

Post-print of: Quarterly Journal of Economics, 131(2), 2016, 633-686.

What drives the gender wage gap? A look at the role of firm and job-title heterogeneity

By Ana Rute Cardoso^a, Paulo Guimarães^b, and Pedro Portugal^c

a Institute for Economic Analysis (CSIC) and Barcelona GSE, Spain.

b Banco de Portugal, Praça da Liberdade 92, 4000-322, Porto, and CEFUP, Universidade do Porto, Porto, Portugal; pfguimaraes@bportugal.pt.

c Banco de Portugal, Lisbon, Portugal.

Abstract

We investigate the mechanisms that shape the gender wage gap in Portugal and provide a clear measure of the impact of the allocation of workers to firms and to jobs. We find that one fifth of the gender gap results from segregation of workers across firms and another fifth from job segregation. We also conclude that the “glass ceiling effect” operates mainly through worker allocation to firms rather than occupations. Our results are based on the application of Gelbach’s decomposition method to the conditional wage gap obtained from a wage equation with high-dimensional fixed effects. This approach may prove to be equally useful in applications to other problems in economics.

JEL Classifications: J31, J16, J24, J71.

1 Introduction

It is well established that firms play a key role shaping wage differences across gender. Indeed, most models of labour 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 [see the earlier work of Blau (1977) and Groshen (1991) on this topic]. Carrington and Troske (1998), Bertrand and Hallock (2001), Bayard et al. (2003), Meng and Meurs (2004), and Woodcock (2008) provided evidence on the relevance of worker allocation to firms as a determinant of gender wage differentials. However, all these studies were based on data sets with important limitations. They either covered a restricted set of firms, occupations, or cities or, when covering broader populations, lacked a reliable method to match workers and firms. 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. Thus, for example, one should 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. Finally, the kind of technology in use may also foster unionisation of the workplace and favour rent seeking. Production activities that imply the concentration of a large number of workers in a single plant (say, in auto or shipbuilding 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. This effect reflects the difficulty faced by women in accessing the higher ranks of the occupational ladder. It can be the 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 (Groshen, 1991; Kunze, 2005; Amuedo-Dorantes and la Rica, 2006). 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 labour market, when hardly any gap exists, until 10 years later.

The strand 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 that 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, p. 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 a 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 analysed – whether a narrow set of firms and occupations or a representative sample of the labour force.

Fortunately, we have access to an unusually rich data set 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. 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*). 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.

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 data set 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 control for these dimensions using sets of high-dimensional fixed effects.

Another strength of our work is its methodological contribution. Our approach combines the estimation of a model with three high-dimensional fixed effects (Guimarães and Portugal, 2010) with Gelbach’s (2016) decomposition, yielding an unambiguous allocation of the gender pay gap into its worker, firm, and job-title components. This approach may prove equally useful in other settings involving estimation of models with high-dimensional fixed effects.

Section 2 presents indicators of female labour 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 labour force participation in Portugal

The pattern of female labour 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 labour force participation (Figure 1). Also, the degree of labour force commitment, evaluated as hours of work is rather similar across gender (Figure 2), with a low incidence of part-time for either gender. Moreover, the observable attributes of males and females in the labour market are similar or favour 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. Also, the fact that female participation patterns are so similar to males' over the lifecycle downplays the potential impact of selection into the labour force on the gender pay gap.¹

3 Data: The *Quadros de Pessoal* Survey

The data set employed in this study is quite unusual. The *Quadros de Pessoal* (QP) is, by construction, a longitudinal matched employer-employee-job title data set. 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, and earnings). The information on earnings is very detailed, precise, and complete. It includes the base wage (gross pay for normal hours of work), regular 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

¹Subject to the short-term maternity leave established by the law, currently set at four months.

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 relevance of progressing from the broad classification of occupations traditionally available in datasets into a richer description of the actual tasks performed by workers has been highlighted in the literature [see for example Autor (2013), or Autor et al. (2006), Goos and Manning (2007), and Dustmann et al. (2009) on job polarisation]. This recent literature illustrates that, in addition to firm and worker heterogeneity, wage outcomes are shaped by task heterogeneity, which should be explicitly accounted for in the analysis (Torres et al., 2013). The variable job category available in our dataset is well suited for an insightful analysis of wage setting mechanisms — this classification of jobs goes beyond a very fine partition of occupations, to consider the complexity of the task performed, the skill level required for the job, the necessary labour market experience, and the hierarchal standing of the worker. Moreover, it is the unit of wage setting in collective bargaining defined within each collective agreement. These job titles are rather disaggregated. For example, in the ship building industry, the painter of the port side of the boat has a distinct job title from a painter of the starboard (because those positions face different risks of accident). Furthermore, similar occupations will be considered distinct job titles if their wages are settled by different collective agreements. For example, a secretary in the banking industry has a job title that is distinct from a secretary in the textile industry. This level of disaggregation implies that, in a typical year, the 300 collective agreements that are settled, determine about 30,000 wage floors corresponding to 30,000 job titles [see Carneiro et al. (2014) and Martins (2014)]. Furthermore, the coverage of the collective agreement reflects as well the degree of relative bargaining power of workers and employers.² Our analysis of the determinants of the gender pay gap will explicitly account for firm, worker and job title heterogeneity.

The data set 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 data set 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 data set. First, we limited our analysis to full-time workers in mainland Portugal, between 1986 and 2008.

²For a detailed description of the Portuguese wage setting system see Cardoso and Portugal (2005) and Addison et al. (2015)

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% of the mandatory minimum wage, which corresponds to the lowest admissible wage for apprentices. Finally, we dropped (around 1% 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 real hourly labour earnings and is constructed as the ratio of the sum of deflated base wages, regular benefits (including seniority payments), and overtime pay over the sum of normal hours of work and overtime hours.

Table A1 in appendix reports descriptive statistics. On one hand, women are slightly younger and have slightly lower tenure than men. On the other hand, they tend to be more educated. The “raw” gender pay gap during the period under analysis was 24 log points.

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 (2016). 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:

$$\ln w_{ifjt} = \mathbf{x}_{ifjt}\boldsymbol{\beta} + \theta_i + \phi_f + \lambda_j + \varepsilon_{ifjt} \quad , \quad (1)$$

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 ($i = 1, \dots, N$) working at firm f ($f = 1, \dots, F$) and holding a job title j ($j = 1, \dots, J$) at year t ($t = 1, \dots, T_i$). There are T_i observations for each individual i and a total of N^* observations. \mathbf{x}_{ifjt} is a row-vector of observed (measured) time-varying characteristics of individual i (a quadratic term on age and a quadratic term on tenure) as well as time dummies that capture growth in real wages. θ_i is the person or worker fixed effect (capturing observed and unobserved individual time-invariant heterogeneity), ϕ_f is the firm fixed effect (capturing observed and unobserved firm time-invariant heterogeneity), and λ_j is the job title fixed effect (capturing observed and unobserved job title time-invariant heterogeneity). According to this equation, there are five components that explain the wage variability:

1. the time-varying characteristics ($\mathbf{x}_{ifjt}\boldsymbol{\beta}$);

2. the workers' permanent heterogeneity or worker fixed effects (θ_i);
3. the firms' permanent heterogeneity or firm fixed effects (ϕ_f);
4. the job titles' permanent heterogeneity or job title fixed effects (λ_j); and,
5. an error term component (ε_{ifjt}), assumed to follow the conventional assumptions.

In matrix notation, the stacked system has the form:

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

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 containing the observed time-varying characteristics of individuals and time dummies, \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 (log) 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 need 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 may be 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. This means that 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 (this group accounts for 99% of our original data set).

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

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

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 time dummies) 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. 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 Guimarães 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 Guimarães 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. More specifically, suppose that we run a simpler model that includes the \mathbf{X} covariates plus \mathbf{G} , a gender dummy variable. Our aim is to understand what happens to the coefficient on the gender dummy when we go from this basic specification

$$\mathbf{Y} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \gamma\mathbf{G} + \varepsilon \quad (3)$$

to the full specification in (2) where all three fixed effects are added. Even though the gender variable is absorbed by the worker fixed effect we are still able to identify the contribution of each fixed effect to the gender gap. To gain some intuition into the workings of Gelbach’s decomposition in this particular setting assume for the moment that the only covariate included in the base specification is \mathbf{G} .³ Now, the estimated coefficient of the regression of \mathbf{Y} on \mathbf{G} , say $\hat{\gamma}_G$, is the difference between the average wage (in logs) of females and males (the “raw” gender gap). In matrix terms, the coefficient $\hat{\gamma}_G$ can be expressed as

$$\hat{\gamma}_G = (\mathbf{G}^*{}'\mathbf{G}^*)^{-1}\mathbf{G}^*{}'\mathbf{Y} \quad (4)$$

where $\mathbf{G}^* \equiv \mathbf{M}_1\mathbf{G}$ and $\mathbf{M}_1 \equiv [\mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']$ is a matrix that demeans a pre-multiplied vector and $\mathbf{1}$ is a conformable vector of ones. When applied to any

³We thank Jonah Gelbach for suggesting this way of presenting the decomposition.

vector the matrix $(\mathbf{G}^* \mathbf{G}^*)^{-1} \mathbf{G}^{* \prime}$ calculates the difference between the averages of males and females. Consider now the fitted regression model that includes only the three fixed effects,

$$\mathbf{Y} = \mathbf{D}\hat{\boldsymbol{\theta}} + \mathbf{F}\hat{\boldsymbol{\phi}} + \mathbf{L}\hat{\boldsymbol{\lambda}} + \hat{\boldsymbol{\varepsilon}} \quad (5)$$

and note that in the above expression $\mathbf{D}\hat{\boldsymbol{\theta}}$, $\mathbf{F}\hat{\boldsymbol{\phi}}$, and $\mathbf{L}\hat{\boldsymbol{\lambda}}$ are column vectors containing the least-square estimates of the fixed effects for the worker, firm, and job titles, respectively. Pre-multiplying all terms of the above equation by $(\mathbf{G}^* \mathbf{G}^*)^{-1} \mathbf{G}^{* \prime}$ we obtain

$$\hat{\gamma}_G = \hat{\delta}_{\theta G} + \hat{\delta}_{\phi G} + \hat{\delta}_{\lambda G} \quad , \quad (6)$$

where we note that $(\mathbf{G}^* \mathbf{G}^*)^{-1} \mathbf{G}^{* \prime}$ is orthogonal to $\hat{\boldsymbol{\varepsilon}}$. The above expression shows that the “raw” gender gap can be partitioned into three components each associated with a fixed effect. These components are simply the difference between the average estimated fixed effects for males and females. If, for example, the average of the fixed effects associated with job-titles is identical for males and females then the allocation into jobs has no contribution to the gender gap. While intuitive, this decomposition ignores time-varying effects that may explain part of the gender gap. Thus, as stated earlier, our interest centers on understanding what happens when we move from the standard equation (3) to the full specification in (2) that includes all three fixed effects. In this latter case the procedure is the same the only difference being the need to partial out the effect of the time-varying covariates. In practical terms this means that we replace \mathbf{M}_1 by $\mathbf{M}_2 \equiv [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ on the derivation above. The decomposition shown in (6) remains valid but now it yields an unambiguous decomposition of the covariate adjusted (i.e. conditional) wage gap. The interpretation of the coefficients in (6) remains clear. For example, the coefficient $\hat{\delta}_{\phi G}$ represents the log point reduction in the gender pay gap that would occur if males and females were equally paid 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. The adjusted gender gap, conditional on the workers’ age, to proxy for labour market experience, and tenure, reduces to 23 log points, reflecting women’s very slight disadvantage in those observable attributes (column 1 in Table A2 in the 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 labour 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 (2016) 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.⁴ 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).

The empirical distribution of the firm fixed-effects evinces the fundamental 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., 2013). 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 penalised 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. 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 labour supply decisions of their husbands, as implied by first generation labour 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,

⁴It is well known that in short panels the fixed effects estimates are inconsistent. However, as suggested to us by Jonah Gelbach, these decompositions are functionals of the full set of estimated fixed effects and thus they should satisfy large sample properties.

there is some evidence that female dominated social networks tend to favour 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 (Card et al., 2015). 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% at the 10th percentile, to -4.9% at the median, to -11.4% at the 90th percentile (Table A3 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 considerable role in wage variation. The notion of job-title summarises 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% difference for the 10th percentile, -3.9% at the median, and just -2.4% 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 labour 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 due mostly 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 fell, 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 earlier methods

The most commonly used procedure to investigate the sources of the gender pay gap has been the familiar Oaxaca-Blinder decomposition into differences in labour market attributes (experience, seniority within the firm, schooling, etc.) and differences in the returns to those attributes. The latter is interpreted as the unexplained part of the gap or “discrimination” component.

Our approach progresses from this tradition along two routes. On one hand, from a methodological point of view, the Oaxaca-Blinder decomposition can be seen as a particular way to establish the Gelbach decomposition, with the additional benefit of providing a framework for statistical inference (for more details, see Gelbach (2016) section 5.2.). On the other hand, the set of variables considered in our analysis provides a richer insight on the current sources of the gender pay gap. Indeed, the set of variables considered in the traditional approach has proven informative at times (and in countries) where there was indeed a gap in the attributes of males and females. However, even then this approach provided weak 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 labour market attributes across genders, the method is no longer as revealing – 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 labour 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 human capital

endowments as determinants of the gender wage gap; while 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 Stanley and Jarrell (1998) and Jarrell and Stanley (2004) provide a particularly interesting core result. Whereas the original study showed that the correction for labour market participation selection led to a lower estimated 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 greater female employment participation, the correction for labour market participation selection tends to have a smaller impact on the gender pay gap than in countries where female participation is low.⁵

In Portugal, the convergence in male and female observable attributes is especially transparent. Even their selection into the labour market at every stage of the life-cycle seems to obey the same rules. This indication raises the need to consider alternative tools of analysis that can better explore the sources of the gender gap. In this paper, the role that the allocation of workers to firms and to jobs—a very fine partition of occupations for purposes of wage bargaining—are thoroughly scrutinised. We therefore highlight the decisions on labour market allocation that may influence the gender pay gap.

7 Conclusion

By combining the estimation of wage regressions with sets of high dimensional fixed effects and Gelbach’s (2016) decomposition method, we provide 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 labour market experience and the same seniority within the firm is due to their allocation to firms of different quality; a similar 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 operates mostly through worker allocation to firms, with lower access of females to higher paying

⁵In their analysis based on survey data, Olivetti and Petrongolo (2008) report that 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, p. 623, Figure 1).

firms, rather than through worker allocation to jobs.

Funding

This work was supported by the Spanish Ministry of the Economy and Competitiveness [CO2012-38460 to A.C.], the Severo Ochoa Programme for Centres of Excellence in R&D [SEV-2011-0075 to A.C.] and the Research Council of Norway [Europe in Transition funding scheme project 227072/F10 at ESOP to A.C.].

Acknowledgements

The authors thank Jonah Gelbach for comments that helped improve the paper and Lucena Vieira for exceptional data support.

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, Washington, D.C.
- Abowd, J. M., Kramarz, F., Margolis, D. N., 1999. High wage workers and high wage firms. *Econometrica* 67, 251–334.
- Addison, J. T., Centeno, M., Portugal, P., 2004. Reservation wages, search duration, and accepted wages in Europe, IZA Discussion Paper 1252, Institute for the Study of Labor (IZA), Bonn.
- Addison, J. T., Portugal, P., Vilares, H., March 2015. Unions and Collective Bargaining in the Wake of the Great Recession. IZA Discussion Papers 8943, Institute for the Study of Labor (IZA), Bonn.
- Altonji, J. G., Blank, R. M., 1999. Race and gender in the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Vol. 3. Amsterdam: 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* 5, 1-34.
- Autor, D. H., 2013. The “task approach” to labor markets : an overview. *Journal for Labour Market Research* 46, 185–199.
- Autor, D. H., Katz, L. F., Kearney, M. S., 2006. The Polarization of the U.S. Labor Market. *American Economic Review* 96, 189–194.
- Baldwin, M. L., Butler, R. J., Johnson, W. G., 2001. A hierarchical theory of occupational segregation and wage discrimination. *Economic Inquiry* 39, 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, 887–922.
- Becker, G. S., 1957. *The Economics of Discrimination*. University of Chicago Press, Chicago, IL.
- Bertrand, M., 2011. New perspectives on gender. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Vol. 4B. Amsterdam: 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, 228–55.

- Bertrand, M., Hallock, K. F., 2001. The gender gap in top corporate jobs. *Industrial and Labor Relations Review* 55, 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.
- Blau, F. D., 1977. *Equal Pay in the Office*. Lexington Books, Lexington, MA.
- 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, 212–228.
- Card, D., Cardoso, A. R., Kline, P., 2015. Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, forthcoming.
- Cardoso, A. R., Portugal, P., 2005. Contractual Wages and the Wage Cushion under Different Bargaining Settings. *Journal of Labor Economics* 23, 875–902.
- Carneiro, A., Portugal, P., Varejão, J., 2014. Catastrophic job Destruction during the Portuguese Economic Crisis. *Journal of Macroeconomics* 39, 444–457.
- Carrington, W. J., Troske, K., 1998. Sex segregation in U.S. manufacturing. *Industrial and Labor Relations Review* 51, 445–464.
- Dustmann, C., Ludsteck, J., Schnberg, U., 2009. Revisiting the German Wage Structure. *The Quarterly Journal of Economics* 124, 843–881.
- Fortin, N. M., Huberman, M., 2002. Occupational gender segregation and women’s wages in Canada: An historical perspective. *Canadian Public Policy* 28, s11–s39.
- Gelbach, J. B., 2016. When do covariates matter? and which ones, and how much? *Journal of Labor Economics* 34, forthcoming.
- Goldin, C., Rouse, C., 2000. Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review* 90, 715–741.
- Goos, M., Manning, A., 2007. Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics* 89, 118–133.
- 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, 457–472.
- Guimarães, P., Portugal, P., 2010. A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal* 10, 628–649.
- Jarrell, S. B., Stanley, T. D., 2004. Declining bias and gender wage discrimination? a meta-regression analysis. *Journal of Human Resources* 39, 828–838.

- Kunze, A., 2005. The evolution of the gender wage gap. *Labour Economics* 12, 73–97.
- Manning, A., Swaffield, J., 2008. The gender gap in early-career wage growth. *Economic Journal* 118, 983–1024.
- Martins, P. S., 2014. 30,000 Minimum Wages: The Economic Effects of Collective Bargaining Extensions. IZA Discussion Papers 8540, Institute for the Study of Labor (IZA), Bonn.
- 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, 189–208.
- Olivetti, C., Petrongolo, B., 2008. Unequal pay or unequal employment? a cross-country analysis of gender gaps. *Journal of Labor Economics* 26, 621–654.
- Ransom, M., Oaxaca, R. L., 2005. Intrafirm mobility and sex differences in pay. *Industrial and Labor Relations Review* 58, 219–237.
- Stanley, T. D., Jarrell, S. B., 1998. Gender wage discrimination bias? a meta-regression analysis. *Journal of Human Resources* 33, 947–973.
- Torres, S., Portugal, P., Addison, J. T., Guimarães, P., 2013. The sources of wage variation: a three-way high-dimensional fixed effects regression model. Working Papers w201309, Banco de Portugal, Lisbon.
- Weeks, D. L., Williams, D. R., 1964. A note on the determination of connectedness in an n-way cross classification. *Technometrics* 6, 319–324.
- Weichselbaumer, D., Winter-Ebmer, R., 2005. A meta-analysis of the international gender wage gap. *Journal of Economic Surveys* 19, 479–511.
- Woodcock, S., 2008. Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. *Labour Economics* 15, 403–418.

Appendix

Table A1: Descriptive Statistics on Population under Study

	Males	Females
Mean Log Real Hourly Wage	0.323	0.081
(standard dev)	(.564)	(.499)
Age	38.1	35.4
(standard dev)	(11.3)	(10.4)
Tenure (months)	107.5	94.6
(standard dev)	(107.8)	(96.2)
Mean Schooling (years)	7.1	7.8
(standard dev)	(3.9)	(4.0)
Share 4 years schooling	38.4	31.2
Share 6 years schooling	21.4	21.6
Share 9 years schooling	16.5	16.8
Share high school	14.9	20.0
Share university	6.4	8.4
Number of observations	17,332,330	10,880,440

Table A2: Regressions (Log) Real Hourly Wages

	(1)	(2)
female	-.230 (.005)	
age	.053 (.001)	.020 (.00006)
age sq.	-.0006 (.00002)	-.0002 (7.88e-07)
tenure	.002 (.0001)	.0006 (4.41e-06)
tenure sq.	-1.32e-06 (1.61e-07)	-1.23e-06 (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

Standard-errors clustered at the firm level.

Table A3: 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

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

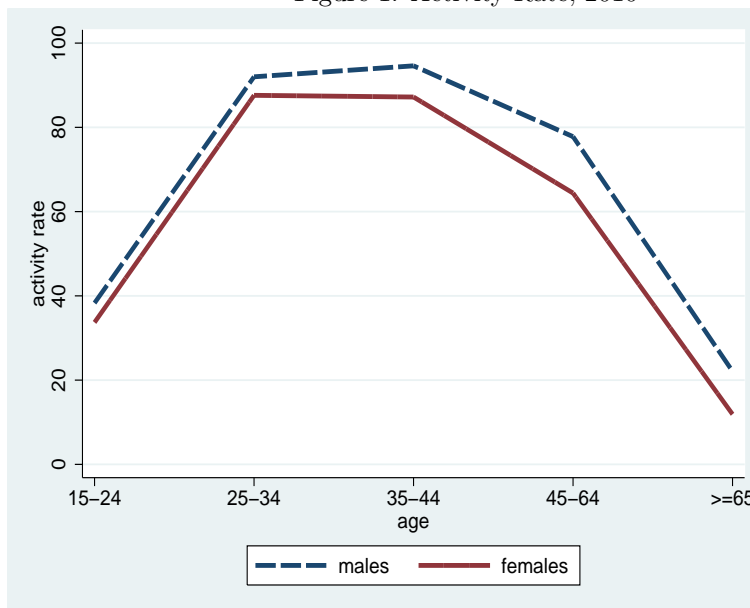
Tables and Graphs

Table 1: Conditional Decomposition of the Gender Wage Gap

Worker FE	Firm FE	Job FE
0.1444	0.0431	0.0425
(0.0001)	(0.0001)	(0.00004)

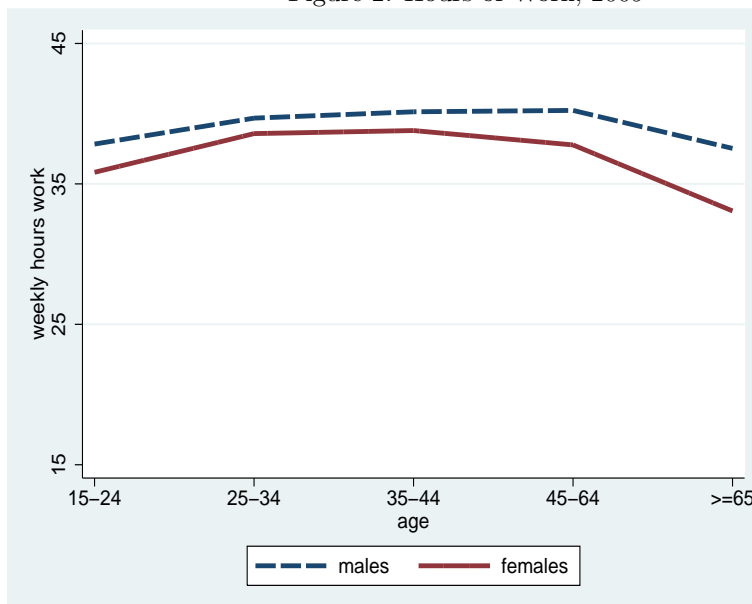
Decompositions based on Gelbach (2014). The basic model includes as regressors a quadratic term on age, a quadratic term on tenure, and year dummies, apart from the gender dummy variable. The full model further includes the sets of fixed effects specified. The number of fixed effects are: 576,459 for the firm; 107,785 for the job; and 4,138,799 for the worker.

Figure 1: Activity Rate, 2010



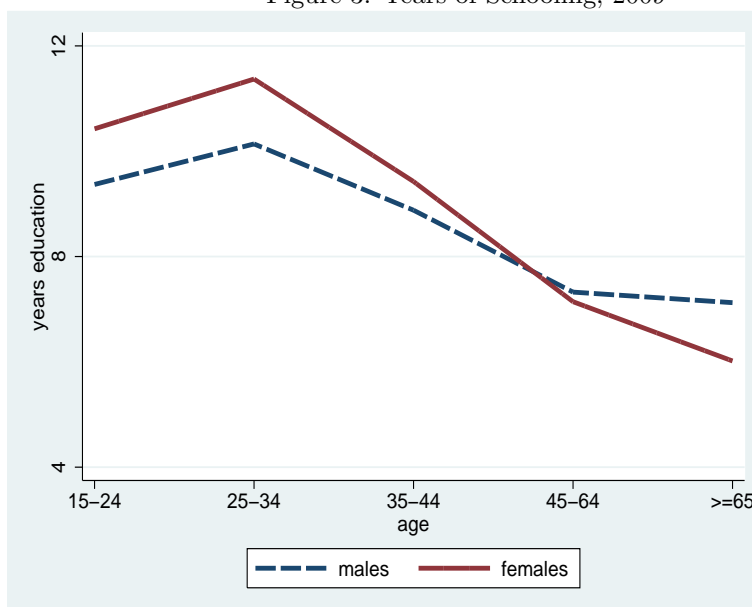
INE, Inquérito ao Emprego, <http://www.ine.pt>, April 3, 2012.

Figure 2: Hours of Work, 2009



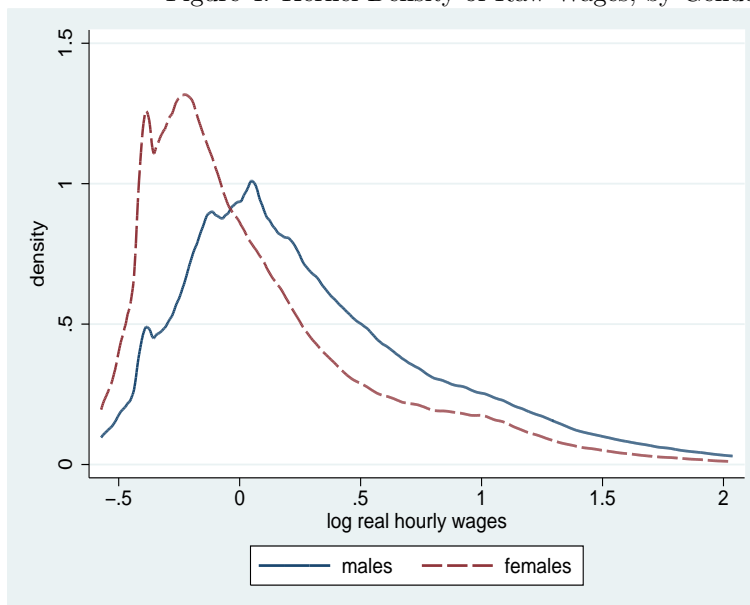
MTSS, *Quadros de Pessoal* 2009.

Figure 3: Years of Schooling, 2009



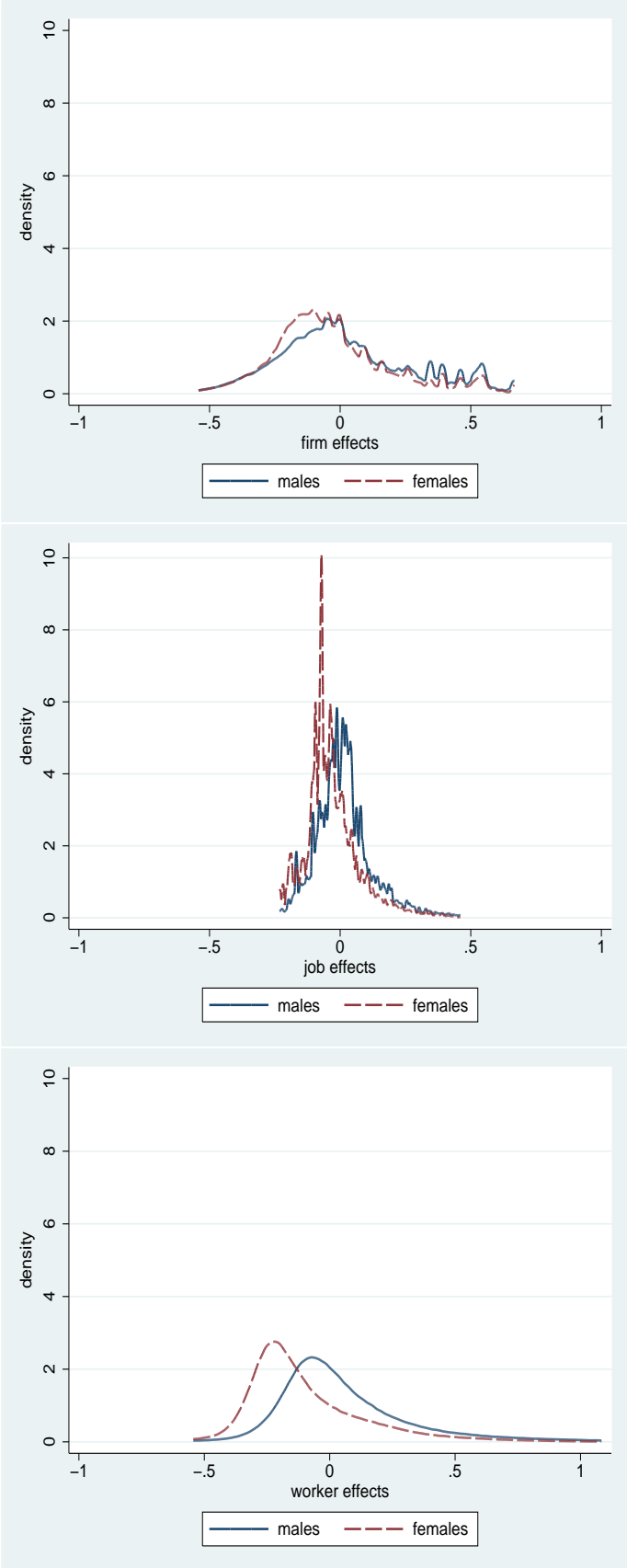
MTSS, *Quadros de Pessoal* 2009.

Figure 4: Kernel Density of Raw Wages, by Gender



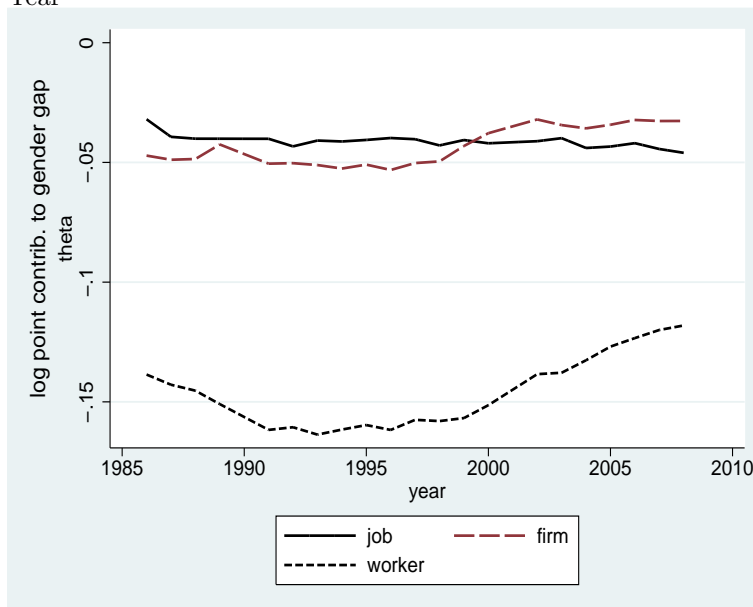
The plot excludes wages below percentile 1 of the female distribution and above percentile 99 of the male distribution. Both kernels rely on a common bandwidth.

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



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



See note to Table 1.