The sources of the gender wage gap

Ana Rute Cardoso
Institute for Economic Analysis, Barcelona GSE

Paulo Guimarães
Banco de Portugal

Pedro Portugal
Banco de Portugal, Universidade NOVA de Lisboa and IZA Bonn

Pedro S. Raposo
UCP - Catolica Lisbon School of Business and Economics

April 2016

Abstract
In Portugal, over the last two decades, the proportion of women among employed workers increased from 35 to 45 percent. This evolution was accompanied by a sharp fall in the gender wage gap from 32 to 20 percent. The improvement in the wage outcome of the women, however, is fully accounted by the catching up of their skills in comparison to males, after two decades of human capital investments. By 2013 women already possess observable characteristics that enhance productivity identical to their male counterparts. This means that gender discrimination remained roughly constant over the 1991-2013 period. In this study, we investigate the sources of the wage gender gap and conclude that sorting among firms and job-titles can explain about two fifths of the wage gender gap. (JEL: J16, J24, J31, J71)

"Um dos aspectos da desigualdade é a singularidade - isto é, não o ser este homem mais, neste ou naquele característico, que outros homens, mas o ser tão-somente diferente dele."

"Os espíritos altamente analíticos vêem quase só defeitos: quanto mais forte a lente mais imperfeita se mostra a cousa observada."

Fernando Pessoa

Introduction

In 1991, the wages of Portuguese women used to be around two thirds of the wages of men. Since 1991, women dominated the labor market inflows, particularly the better skilled women. This evolution translated into a 10

Acknowledgements: The authors are grateful for all the help provided by António Antunes, Hugo Reis, Lucena Vieira and Hugo Vilares and benefit from insightful discussions with the students from the Labor Economics class at NOVA SBE.

E-mail: anarute.cardoso@iae.csic.es; pfguimaraes@bportugal.pt; pportugal@bportugal.pt; pedro.raposo@ucp.pt

percentage point increase in the feminization rate of the stock of employed workers (see figure 1) and a 12 percentage point decrease in the raw wage gender gap (blue line, in figure 2). By 2013, the average wages of women represented about four fifths of the wages of men.

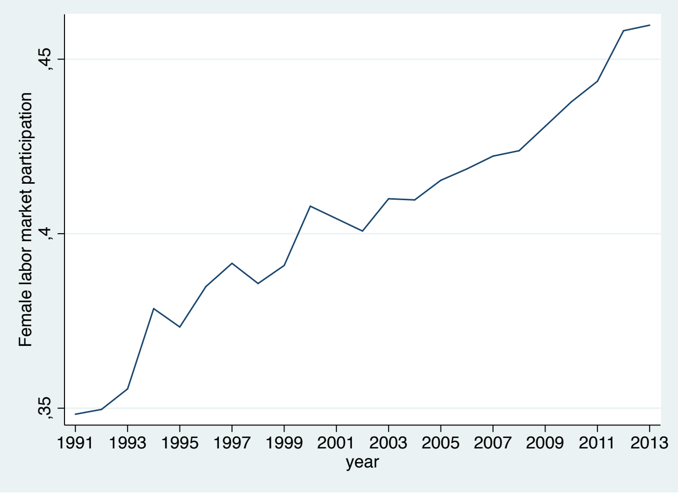

FIGURE 1: Female labor market participation

If, however, we take into account the characteristics of the workers to compute an "adjusted" gender wage gap, there is no longer an indication of improvement (red line, figure 2). In other words, the wage gain achieved by women over this period is due to the catching up of their skills (labor market experience, seniority, etc.) in comparison to males not by a reduction in unexplained component of the wage difference, which is conventionally equated to gender discrimination. Gender discrimination, in this sense, did not ameliorate, it slightly deteriorate over the 22 years period.

In this study we aim to study what hides behind the gender wage gap by executing a number of wage decomposition exercises. Firstly, we shall exploit the Machado and Mata (2005) quantile decomposition methodology to disentangle the role of the structural from the composition effects along the quantiles of the wage distribution. Secondly, we will combine the estimation of high-dimensional fixed effects regression models with the omitted variable bias decomposition suggested by Gelbach (2016) to access the importance of sorting into firms with heterogeneous wage policies and into job titles which are associated with different wage differentials. In this sense, we are updating and extending the work of Cardoso, Guimarães, and Portugal (2016).Thirdly, and finally, we will adapt the methodology of Guimarães and Portugal (2010) to incorporate the notion of high-dimensional slope effects, measuring gender wage gaps at the firm and the job-title levels.

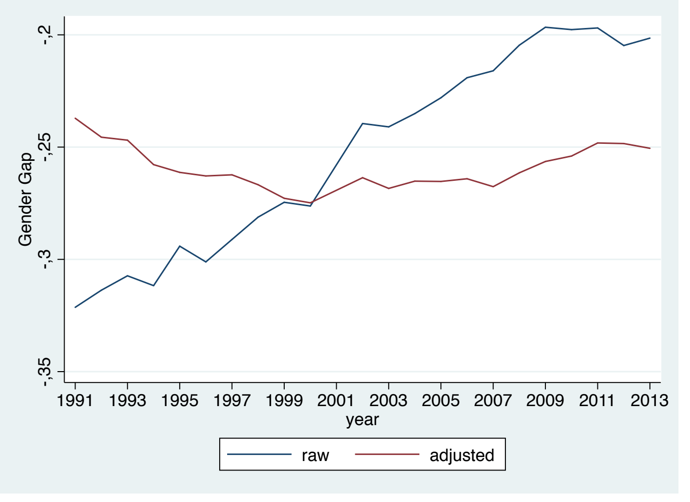

FIGURE 2: Gender wage discrimination

For this purpose, we rely upon an unusually rich data set, the "Quadros de Pessoal" survey, a longitudinal matched employer-employee- job title dataset, which covers all the establishment with at least one wage earner. The information about wages is provided by the employer annals. It is a complete and reliable source, because the main reason for its existence is to allow the officials from the Ministry of Employment to verify whether the employers are complying with the wage floors established by the collective agreement for the job-title of the worker.

The next section makes a brief literature review. 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. Section 6 concludes.

Literature review

It seems fair to claim that we are witnessing a revival of interest in the search for the determinants of the gender pay gap, under new empirical approaches, richer data, and renewed theoretical perspectives. Indeed, traditional economic analysis had focused primarily on the importance of female labor force participation and differences in observable attributes between men and women on the gender pay gap. Either of these two mechanisms can be understood intuitively. If the female participation rate is low, there is scope for the attributes of employed women to be unrepresentative of those of the female population in general. This selection could operate to raise or lower females' wages relative to males, depending on whether social norms, preferences, economic conditions and public policies disproportionately attract into the labor market more or less qualified women (along dimensions that can be observable or unobservable). In any case, as the female participation rate increases, the importance of selection influencing the gender pay gap is expected to decline (see the cross-country evidence in Olivetti and Petrongolo (2008) or the evidence over time for the US in Stanley and Jarrell (1998) and Jarrell and Stanley (2004)). Concomitantly, the qualifications of females and males in the labor market will influence their relative pay (see the ample evidence that education and experience contribute to shape the gender pay gap, in the review by Altonji and Blank (1999). Under this strand of literature, the convergence in schooling achievement across males and females (if not the reversal of the gap, in favor of women) and the increased female labor force attachment would lead us to expect the closing of the gender pay gap. Strikingly, a question lingers on: Why is the gender pay gap so persistent, despite a marked convergence in the participation rates and observable labor market attributes of men and women, in particular in developed economies?

A recent surge of literature addresses that question. Blau and Kahn (2016) report on the partial closing of the gender pay gap in the US in recent decades, remarkably so during the 1980s. Their empirical analysis, together with a review of the recent literature for other countries, points to a set of stylized facts and remaining challenges.

First of all, the convergence in attributes such as education and experience played a key role reducing the gender pay gap. These factors have currently a muted impact on pay differences between men and women. On the contrary, the industry and the occupation strive as factors generating pay differences across gender. Further research is thus needed to fully understand the allocation of gender across industries and occupations and their associated pay. Most likely, a better understanding of firm recruitment and pay policies will be helpful. A third noteworthy fact is that the gender pay gap is persistently larger at the top of the skill and wage distribution. The sources of this "glass ceiling effect" are also not yet fully understood. Plausible explanations highlighted by Blau and Khan include differences in psychological attributes (for example, competitiveness and bargaining power) that would penalize women at the top of the skill and job ladder, compensating differentials for characteristics of the top jobs (for example, longer and less flexible working hours), and pure discrimination.

Progress on some of the pending issues has recently been facilitated by availability of large longitudinal linked employer-employee datasets. Cardoso *et al.* (2016) (CGP) quantify the impact of access to firms and detailed jobs on the gender pay gap. They depart from the idea that different firms adopt different pay standards and assume that this generosity of the firm pay policy is common across gender and can be captured by a firm-specific fixed effect in a wage regression. Their subsequent step is to compare the average firm wage effect for males and females. They conclude that gender allocation to firms of different pay standards accounts for 20% of the overall gender pay gap. Similarly, the sorting or segregation of males and females across job titles accounts for approximately another 20% of the gender pay gap. Their quantification of the impact of worker allocation to firms and to jobs takes into due account the heterogeneity in worker quality. Their exercise is accomplished by adapting the methodology in Gelbach (2016), which allows for an unambiguous decomposition of the gender pay gap.

Card *et al.* (2016) (CCK) progressed along a different dimension. They aimed at formally testing a hypothesis long discussed in other fields of science, which made its entry into economic analysis more recently, namely, that females would have non-observable skills (such as competitiveness and bargaining attitudes) that would penalize them in the labor market vis a vis men. If so, women would extract lower rents from their employer than men working for the same firm. Accordingly, CCK allow for gender-specific firm wage premia and link these premia to measures of firm performance. Their analysis thus uncovers two channels contributing to the gender pay gap: the allocation of workers to firms (sorting or segregation channel) and the bargaining channel. Their decomposition of the pay gap is performed by relying on the following counterfactual exercises: by imposing the male firm wage premium on females in the same firm, they "shut down" the bargaining channel; similarly, by imposing an even distribution of males and females across firms, they "shut down" the allocation channel. The exercise requires firms that employ both males and females and it thus excludes single-gender firms.[1] They conclude, on one hand, that the bargaining effect accounts for 5% of the overall gender pay gap in Portugal. On the other hand, they confirms the relevance of the firm sorting channel, as it accounts for 15% of the overall pay gap.

Another recent strand of literature explores the role of compensating differentials for characteristics of the top jobs, in particular longer and less flexible working hours. Goldin (2014) and Bertrand and Katz (2010) are among the studies that present compelling evidence on the importance of this channel.

The aim of the current paper is to progress along the new strand of literature that relies on large longitudinal linked employer-employee data to evaluate the role of the firm shaping the gender pay gap.

1. A further requirement is that these firms are "connected" by workers of either gender moving across firms.

Data

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, and covers virtually all firms employing paid labor in Portugal. Due to the mandatory nature of the survey, problems commonly associated with panel data sets, such as panel attrition, are considerably attenuated.

The data set includes both firm-specific information (location, industry (SIC codes), legal setting, foreign ownership, employment, sales, ownership type) and and each and every one of its workers (labor earnings, worker qualifications, gender, age, tenure, hours of work, etc.). 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 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 Goos and Manning (2007), Autor *et al.* (2006) and Dustmann *et al.* (2009) on job polarization]. 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).

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 2013.[2] 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. Fifth, we excluded observations whose firm-job-title match included only one worker. Finally, we dropped (around 1 percent of the total number of) observations

2. The years between 1986 and 1989 were only used in order to obtain with more precision the estimates of the three high dimensional fixed effects in equation 3.

that did not belong to the largest connected group. Our final sample included 27,921,002 observations (338,580 firms; 5,126,998 workers; 95,196 job titles).

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.

High-Dimensional fixed effects and Gelbach's decomposition

In this section we follow closely the empirical approach of Cardoso *et al.* (2016). The idea consists of employing Gelbach's (2016) decomposition to help sort out the root causes of the observed gender wage gap. The novelty here is the application of Gelbach's decomposition to a linear regression model that accounts for the main sources of variation including unobserved components that are captured by the inclusion of multiple high-dimensional fixed effects. Our departure point is the traditional workhorse Mincerian wage equation:

$$\ln w_{ifjt} = \mathbf{x}_{ift}\boldsymbol{\beta} + \gamma g_i + \varepsilon_{ifjt} \quad . \tag{1}$$

In the above equation, $\ln w_{ifjt}$ stands for the natural logarithm of the real hourly wage. The various indices attached to w serve to emphasize all potential sources of wage variation. The index i $(i = 1, ..., N)$ stands for the worker, f $(f = 1, ..., F)$ accounts for firms while j reflects the variation accrued by differences in job titles. The index t stands for time $(t = 1, ..., T)$. The vector of explanatory variables, $\mathbf{x}$, comprises both observed characteristics of the worker and of the firm. These include variables such as worker education and tenure as well as firm size. Intentionally, we leave out of the vector $\mathbf{x}$ the variable g, a dummy that accounts for gender differences. The coefficient associated with this variable is the focus of our analysis as it provides the standard estimate for the gender wage gap. Finally, it is assumed that the error term, ε_{ifjt} follows the conventional assumptions.

It is more convenient to express the above equation in matrix terms. In doing so we obtain

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{G} + \boldsymbol{\varepsilon} \tag{2}$$

where the symbology used is quite obvious. The above specification is what we call the base model and is the regression typically used the ascertain the size of the gender wage gap. Basically, it estimates the percentage difference between the wages of men and women once we take into account the observed characteristics of the workers such as their education level and tenure and important firm characteristics such as size. However, in line with the work of Abowd *et al.* (1999), we recognize the need to explicitly account for all wage variation emanating from factors that are specific to the worker and the firm. This can only be accomplished with employer-employee data. As shown by Abowd *et al.* (1999) with the introduction of fixed-effects for firm and worker we are able to control for time-invariant characteristics of workers and firms whether or not we are able to observe them. In this framework, things such as worker ability, family background, risk aversion, etc. are all accounted for. The same applies to firm unobserved characteristics, such as managerial ability and organization, location, etc. The richness of our data allows us to go a step further. As explained earlier, since we have detailed job title information we are also able to introduce a fixed effect that absorbs all time-invariant characteristics of a specific job-title.

Adding firm or job-title fixed effects to the base equation in 2 should not affect the estimate of γ unless there is an uneven distribution of gender across firms and job-titles. Put differently, if γ changes when we fully control for firm and job-title effects than this means that the sorting of females/males across firms or jobs is a factor that is contributing to the gender wage gap. But, to account for the main sources of variation the full regression model needs to also include a worker fixed effect. With the introduction of an individual specific fixed effect we will absorb all time-invariant individual specific characteristics, including the gender dummy variable (G). As we will see below, this does not prevent us from understanding what happens to γ when we control for all three additional sources of variation (worker, firm and job title). In order to do this we need to estimate a full model, one that includes the three fixed effects. This model is simply

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\varphi} + \mathbf{L}\boldsymbol{\lambda} + \boldsymbol{\varepsilon} \tag{3}$$

where we have added three high-dimensional fixed effects to the equation in (2). $\mathbf{D}$ is a design matrix for the worker effects, $\mathbf{F}$ is design matrix for the firm effects while $\mathbf{L}$ is a design matrix for the job title effects. As usual, we maintain the assumption of strict exogeneity of the error term.

The large size of our data, with around 28 million observations, more than 5 million workers and 400 thousand firms, and around 95,000 distinct job-titles, raises some econometric challenges. Of particular concern is the high-dimensionality of the fixed effects. Estimation of a regression with three high-dimensional fixed effects is a non-trivial issue given the size of the matrices involved. The within transformation can absorb one of the fixed effects but the large dimension of the remaining fixed effects prevents the application of the conventional OLS formula. Estimation of this model is possible if we resort to the algorithm of Guimarães and Portugal (2010). This algorithm is able to provide the exact OLS solution without requiring the inversion of large matrices. [3]

3. We used the user-written command `reghdfe` coded by Sergio Correia which implements an improved version of the Guimarães and Portugal (2010) algorithm.

Since we provide secondary analysis of the estimates of the fixed effects we have to make sure that they are identifiable. This is done by restricting our data set to a connected subset. We accomplish this by using an algorithm proposed by Weeks and Williams (1964). Application of this algorithm to our data permits the identification of a subset of the data that comprises 99% of our original data set. Within this subset of data the estimates of all fixed effects are comparable up to an additive scalar factor.

The Gelbach (2016) decomposition can help us understand what happens to the estimate of γ when we move from the basic equation in (2) to the full equation (3) where the three fixed effects are simultaneously added. The approach is based on the OLS formula for omitted variable bias and has the advantage of providing an unequivocal way to quantify the parcel of change that can be attributed to the inclusion of each individual fixed effect. To see how the decomposition can be employed in this context we recall that by the Frisch-Waugh-Lovell (FWL) theorem it is possible to obtain an estimate of the γ in the base model by running a two step regression. First, we regress $\mathbf{Y}$ on $\mathbf{X}$ and calculate the residual of that regression. If we let $\mathbf{M} \equiv [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ be the residual-maker matrix then this amounts to calculating the vector $\mathbf{MY}$. Similarly, we calculate the residual of the regression of $\mathbf{G}$ on $\mathbf{X}$, that is, $\mathbf{MG}$. With this procedure we have expurgated the effect of the $\mathbf{X}$ variables from $\mathbf{Y}$ and $\mathbf{G}$. Thus, if we now run a simple linear regression of $\mathbf{MY}$ on $\mathbf{MG}$ we know by the FWL theorem that we obtain the OLS estimate for the γ in our base model. That is,

$$\widehat{\gamma} = (\mathbf{G}'\mathbf{MG})^{-1}\mathbf{G}'\mathbf{MY} = \mathbf{M_G Y} \tag{4}$$

where we note in passing that $\mathbf{M_G} \equiv (\mathbf{G}'\mathbf{MG})^{-1}\mathbf{G}'\mathbf{M}$ and $\mathbf{M}$ is an idempotent matrix. We now turn to the full version of the wage equation model in (3). The fitted version of this model can be expressed as

$$\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{D}\widehat{\boldsymbol{\theta}} + \mathbf{F}\widehat{\boldsymbol{\varphi}} + \mathbf{L}\widehat{\boldsymbol{\lambda}} + \widehat{\varepsilon} \tag{5}$$

where we have replaced the coefficients and error term by their OLS estimates. Note that $\mathbf{D}\widehat{\boldsymbol{\theta}}$, $\mathbf{F}\widehat{\boldsymbol{\varphi}}$ and $\mathbf{L}\widehat{\boldsymbol{\lambda}}$ are the column vectors containing the estimates of the fixed effects. To implement Gelbach's decomposition we simply have to pre-multiply the above expression by $\mathbf{M_G}$. When we do this we obtain on the left-hand side the formula for the OLS estimate of γ while on the right-hand side the terms associated with $\mathbf{X}$ and $\widehat{\varepsilon}$ disappear. [4] We are left with three components, each one associated with one of the fixed effects, that add up to the observed gender wage gap, $\widehat{\gamma}$. That is,

$$\widehat{\gamma} = \widehat{\delta}_{\boldsymbol{\theta}} + \widehat{\delta}_{\boldsymbol{\varphi}} + \widehat{\delta}_{\boldsymbol{\lambda}} \quad . \tag{6}$$

4. By construction $\widehat{\varepsilon}$ is orthogonal to $\mathbf{X}$ and to $\mathbf{D}$ meaning that it is also orthogonal to $\mathbf{G}$. It follows that $\mathbf{M_G}\widehat{\varepsilon} = \mathbf{0}$. Using the fact that $\mathbf{MX} = \mathbf{0}$ it is easy to show that $\mathbf{M_G X} = \mathbf{0}$.

In practical terms each $\widehat{\delta}$ in the left-hand side is the coefficient of a regression between the respective fixed effect and the gender variable adjusting for the $\mathbf{X}$ covariates. If, conditional on the $\mathbf{X}$ variables, the distribution of females across firms was absolutely random then we would expect $\widehat{\delta_{\varphi}}$ to be zero. This would mean that the sorting of females/males across firms was not a contributing factor to the gender pay gap. A similar reasoning can be applied to the sorting of gender across jobs.

Discussion of the results

The Machado and Mata decomposition

We rely on quantile regression methods to analyse the changes in the wage distribution between gender over a 22 year period. To that end, we use the Machado and Mata decomposition method which enables us to identify the sources of the changes in the distribution of wages between females and males. We repeat the exercise in 1991 and in 2013 in order to compare how the sources of variation have evolved between the beginning of the period (1991) and 22 years later (2013).

Gender differences in the distribution of wages may result from changes in the distribution of the conditioning variables (changes in terms of the characteristics of the population, e.g. labor force characteristics such as education and age) or from changes in the conditional distribution of wages itself (which may be thought of as changes in the way characteristics impact wages, the "coefficients"). The first is a "composition effect" and the second may be thought of as a "structural effect" (Autor *et al.* (2008)). We build the counterfactual exercise by estimating the marginal distribution of wages that would have prevailed for male if they had the characteristics of females ("composition effect"). Subsequently, we estimate the marginal distribution of wages that would have prevailed for female if they had the same returns than males ("structural effect").

In 1991, men earned more than women, most notably at higher percentiles. Whereas males earned more 35.1 log points than females at the median, the difference was 41.7 log points at the 8th decile (see the third column of Table 1). It is clear from columns 4th and 5th that (aggregate) differences in the coefficients were more influential driving the overall shift in the wage distribution than (aggregate) differences in the covariates. At the median, the gender wage gap was 10.9 log points due to changes in covariates and it was 24.2 log points due to changes in the coefficients. Interestingly, "covariate changes" are larger at the 1st decile but "coefficient differences" become more influential as we move up the wage distribution. The "coefficient changes" generated a larger gender gap at the highest percentiles.

	(1) Female x0b0	(2) Male x1b1	(3) (2)-(1)	(4) Aggregate composition effect (x1b1-x0b1)	(5) Aggregate structural effect (x0b1-x0b0)
10 percentile	-0.433***	-0.268***	0.165***	0.090***	0.074***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
20 percentile	-0.351***	-0.116***	0.235***	0.096***	0.139***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
30 percentile	-0.274***	0.010***	0.284***	0.101***	0.183***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
40 percentile	-0.191***	0.130***	0.322***	0.105***	0.216***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
50 percentile	-0.099***	0.251***	0.351***	0.109***	0.242***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
60 percentile	0.008***	0.384***	0.375***	0.111***	0.264***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
70 percentile	0.142***	0.539***	0.397***	0.114***	0.282***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
80 percentile	0.321***	0.737***	0.417***	0.117***	0.300***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
90 percentile	0.602***	1.036***	0.433***	0.123***	0.310***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)

TABLE 1. Gender wage discrimination decomposition (1991)

	(1) Female x0b0	(2) Male x1b1	(3) (2)-(1)	(4) Aggregate composition effect (x1b1-x0b1)	(5) Aggregate structural effect (x0b1-x0b0)
10 percentile	-0.163***	-0.064***	0.099***	0.021***	0.078***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
20 percentile	-0.063***	0.067***	0.130***	0.014***	0.117***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
30 percentile	0.028***	0.186***	0.158***	0.008***	0.150***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
40 percentile	0.124***	0.306***	0.183***	0.002**	0.181***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
50 percentile	0.229***	0.434***	0.205***	-0.004***	0.209***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
60 percentile	0.349***	0.575***	0.226***	-0.009***	0.235***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
70 percentile	0.495***	0.739***	0.244***	-0.015***	0.259***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
80 percentile	0.684***	0.947***	0.262***	-0.016***	0.279***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
90 percentile	0.968***	1.256***	0.288***	-0.013***	0.301***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)

TABLE 2. Gender wage discrimination decomposition (2013)

	1991		2013	
	Female	Male	Female	Male
Age	33,98	38,27	40,25	40,73
Tenure	8,87	10,17	9,47	9,55
Firm size	5,09	5,50	4,82	4,77
Education	6,36	6,27	9,86	9,29

TABLE 3. Gender wage discrimination: Summary statistics (Composition)

	1991		2013	
	Female	Male	Female	Male
Age	0,0190	0,0470	0,0267	0,0452
Age^2	-0,0001	-0,0005	-0,0002	-0,0004
Tenure	0,0095	0,0104	0,0171	0,0216
$Tenure^2$	-0,0002	-0,0001	-0,0002	-0,0002
Firm size	0,0422	0,0629	0,0258	0,0426
Education	0,0739	0,0810	0,0723	0,0783
Fraction of Females	-0,2527	-0,1031	-0,2537	-0,1023
Constant	-1,0997	-1,6832	-1,2180	-1,7495

TABLE 4. Gender wage discrimination: Quantile regressions (β)

In 2013, the gender gap is still positive and statistically significant but its magnitude was reduced. Although males earn more 20.5 log points than females at the median, the difference between the highest and the lowest percentiles was reduced. In 2013 the "coefficient differences" are everywhere larger, in absolute magnitude, than "covariate differences" (Table 2).

Females in 2013 are not only more similar to their male counterparts but also show better characteristics (Table 3). Women in 2013 are older and more experienced reflecting the increase in their labor market participation rates. The educational level of the labor force increased considerably during this period reflecting the aging of the baby-boom generation. Women in 2013 are working in larger firms and they are clearly more educated than male. There are significant differences in the returns to education, both in 1991 and in 2013. Despite having similar characteristics the return to general and specific human capital is much lower for women in comparison with their male counterparts (Table 4). The high paying policies by large firms benefit males in a much larger extent than females. Finally, firms whose workforce is more heavily populated by women (more segregated) generate a wage penalty, most notably, for females.

The Gelbach decomposition

The sizable gender wage gap for total hourly earnings that we have estimated constitutes an average differential between the wages of two otherwise observably identical workers. A key question concerns the potential sources of the unobserved heterogeneity behind these differentials (see figure 3). We next consider how sorting among firms with different compensation policies, the assignment to distinct job titles, and the allocation of workers with different unobserved ability drive the gender wage gap. Our focus in decomposing the gender wage gap is therefore upon the contributions of each of these three sources of unobserved heterogeneity.

Before proceeding, it is worth to discuss the interpretation of the three high-dimensional fixed effects added in equation (3). The firm fixed effect, in essence, captures the (constant) wage policy of the firm. Firms with generous compensation policies will exhibit positive firm fixed effects, low-wage firms will generate negative fixed effects. In Figure 4 we contrast the distribution of the firm fixed effects for workers by gender.[5] It is very clear from the picture that males disproportionally populate high paying firms.

In Figure 5 the empirical distribution of the worker fixed-effects are presented. The worker fixed effects condense the influence of constant characteristics (observed and non-observed) of the individuals on their wages. They can be a proxy for the portable human capital (or productivity) of the worker or they may simply reflect gender discrimination that is not associated with sorting of workers across firms and job titles. The picture shows the wage gap between males and females is firmly rooted in the individual component of wages, more notably in the upper tail of the distribution. This outcome can be the result of observed or unobserved characteristics (say, schooling or ability). We shall, below, identify the specific role of unobserved skills.

Finally, we show the empirical distribution of the job title fixed effects. Job title fixed effects largely reflect the remuneration status of disaggregated occupations. In a way, the inclusion of job title effects builds upon first generation Mincerian wage equation which included broad definition of occupations. In the current setup, we provide an unusually fine accounting of the tasks required to fill a job. The distributions of the job title fixed effects given in Figure 6 do exhibit a discernible difference in terms gender, suggesting that the allocation of workers across job titles significantly disfavors women.

Results for the Gelbach decomposition are given in table 5. It can be seen that the wage penalty of 25.6 log points (arriving from the estimation of equation 1) can be decomposed into the contribution of three parts: worker, firm, and job title unobserved heterogeneity. A significant fraction of the gender wage gap is explained by the heterogeneity of the firms' compensation policies. The allocation of workers into firms is responsible for 5.8 out of 25.6 log points of the gender wage gap. This means that females disproportionally belong to firms firms with less generous wage policies. Put differently, if workers were randomly assigned to firms, the gender wage gap would be reduced by about one fifth. We also find that the attribution of job-titles, either through promotion policies or through initial assignments, is significantly influenced by gender, contributing 4.3 log points to wage gap. Together, the process of sorting into firms and job titles accounts for around 40 percent of the gender wage gap. The unobserved (permanent) characteristics of the individuals is responsible for the remaining 60 percent. These unobserved (to

5. Notice, however, that in this comparison the influence of variables such as industry or firm size are still subsumed in the firm fixed effect.

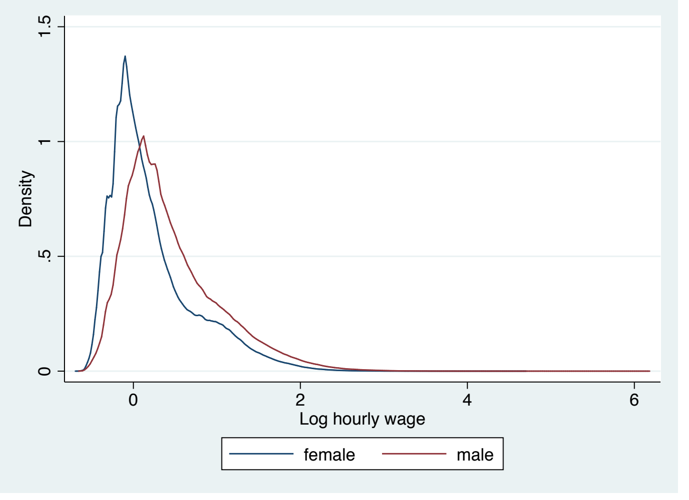

FIGURE 3: Gender wage discrimination: (log) real hourly wage

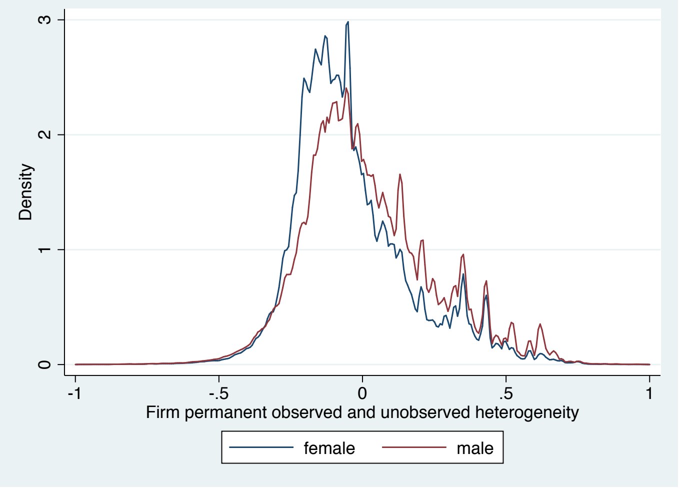

FIGURE 4: Gender wage discrimination: Firm

the researcher) worker characteristics can be equated either with unobserved skills or, simply, to some form of gender discrimination.

Figure 7 display the gender gap decomposition over time. The allocation of female workers into firms and job-titles did not improve over the last two decades. If anything, the sorting into firms and job-titles is now slightly less favorable for women (-1.7 and -1.0 log points for firms and for job-titles, respectively, over the 1991-2013 period). In compensation, the wage penalty resulting from the role of unobserved individual heterogeneity was visibly attenuated (3.2 log points), in particular since the beginning of the century.

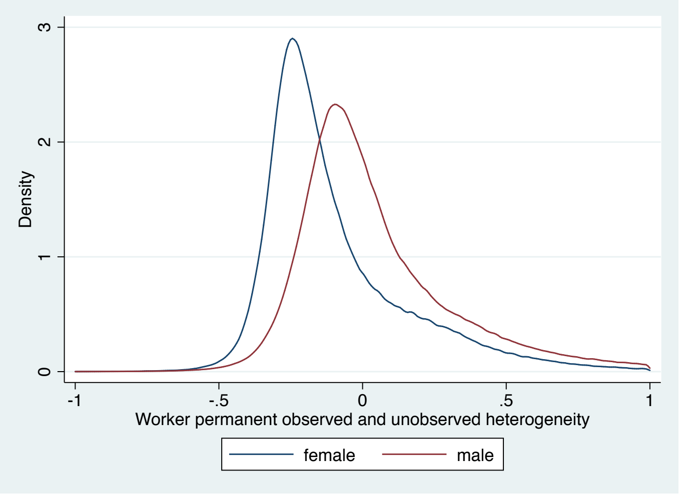

FIGURE 5: Gender wage discrimination: Worker

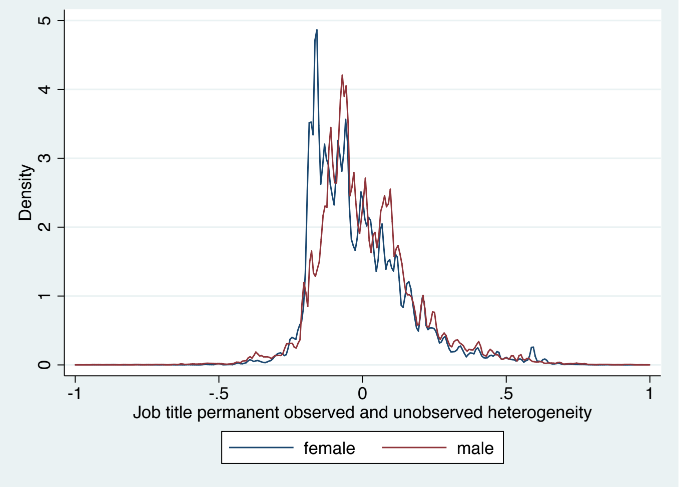

FIGURE 6: Gender wage discrimination: Job title

gap	worker fe	firm fe	job fe
-0.2560	-0.1547	-0.0580	-0.0433

Note: Decompositions based on Gelbach (2016).

TABLE 5. Conditional Decomposition of the Gender Wage Gap

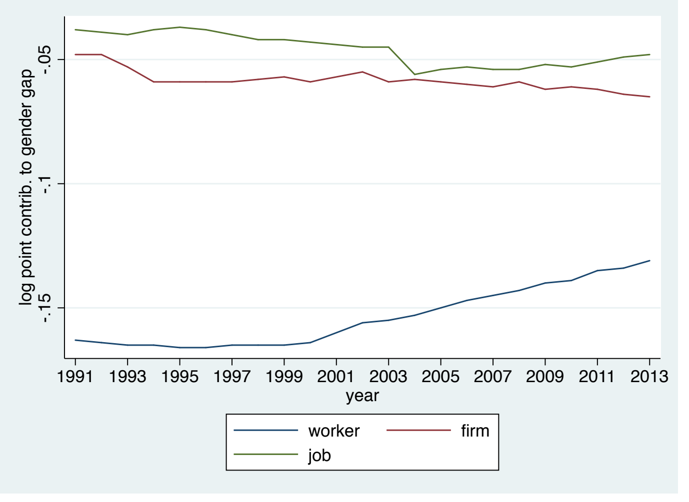

FIGURE 7: Conditional decomposition of the gender wage gap, separately by year

Overall, the combination of the evolution of the three sources of heterogeneity resulted in a tiny (0.5 log points) decrease in the gender wage gap over 22 years.

The gender gap heterogeneity

Whereas the approach based upon the high-dimensional fixed effect regression fully accounts for the distribution of workers across firms and job title cells, it is silent regarding the heterogeneity of gender gaps within the firm and job title. Firm specific gender gaps have been interpreted as evidence of gender discrimination that emerges from the shortfall in women's relative bargaining power (Card *et al.* (2016)). Here we extend our previous approach to accommodate the estimation of firm specific and job specific gender gaps. In essence, for the firm case, we estimate the following regression model:

$$\ln w_{ifjt} = \mathbf{x}_{ift}\boldsymbol{\beta} + \varphi_f + \gamma_f g_i + \varepsilon_{ifjt} \quad . \tag{7}$$

where equation (1) is generalized to include a firm fixed effect (φ_f) and a firm-specific gender effect (γ_f). It should be noted that we are not including a worker fixed effect and so the firm gender gap is not filtered from the presence of unobserved individual heterogeneity. The identification of the firm gender parameter in conjunction with the worker fixed effect would require additional normalization restrictions in order to retain a common scale.

The results from this procedure are exhibited in figure 8, where the empirical distribution of firm specific wage gender gaps for 1991 are contrasted with those of 2013. The histogram may be interpreted as the distribution of discriminating employers (or, in the sense of Card *et al.* (2016), as reflexion of the relative bargaining power of women). The graph indicates that most employers have negative gender wage gaps and that the distribution of the gender gaps only mildly improved from 1991 to 2013. It is interesting to notice that a non-negligible fraction of employers has positive gender gaps.

Whether this outcome signals the true distribution of discriminating employers or is just a product of sampling variation remains to be solved. An indication that it is not simply the consequence of sampling variation can be argued from the fact that firm specific gender gaps are highly correlated with the firm level segregation (-0.476). The notion that higher proportion of females leading to more negative firm gender gaps is consistent with the idea of a shortfall in women bargaining power.

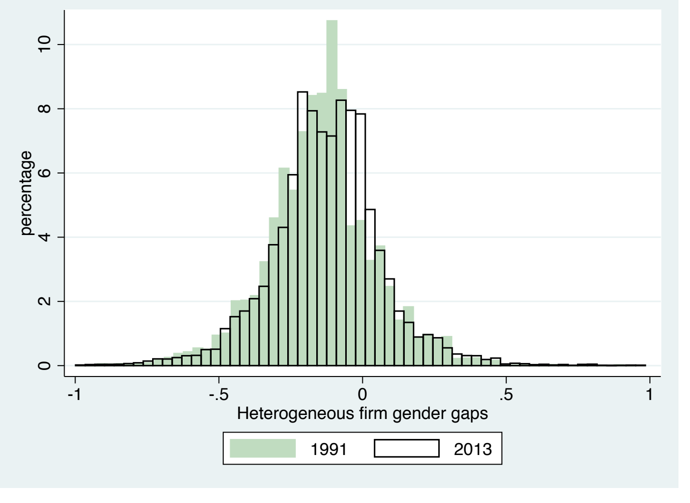

FIGURE 8: Heterogeneous firm gender gaps

The distribution of job title specific wage gender gaps is much less dispersed, in particular in 2013. In contrast with the firm gender gaps which are sensitive to the segregation at the firm level, these are poorly predicted by the measure of job title segregation (correlation equals 0.006). Whereas job title segregation leads to lower mean wages, it does not lead to larger gender gaps along the job title dimension. Put differently, whereas firm segregation leads to higher gender gaps, job title segregation leads to lower wages. This latter result is in line with Groshen (1991) and is consistent, for example, with the idea that some occupations may be overcrowded by women.

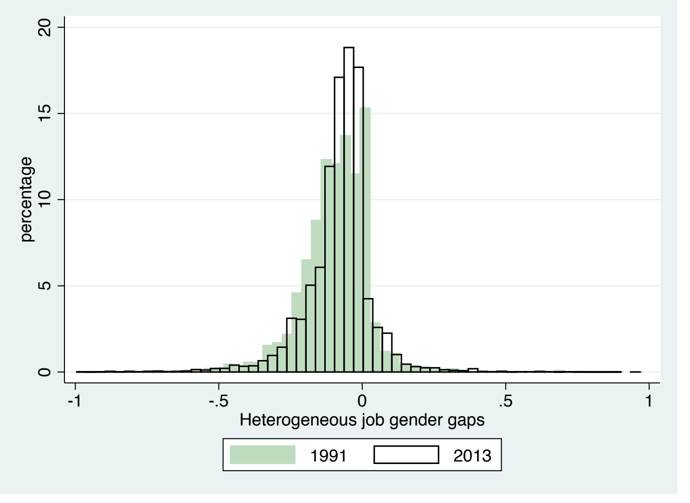

FIGURE 9: Heterogeneous job title gender gaps

Conclusions

Over the 1991-2013 period, there was a notable increase in the feminization rate of Portuguese labour market. At the same time, the average wages of women approached significantly those of the men. In this study, we argue that the fall in the gender wage gap is largely the result of a compositional change (not a structural effect), due to the fact that the women that joined the labor market detained higher level of general and specific human capital.

This means that the adjusted gender wage gap remained roughly constant at around 25 percent over the period. We show that gender plays an important role in the allocation of workers across firms with distinct wage policies. Indeed, if workers were randomly allocated to firms, the gender gap would be reduced by 5.8 percentage points. Similarly, if workers were randomly selected into job titles, the gender gap would be reduced by 4.3 percentage points. Overall, if workers were randomly sorted into firms and job titles, the gender gap would be reduced by about two fifths.

The allocation of female workers to firms and job-titles did not improve over the last two decades. In fact it deteriorate somewhat, since in 2013 females tend to be less present in firms and job titles with more generous wage policies. In compensation, the role of unobserved skills favored a small decrease of the gender gap. This may either reflect less gender discrimination or improved ability.

Firm segregation, that is, the feminization rate at the firm level, leads to higher firm specific gender gaps. In contrast, job title segregation leads to lower wages but not larger job title specific gender gaps.

References

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). “High Wage Workers and High Wage Firms.” *Econometrica*, 67(2), 251–334.

Altonji, Joseph G. and Rebecca M. Blank (1999). “Race and gender in the labor market.” In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, pp. 3143–3259. Elsevier.

Autor, David H. (2013). “The “task approach" to labor markets : an overview.” *Journal for Labour Market Research*, 46(3), 185–199.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). “The Polarization of the U.S. Labor Market.” *American Economic Review*, 96(2), 189–194.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). “Trends in U.S. Wage Inequality: Revising the Revisionists.” *The Review of Economics and Statistics*, 90(2), 300–323.

Bertrand, C. Goldin, M. and L. F. Katz (2010). “Dynamics of the gender gap for young professionals in the financial and corporate sectors.” *American Economic Journal: Applied Economics*, 2(3), 228–255.

Blau, F. and L. Kahn (2016). “The gender wage gap: Extent, trends and explanations.” *Journal of Economic Literature*, forthcoming.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). “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, Ana Rute, Paulo Guimarães, and Pedro Portugal (2016). “What drives the gender wage gap? A look at the role of firm and job-title heterogeneity.” *Oxford Economic Papers*, 68(2), 506–524.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). “Revisiting the German Wage Structure.” *The Quarterly Journal of Economics*, 124(2), 843–881.

Gelbach, Jonah (2016). “When do covariates matter? and which ones, and how much?” *Journal of labor Economics*, forthcoming.

Goldin, C. (2014). “A grand convergence: Its last chapter.” *American Economic Review*, 104(4), 1091–1119.

Goos, Maarten and Alan Manning (2007). “Lousy and Lovely Jobs: The Rising Polarization of Work in Britain.” *The Review of Economics and Statistics*, 89(1), 118–133.

Groshen, Erica 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.

Guimarães, Paulo and Pedro Portugal (2010). “A simple feasible procedure to fit models with high-dimensional fixed effects.” *Stata Journal*, 10(4), 628–649.

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

Machado, José A. F. and José Mata (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics*, 20(4), 445–465.

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

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

Torres, Sónia, Pedro Portugal, John T. Addison, and Paulo Guimarães (2013). "The sources of wage variation: a three-way high-dimensional fixed effects regression model." Working Papers w201309, Banco de Portugal, Economics and Research Department, URL `http://ideas.repec.org/p/ptu/wpaper/w201309.html`.

Weeks, David L. and Donald R. Williams (1964). "A Note on the Determination of Connectedness in an N-Way Cross Classification." *Technometrics*, 6(3), 319–324.