas.repec.org/p/cen/tpaper/2002-06.html`

Abowd, J. M., Kramarz, F., Margolis, D. N., 1999. High wage workers and high wage firms. Econometrica 67 (2), 251–334.

Addison, J. T., Centeno, M., Portugal, P., 2004. Reservation wages, search duration, and accepted wages in Europe, IZA Discussion Paper 1252.

Altonji, J. G., Blank, R. M., 1999. Race and gender in the labor market. In: Ashenfeher, O., Card, D. (Eds.), Handbook of Labor Economics. Vol. 3. Elsevier, pp. 3143–3259.

Amuedo-Dorantes, C., la Rica, S. D., 2006. The role of segregation and pay structure on the gender wage gap: Evidence from matched employer-employee data for Spain. The B.E. Journal of Economic Analysis & Policy 0 (1), 10.

Baldwin, M. L., Butler, R. J., Johnson, W. G., 2001. A hierarchical theory of occupational segregation and wage discrimination. Economic Inquiry 39 (1), 94–110.

Bayard, K., Hellerstein, J., Neumark, D., Troske, K., 2003. New evidence on sex segregation and sex differences in wages from matched employee-employer data. Journal of Labor Economics 21 (4), 887–922.

Becker, G. S., 1957. The Economics of Discrimination. University of Chicago Press.

Bertrand, M., 2011. New perspectives on gender. In: Ashenfeher, O., Card, D. (Eds.), Handbook of Labor Economics. Vol. 4B. Elsevier, pp. 1543–1590.

Bertrand, M., Goldin, C., Katz, L. F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. American Economic Journal: Applied Economics 2 (3), 228–55.

Bertrand, M., Hallock, K. F., 2001. The gender gap in top corporate jobs. Industrial and Labor Relations Review 55 (1), 3–21.

Blackaby, D., Booth, A. L., Frank, J., 2005. Outside offers and the gender pay gap: Evidence from the UK academic labour market. Economic Journal 115, F81–F107.

Cannings, K., Montmarquette, C., 1991. Managerial momentum: A simultaneous model of the career progress of male and female managers. Industrial and Labor Relations Review 44 (2), 212–228.

Carrington, W. J., Troske, K., 1998. Sex segregation in U.S. manufacturing. Industrial and Labor Relations Review 51 (3), 445–464.

Fortin, N. M., Huberman, M., 2002. Occupational gender segregation and women's wages in Canada: An historical perspective. Canadian Public Policy 28 (s1), 11–39.

Gelbach, J. B., 2009. When do covariates matter? and which ones, and how much?, Department of Economics Working Paper 2009-07, University of Arizona, Tucson.

Goldin, C., Rouse, C., 2000. Orchestrating impartiality: The impact of "blind" auditions on female musicians. American Economic Review 90 (4), 715–741.

Groshen, E. L., 1991. The structure of the female/male wage differential: Is it who you are, what you do, or where you work? Journal of Human Resources 26 (3), 457–472.

Guimaraes, P., Portugal, P., 2010. A simple feasible procedure to fit models with high-dimensional fixed effects. Stata Journal 10 (4), 628–649.

Jarrell, S. B., Stanley, T. D., 2004. Declining bias and gender wage discrimination? a meta-regression analysis. Journal of Human Resources 39 (3), 828–838.

Kunze, A., 2005. The evolution of the gender wage gap. Labour Economics 12 (4), 73–97.

Manning, A., Swaffield, J., 2008. The gender gap in early-career wage growth. Economic Journal 118, 983–1024.

Mencken, F. C., Winfield, I., 2000. Job search and sex segregation: Does sex of social contact matter? Sex Roles 42, 847–864.

Meng, X., Meurs, D., 2004. The gender earnings gap: effects of institutions and firms–a comparative study of French and Australian private firms. Oxford Economic Papers 56 (2), 189–208.

Olivetti, C., Petrongolo, B., 2008. Unequal pay or unequal employment? a cross-country analysis of gender gaps. Journal of Labor Economics 26 (4), 621–654.

Ransom, M., Oaxaca, R. L., 2005. Intrafirm mobility and sex differences in pay. Industrial and Labor Relations Review 58 (2).

Stanley, T. D., Jarrell, S. B., 1998. Gender wage discrimination bias? a meta-regression analysis. Journal of Human Resources 33 (4), 947–973.

Torres, S., Portugal, P., Addison, J. T., Guimaraes, P., 2012. The sources of wage variation: A three-way high-dimensional fixed effects model, miimeo.

Weeks, D. L., Williams, D. R., 1964. A note on the determination of connectedness in an n-way cross classification. Technometrics 6 (3), 319–324.

Weichselbaumer, D., Winter-Ebmer, R., 2005. A meta-analysis of the international gender wage gap. Journal of Economic Surveys 19 (3), 479–511.

Woodcock, S., 2008. Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. Labour Economics 15 (4), 403–418.