

GENDER WAGE DISCRIMINATION IN INDIA: GLASS CEILING OR STICKY FLOOR?

SHANTANU KHANNA

Email: shantanu@econdse.org

Department of Economics

Delhi School of Economics

Working Paper No. 214

Centre for Development Economics

Department of Economics, Delhi School of Economics

Gender wage discrimination in India: Glass ceiling or sticky floor?

Shantanu Khanna

Email: shantanu@econdse.org

Department of Economics¹
Delhi School of Economics
University of Delhi
New Delhi - 110007

June 2012

¹I am grateful to Professor Ashwini Deshpande and Dr. Deepti Goel of the Department of Economics, Delhi School of Economics for their detailed comments on this paper.

ABSTRACT

Traditional analysis of gender wage gaps has largely focused on average gaps between men and women, and mean wage decompositions such as the Blinder-Oaxaca (1973) decomposition method. To answer the question of whether there is a “glass ceiling” or a “sticky floor”, i.e. whether wage gaps are higher at the upper or lower ends of the wage distribution, this paper examines the wage gaps across different quantiles of the wage distribution. These gender wage gaps are analysed for regular wage workers in India using the 66th round of the National Sample Survey’s Employment - Unemployment Schedule (2009-2010). The paper finds evidence of a “sticky floor”. In addition to estimating the standard OLS wage equations for men and women, quantile regressions are used to assess how different covariates such as education, union membership, and occupations, affect within and between group (gender) inequalities. Finally, the Machado-Mata-Melly (2006) decomposition method is used to decompose gender wage gaps at different quantiles to determine whether it is the differences in characteristics (levels of covariates) or the unexplained (discrimination) component that drives the sticky floor effect. The paper concludes with a discussion on the possible reasons for observing a sticky floor phenomenon in India.

JEL Codes: J16, J31, J71, C21

Keywords: *wage decompositions, gender discrimination, glass ceiling, sticky floor, quantile regressions*

1 Introduction

Where in the wage distribution are wage gaps higher? Do low wage earners face higher wage gaps, or are wage gaps higher at the upper end of the distribution? In other words, do women face a sticky floor or glass ceiling? To what extent do differences in the labour market characteristics explain the gender wage gap in India, both at the mean, and at other points of the wage distribution? Do poorer women face greater gender discrimination in wages as compared to high wage earners? The motivation of this paper is to answer such questions.

While there are several papers that deal with gender wage inequality, most analyses focus on the average difference in wages between men and women. In this paper, I undertake a distributional analysis of gender wage gaps using wage data for regular wage workers from the National Sample Survey (NSS) Employment-Unemployment schedule (EUS) for the year 2009-2010. I find evidence of the sticky floor phenomenon, where gender wage gaps are higher at the bottom end of the distribution, and fall to smaller levels at the top. In order to design suitable policies to combat gender discrimination, it is instructive to know the distribution of gender discrimination over the wage distribution. Since gender gaps are higher at the lower end of the distribution in India, the impact on equity and poverty reduction by correcting these disparities is potentially substantial.

Apart from the OLS log wage regressions, I use quantile regressions to determine how returns to covariates (or wage determining characteristics) differ over the wage distribution, and how they differ between men and women as well. These regressions help us draw conclusions on the impact of different covariates on gender inequality. To answer the question of whether the sticky floor is due to differences in labour market characteristics of men and women or due to differences in returns to these characteristics, I use recent advances in decomposition techniques that focus on decomposing wage gaps at different parts of the distribution. Results indicate that the sticky floor effect persists even after controlling for various personal and job market characteristics, implying that poorer women are heavily discriminated against in India.

This paper begins with a literature review which focuses on decomposition methods and gender discrimination studies on the glass ceiling/ sticky floor phenomenon. I discuss papers that propounded these techniques as well as those that use them in the Indian context. The data and methodology section follow, with the latter explaining the basics of the decomposition techniques and the quantile regression techniques applied later in the paper. Section 5 of the paper gives results and section 6 concludes with some possible interpretations of the results.

2 Literature review

In their seminal papers, Blinder (1973) and Oaxaca (1973) were the first economists who made an attempt to arrive at a quantitative measure of wage discrimination. The Blinder-Oaxaca decomposition (hereafter, B-O) has been used in several studies on wage discrimination and retains much of its relevance even today as a summary measure of average wage discrimination. Details of the B-O decomposition technique are given in section 3.1.

Since the mid-1990s, newer decomposition methods have gone beyond the focus on single summary measures. Studies now seek to go beyond the mean and answer questions like *what happened where* in the distribution, rather than decomposing only differences at the means. Additionally, since policies could have different effects at different parts of the wage distribution, this nuanced understanding is needed to

assess their impact on wage inequalities. For example, a change in minimum wage regulations is more likely to influence the lower part of the wage distribution.

When considering wage gaps among different social groups, the mean gap may not turn out to be particularly representative of the gaps across different quantiles of the distribution. Thus, Jenkins (1994) criticizes the “almost universal practice to analyse discrimination at the average proportionate wage differential” (p. 82). While the summary scalar measure of discrimination from the B-O decomposition of wage inequality may be the same for two populations, it is possible that the distribution of discrimination across different parts of the wage distribution may be very different for the two. As with the B-O decomposition, the starting point of these decompositions is the creation of a counterfactual. The methods which study wage gaps throughout the wage distribution rely on estimation of counterfactual wage distributions using various methods.

Machado and Mata (2005) put the quantile regression (QR) technique to great use in assessing how different variables affect inequality. For instance, how higher education affects the top of the distribution may be very different from how it affects the bottom and this has implications for influencing overall wage inequality. This is where the QR technique can be far more informative than the Ordinary Least Squares (OLS) regression which only looks at mean effects. The Machado Mata (MM) decomposition splits the changing wage inequalities in Portugal over the period 1986-1995 into two components, one representing the changing characteristics and the other the changing returns to them. Details of the MM estimator are given in section 3.3.1. In this paper I implement the decomposition given in Melly (2006) which is a modification of the MM methodology and addresses some of the issues with the MM (hereafter, MMM). The MMM decomposition is discussed in section 3.3.2. Application of these techniques to the distribution of gender gaps can answer questions about where discrimination was higher in the wage distribution and whether women faced a “glass ceiling”. The early papers focused mostly on European countries, though recently there have been a few studies on Asian economies as well.

Albrecht et al. (2003) was the first paper to introduce a concrete way of defining the glass ceiling phenomenon in wages. They interpret a glass ceiling as an increasing gap in log wages of two groups as one moves from lower to higher quantiles, with a sharp acceleration at the upper tail of the distribution. They found strong evidence of the emergence of a glass ceiling in Sweden, especially in the 1990s. They also found that the glass ceiling effect persists even when controlling for individual differences in characteristics such as age, education, industry/sector of employment and immigrant status. As expected, including detailed occupational controls did lead to decline of the wage gap at the top of the distribution indicating a dampening of the glass ceiling effect. The final section of their paper uses the MM methodology to decompose the log wage gaps at various quantiles into the explained and unexplained components, allowing for different returns to covariates for men and women. The evidence shows that the glass ceiling effect in Sweden is driven by differences in returns to the characteristics (unexplained/discrimination) rather than the endowment differences. de la Rica et al. (2005) considered a similar question for Spain and found that there was a hidden composition effect behind the flat gender wage profile across the distribution. They concluded that the glass ceiling effect was present in the more educated groups, whereas the “glass floor” effect was more apparent for the less educated. This glass floor effect is also referred to as the sticky floor effect in the literature. The sticky floor effect is characterized by larger gender (log) wage gaps at lower quantiles of a distribution and smaller (log) wage gaps at the top of the distribution i.e. a declining wage gap across the distribution. Arulampalam et al. (2007) analyse the question of whether there is a glass ceiling over Europe by studying wage data for 11

countries for the years 1995-2001. They found considerable variations in the raw gender wage gap between European nations. They estimated the quantile regressions separately for three sub-samples (public sector, private sector and both combined) for all countries. Like most studies estimating wage equations, the analysis is carried out including industry and occupation in one model and excluding them in the other specification. Using MM decompositions, they conclude that for all countries and across all quantiles, gender wage gaps exist even when men and women are assumed to have the same characteristics. Some evidence of a sticky floor is found for Italy and Spain, whereas for Austria, Belgium, Britain, Denmark, Finland, France, Germany, Italy and Netherlands, women faced a glass ceiling. Chi and Li (2008) use the RIF decomposition technique to study gender wage differences in urban China from 1987-2004 and find that the sticky floor effect is present in China as opposed to the glass ceiling and that this effect has become stronger over the time period under the study. This effect is mainly responsible for the increasing mean gender wage gaps in China. Both the composition effects as well as the structure effect (unexplained/discrimination component) have increased over time.

There are a few important studies on gender wage discrimination in India. Reilly and Dutta (1996) find that the mean wage differential between men and women was relatively stable in the 1980s and the 1990s. More recently, a study on urban labour markets conducted by Madheswaran and Khasnobis (2007) focused on decomposition of gender wage differentials. The objective of the paper was to study wage determination for regular wage as well as casual workers, and the wage gap between males and females for both these categories of workers. They used three rounds of the EUS, namely the 38th (1983), the 50th (1993-94) and the 55th (1999-00). They implemented several variants of the B-O technique, as well as the Banerjee and Knight (1983) technique to assess job discrimination in the labour market. They found that the gross raw wage differential between men and women declined over time, in both the regular and casual labour markets. Decomposing the gender wage differential, the paper found that both the endowment component and the discrimination component narrowed down for the regular workers over 1983 and 1999-2000, though the decline in endowment difference largely contributed to the decline in the raw wage differentials. The detailed decomposition revealed that differences in education played an important role in explaining the wage differential for regular workers. The casual labour market saw a widening of endowment differences. The discrimination component of the casual labour market however, declined. On the whole, the discrimination against women in the labour market persisted. In addition, for 1999-00, Madheswaran & Khasnobis (2007) also find that for all workers the “explained” component accounts only for about one-fifth of the wage differential, and this proportion is even smaller for regular wage workers, which is a remarkable finding in itself. Quantitative literature on gender wage discrimination in India includes also Duraisamy and Duraisamy (1996) and Kindon and Unni, (1997).

Recent literature on gender wage gaps in India is quite limited. Most papers studying this question rely on the standard mean wage decompositions. My paper addresses the question of whether Indian women face the glass ceiling or sticky floor by analysing the distribution of gender wage gaps across quantiles for different education levels, occupation groups, sectors, age cohorts and caste categories. Quantile regressions will help us understand within and between group gender inequality, as well as how returns to covariates that affect wages, change across quantiles and how these differ by gender. Apart from the standard B-O decomposition, the MMM technique is applied to decompose the wage gaps across different quantiles to ascertain which factor is driving the overall sticky floor (or glass ceiling) effect.

3 Methodology

3.1 The Blinder – Oaxaca decomposition

The wage gap between men and women can exist not only because women are disadvantaged in terms of education, work experience and occupation as compared to men, but also because the returns to these characteristics are different for men and women. The B-O decomposition breaks down the mean gender wage gap into two components: The first is due to differences in covariates and is termed as the “explained” gap (composition effect), and the second is due to differences in returns and is termed as the “unexplained” component (or the wage structure effect). The second component is attributed to discrimination.

The gender wage gap is define as $T = \frac{GM_m - GM_w}{GM_w}$

where GM stands for the geometric mean of wages of the relevant group (m for men and w for women).

Consider the following OLS log wage regression on the pooled sample:

$$\ln(W_i) = \alpha + \beta D_{1i} + u_i \quad (1)$$

where i stands for the individual observation, W refers to wage, u refers to the OLS residuals, α is the intercept term and β is the coefficient of the only covariate in this regression, D_1 , the gender dummy. The estimated coefficient of the gender dummy, $\hat{\beta}$ can be expanded into the following expression,

$$\begin{aligned} \hat{\beta} &= \overline{\ln(W_m)} - \overline{\ln(W_w)} = \frac{1}{n_m} [\sum_{i=1}^{n_m} \ln(W_i)] - \frac{1}{n_w} [\sum_{j=1}^{n_w} \ln(W_j)] \\ &= \frac{1}{n_m} [\ln \prod_{i=1}^{n_m} W_i] - \frac{1}{n_w} [\ln \prod_{j=1}^{n_w} W_j] = \ln(\prod_{i=1}^{n_m} W_i)^{\frac{1}{n_m}} - \ln(\prod_{j=1}^{n_w} W_j)^{\frac{1}{n_w}} \\ &= \ln(GM_m) - \ln(GM_w) = \ln\left(\frac{GM_m}{GM_w}\right) \end{aligned}$$

Therefore,

$$e^{\hat{\beta}} = \frac{GM_m}{GM_w} \text{ and, } T = \left[\frac{GM_m - GM_w}{GM_w} \right] = e^{\hat{\beta}} - 1$$

In fact, when the wage gap T is small (less than 10 percent in magnitude), it can be approximated by $\hat{\beta}$. When wage gaps are larger, it is important to bear in mind the above transformation².

In what follows, I explain the B-O method in brief. The B-O decomposition method is a mean wage decomposition method. It decomposes the difference in the arithmetic mean of a variable (in this case, logarithm of wages) between two groups (in this case, men and women). To understand the method, let us suppose $Se\{m, w\}$ represents the two groups under study here, where m stands for men and w for women. The following semi-log regression equation is run for both groups separately

$$\ln(W_{si}) = \beta_{s0} + \sum_{k=1}^K X_{ski} \beta_{sk} + u_{si} \quad (2)$$

where i stands for individual observation, W for wage, X for covariate (characteristics). There are K covariates such as age, education and caste status and these are indexed by k . β represents the coefficients (returns to covariates). We assume that $E(u_{si}|X_i) = 0$.

² In my paper, given that the wage gaps are large in magnitude, I always apply the transformation to arrive at T .

Given that the residuals from an OLS have zero arithmetic mean:

$$\overline{\ln(W_m)} = \widehat{\beta}_{m0} + \sum_{k=1}^K \overline{X_{mk}} \widehat{\beta}_{mk} \text{ and for women, } \overline{\ln(W_w)} = \widehat{\beta}_{w0} + \sum_{k=1}^K \overline{X_{wk}} \widehat{\beta}_{wk} \quad (3)$$

Therefore,

$$\overline{\ln(W_m)} - \overline{\ln(W_w)} = [\widehat{\beta}_{m0} + \sum_{k=1}^K \overline{X_{mk}} \widehat{\beta}_{mk}] - [\widehat{\beta}_{w0} + \sum_{k=1}^K \overline{X_{wk}} \widehat{\beta}_{wk}] \quad (4)$$

To decompose this total difference, we add and subtract a counterfactual average wage. Let us assume that the true or the non-discriminatory wage structure prevailing in the market is the one faced by men. If women were to be paid according to this wage structure, then the average wage for a woman would be represented by,

$$CF_1 = [\widehat{\beta}_{m0} + \sum_{k=1}^K \overline{X_{wk}} \widehat{\beta}_{mk}] \quad (5)$$

Adding and subtracting CF_1 to equation (4) above, we get,

$$\overline{\ln(W_m)} - \overline{\ln(W_w)} = \underbrace{[\widehat{\beta}_{m0} - \widehat{\beta}_{w0}] + \sum_{k=1}^K \overline{X_{wk}} (\widehat{\beta}_{mk} - \widehat{\beta}_{wk})}_{\text{"D" Unexplained (Discrimination)}} + \underbrace{\sum_{k=1}^K (\overline{X_{mk}} - \overline{X_{wk}}) \widehat{\beta}_{mk}}_{\text{"E" Explained (characteristics)}} \quad (6)$$

The first two terms are that part of the total differential which arise out of the differing returns of men and women to the labour market characteristics that they possess. This difference in coefficients can be thought of as the discrimination component, as it will lead to a wage differential between two groups even if both groups possess exactly the same vector of (average) covariates. The final term represents the difference in average logarithm of wages that is due to different average levels of covariates between men and women. This is the explained component of the average gender log wage gap.

Depending on the assumption about the non-discriminatory wage structure, we could have done the above decomposition differently. If the assumption is wage structure that will prevail in the absence of discrimination is the one faced by women, then the relevant average counterfactual wage would be the wage paid to a person who possesses the characteristics of a representative male worker but is paid according to the wage structure that females face (say, CF_2). This way we get the following alternative decomposition,

$$\overline{\ln(W_m)} - \overline{\ln(W_w)} = \underbrace{[\widehat{\beta}_{m0} - \widehat{\beta}_{w0}] + \sum_{k=1}^K \overline{X_{mk}} (\widehat{\beta}_{mk} - \widehat{\beta}_{wk})}_{\text{"D" Unexplained (Discrimination)}} + \underbrace{\sum_{k=1}^K (\overline{X_{mk}} - \overline{X_{wk}}) \widehat{\beta}_{wk}}_{\text{"E" Explained (characteristics)}} \quad (7)$$

This problem of ending up with different decompositions depending on the assumption made about what is the true/non-discriminatory wage structure is known as the index number problem. The two examples of counterfactuals can be thought of as two extremes, one representing an upper limit and the other a lower limit of discrimination. Other assumptions of non-discriminatory wage structures can be thought of as weighted averages of the two sets of coefficients.³

³ Cotton (1988) used sample proportions of each group as weights, that is, $\widehat{\beta}_C = \rho_M \widehat{\beta}_M + \rho_W \widehat{\beta}_W$, where ρ_M and ρ_W are sample proportions of men and women. The decomposition is done as following:

$$\overline{\ln(W_m)} - \overline{\ln(W_w)} = \underbrace{[\widehat{\beta}_{M0} - \widehat{\beta}_{C0}] + \sum_{k=1}^K \overline{X_{MK}} (\widehat{\beta}_{MK} - \widehat{\beta}_{CK})}_{\text{Male Treatment Advantage}} + \underbrace{[\widehat{\beta}_{C0} - \widehat{\beta}_{W0}] + \sum_{k=1}^K \overline{X_{WK}} (\widehat{\beta}_{CK} - \widehat{\beta}_{WK})}_{\text{Female Treatment Disadvantage}} + \underbrace{\sum_{k=1}^K (\overline{X_{MK}} - \overline{X_{WK}}) \widehat{\beta}_{WK}}_{\text{"E" Explained (characteristics)}}$$

In the results section 5.3, I provide estimates using the two counterfactuals detailed as above and a third structure that uses coefficients from the pooled (men and women combined) OLS regression as the true wage structure. I present the results for the aggregate B-O decomposition in Table 5.

3.2 Quantile regressions

I use quantile regressions extensively in this paper. This section gives a brief outline of the quantile regression methodology. Quantile regressions were introduced by Koenker and Bassett (1978). Just as OLS regressions seek to estimate conditional mean functions, quantile regressions estimate conditional quantile functions. For instance, in this paper I estimate quantiles of the conditional distribution of log wages expressed as a function of covariates such as age, education, caste and job characteristics. The QR offers more flexibility than the OLS in the sense that it allows the effects of a covariate on the dependent variable to change at different points (quantiles) of the distribution. The assumption is that the conditional quantiles of the dependent variable y is linear in covariates X . Therefore, the θ^{th} quantile of the conditional distribution is given by

$$Q_{\theta}(y_i|X_i) = X_i\beta_{\theta}, \quad \theta \in (0,1) \quad (8)$$

For a given θ , the estimate of β_{θ} solves the following minimization problem:

$$\sum_{i=1}^n \rho_{\theta}(y_i - X_i\beta_{\theta}) \quad (9)$$

Where,

$$\rho_{\theta}(u) = \begin{cases} \theta * u, & u \geq 0 \\ (\theta - 1) * u, & u < 0 \end{cases} \quad (10)$$

$\rho_{\theta}(u)$ is also known as the check function. In effect, just as minimizing the sum of squares of residuals in an OLS regression with just the intercept gives the sample mean, minimizing the sum of absolute residuals yields the median. Estimating a conditional median function therefore means minimizing equation (9) with $\theta=0.5$. For other quantiles, the problem boils down to minimizing the sum of asymmetrically weighted absolute residuals, where the weights are given by the function $\rho_{\theta}(u)$. The quantile regression allows us to estimate the marginal effect of various personal and job characteristics which affect wages at various deciles (see section 5.4 for QR results).

3.3 The Quantile regression decomposition methods

3.3.1 The Machado Mata decomposition

The B-O decomposition method outlined in section 3.1 decomposed the average gender log wage gap into a “discrimination” and an “explained” component. Just as the B-O decomposition relied on the construction of a counterfactual average log wage (CF_1 or CF_2), in the MM decomposition, in order to decompose the difference between two distributions we need to construct an entire counterfactual distribution. Essentially, the MM decomposition is a generalization of the B-O mean decomposition to decomposing quantiles. Gender wage gaps at each quantile, which represent the differences between the male log wage and the female log wage distribution at that quantile, can then be decomposed using a counterfactual distribution. This counterfactual distribution could be constructed using the covariates of men and the returns to characteristics for women, or vice-versa (the familiar index-number problem).

As outlined in the literature review, there are several methods of estimating counterfactual distributions. Our focus is on the MM methodology that relies on quantile regressions. We first outline the MM

methodology and then focus on Melly's modification of the same (MMM). The MMM decomposition will then be applied in 6.7 to obtain my main result.

The MM decomposition estimates the entire conditional distribution using quantile regressions. This is a parametric approach as it assumes the linearity of the conditional quantile in covariates (as in equation (8)). The four steps of the MM procedure to generate a counterfactual log wage distribution are:

- 1) A random sample of size n is generated from a uniform distribution $U[0,1]: u_1, u_2, \dots, u_n$
- 2) For men and women separately, estimate n QRs using the draw values as the quantile at which to estimate the QRs. Thus, we get $\{\hat{\beta}_{u_j}^M\}_{j=1}^n$ and $\{\hat{\beta}_{u_j}^W\}_{j=1}^n$. In other words, we end up with n coefficient vectors each, for men and women.
- 3) Take a random sample with replacement of size n from the covariate distribution of men and women separately. Let this random sample be denoted by $\{\tilde{X}_j^M\}_{j=1}^n$ and $\{\tilde{X}_j^W\}_{j=1}^n$
- 4) The counterfactual distributions are estimated as $\{Y_j^{cf} = \tilde{X}_j^M \hat{\beta}_{u_j}^W\}$ or $\{Y_j^{cf} = \tilde{X}_j^W \hat{\beta}_{u_j}^M\}$ for $j = 1, 2, \dots, n$

The first counterfactual distribution represents a distribution of log wages of men being paid according to the female wage structure, as it uses the covariates of men and returns of women. The second counterfactual represents the case where women retain their characteristics and get "paid like men".

3.3.2 The Machado-Mata-Melly (MMM) decomposition

The MMM decomposition is very similar to the Machado and Mata (2005) decomposition and the principles behind the two decompositions are identical. In the MMM procedure, parametric linear conditional quantile regressions are estimated in exactly the same manner as outlined in steps one and two of the MM procedure, to obtain n sets of coefficients at various quantiles, for men and women separately. In my paper I estimate the conditional distribution using 100 quantile regressions (i.e. $n=100$). Increasing the number of estimated QRs increases precision of the estimated conditional distribution but this comes at the cost of very large computation time. Thereafter, unlike the MM procedure that relies on a random draw of n vectors from the distribution of covariates, the MMM uses all observations on covariates and combines with each observation the n quantile regression coefficients to generate the unconditional (marginal) distribution of log wages. Thus, the conditional distribution is integrated over the entire range of covariates to obtain the unconditional distribution. Estimating the unconditional distribution this way has the advantage of using all the information contained in the regressors and this makes the MMM estimator more efficient than the MM estimator. Also, since MMM does not rely on simulations to obtain covariate vectors, it is computationally less intensive and as a consequence faster to implement.

The MM is numerically equivalent to the MMM decomposition when the number of simulations (n) used in the MM decomposition goes to infinity. Using 5000 Monte Carlo replications, Melly (2006) demonstrates that the mean square error of the MM is greater than that of the MMM procedure, and the MM mean square error converges to the MMM mean square error only when the number of simulations in MM become very large. Melly also proves in his paper that his estimator is consistent and asymptotically normally distributed.

At the θ^{th} quantile, the difference between the estimated unconditional quantile of log wage for men, $\hat{Q}_m(\theta)$, and the estimated unconditional quantile of log wage for women, $\hat{Q}_w(\theta)$, can be decomposed as,

$$\hat{Q}_m(\theta) - \hat{Q}_w(\theta) = \underbrace{[\hat{Q}_m(\theta) - \hat{Q}_{cf}(\theta)]}_{\text{Effects of Coefficients}} - \underbrace{[\hat{Q}_{cf}(\theta) - \hat{Q}_w(\theta)]}_{\text{Effects of characteristics}} \quad (11)$$

where $\hat{Q}_{cf}(\theta)$ is the estimated counterfactual unconditional quantile of log wage for men created using the coefficients of women. Results based on Equation (11) are presented in Table 8 and Figure 13.

4 The data

For the estimation, this paper uses the “thick” (large sample) round of the NSS Employment-Unemployment Survey (EUS) for the year 2009-2010 (66th round). I also briefly examine the data for the year 1999-2000 and 2004-2005 (the 55th and 61st rounds, respectively). The EUS provides wage information for casual labourers and regular wage workers (it does not include earnings of the self-employed). In this paper, like most other studies estimating wage equations using EUS data, the sample is restricted to the regular wage (or salaried) workers and excludes casual labourers. The regular workers roughly correspond to those working in the formal sector. NSS defines these workers as those who worked in others’ farm or non-farm enterprises (both household and non-household) and received salary or wages on a regular basis (as opposed to the daily or periodic renewal of work contract). One reason I exclude casual labourers is that covariates like education and experience do not explain the variation in wage rates of casual workers in the same manner as they do for regular wage workers. The other reason is that I am interested to know whether gender discrimination is practiced in the formal labour market which is supposed to be governed by labour regulations which prohibit unequal pay for equal work. It is sometimes argued that the women are under-represented within regular wage workers as most of them work as casual labourers. In the dataset which I use, about 20 percent of the regular wage workers are women. For casual labourers, however, this ratio is not very different (75 percent are men and 25 percent are women)⁴. The EUS asks questions on weekly wages, which I use to estimate the daily wage rate using number of days worked in the survey week. Throughout this paper the raw wages refer to nominal daily wage rates measured in rupees. The wage distribution is trimmed by 0.1 percent from the top and the bottom, in order to get rid of outliers and potentially anomalous wages at the extreme ends of the distribution. The focus of the paper is on full time workers with ages between 15 and 59⁵. Making all these restrictions I lose around 4 percent of the observations. The sample size I am left with is 35,363 observations which represents 59 million workers in the population working in the regular wage market, comprising of 48.6 million men and 10.4 million women.

5 Results

5.1 The sticky floor effect

Figure 1 shows the log wage gaps at each quantile (in this case, each percentile), calculated as the difference between the log wage of men and log wage of women at that particular quantile. For instance, the log daily wage at the tenth percentile of the male wage distribution is 4.34. In the female wage distribution, the log daily wage at the tenth percentile is 3.6. Their difference, 0.74, is therefore, the gender log wage gap at the tenth percentile⁶. Similar calculations are carried out for each percentile. At

⁴ I do not correct for sample selection into employment when estimating the wage equations.

⁵ The minimum legal age for working is 15. At the age of 60 we notice a sharp drop in average wages because the retirement age for regular wage workers is usually 60. Evidently, most people do not continue working after this age or switch jobs and move to something more informal

⁶ In a QR of log wage on gender alone, let $\hat{\beta}_0$ represent the QR coefficient of the gender dummy.

each percentile the (log) wages of men exceeded that of women, i.e., the (log) wage gap is positive throughout the distribution.

An important point to note is that these gaps are also the quantile regression coefficients of the gender dummy in a log wage regression evaluated at different quantiles without any controls. For example, the median gap of 0.43 is also the coefficient of the gender dummy of the median regression. The advantage of calculating the log wage gaps using quantile regressions is that we can obtain standard errors for these gaps and are therefore able to check their significance. The average log wage gap, as represented by the coefficient of the gender dummy in an ordinary least squares regression of log wages on a gender dummy (without including other controls), is represented by the solid horizontal line. The value of this average gender log wage gap is 0.4 and corresponds to a 50 percent raw wage differential where the latter is the difference in the geometric means of the raw wages of men and women expressed as a percentage of the geometric mean of women's wages⁷. The dotted horizontal lines represent the 95 percent confidence interval of the OLS coefficient.

As is evident from Figure 1, gender log wage gaps in India in the salaried sector are characterized by a “sticky floor” effect. This means that gaps are much higher at the lower percentiles of the log wage distribution and decrease as we move to higher percentiles. To put it simply, women who are low wage earners face a greater gender gap as compared to women at the higher end of the wage distribution. While OLS estimates reveal an average log wage gap of 0.4, the log wage gap is as high as 0.79 at the 7th percentile and it is as low as 0.04 at the 84th percentile. Only around the middle of the distribution between the 40th and the 60th percentiles do we find gender log wage gaps that are not significantly different from the average gap - all other gender log wage gaps lie beyond the confidence interval of the OLS coefficient. Thus, the average gap is hardly representative of the gaps at different points of the distribution. To get a better idea of how large these gaps are at the bottom end, note that a log wage gap of about 0.79 at the 7th percentile corresponds to a raw wage gap of 120 percent, while the average (OLS) gap of 0.4 corresponds to a raw wage gap of 50 percent (see Footnote 6 and 7). This means that men get paid more than twice as much as women at the bottom of the distribution. At the 84th percentile, the gap is as small as 4 percent. While for most of the distribution the gap is highly statistically significant, at some of the top percentiles, this gap becomes statistically insignificant. Whether this sticky floor effect is present across time, and whether it persists within occupations, education levels, industries, age cohorts and sectors (urban/rural) will be examined in the following sections.

5.1.1 The sticky floor effect across time

The average wages in Table 1 is the arithmetic mean of the raw wages, where raw wages refer to wages expressed in Rs. per day, without the log transformation. The average wage differential in Table 1 is the difference of the arithmetic means of raw wages of men and women expressed as a percentage of women's mean wage. I follow this convention in Table 2 as well. The figures, however, always depict log wage differences.

For each quantile θ , $\hat{\beta}_\theta = \ln w_{m\theta} - \ln w_{w\theta} = \ln\left(\frac{w_{m\theta}}{w_{w\theta}}\right)$. And therefore, $[e^{\hat{\beta}_\theta} - 1] * 100 = \left[\frac{w_{m\theta} - w_{w\theta}}{w_{w\theta}}\right] * 100$. Given that a log transformation is monotonic, $\log w_{m\theta} = [\log w_m]_\theta$, that is, taking the log of the θ^{th} quantile of the raw wage distribution is equivalent to the θ^{th} quantile of the log wage distribution. Hence a difference in the gender log wage gap of 0.74 corresponds to a raw wage gap differential of 107 percent at the tenth percentile.

⁷ The values 0.4 and 50 % correspond to $\hat{\beta}$ and to $T * 100$, respectively (where $\hat{\beta}$ and T are defined as in Section 3.1)

Table 1 shows that average gender wage differential worsened from 30 percent in 1999-2000 to about 34 percent in 2004-2005, followed by a sharp fall to 25 percent in 2009-2010. While average wage differentials are typically reported as summary measures, I have demonstrated above that average log wage gaps are not representative of log wage gaps at various quantiles of the distribution. Figure 2 shows the evolution of the gender log wage gap across quantiles over the ten year period from 1999-2000 to 2009-2010. The solid dark line corresponds to 2009-2010, the dark dotted line to 2004-2005 and the light double line to 1999-2000. The sticky floor effect is present for all the three years plotted in the figure. As is evident from the figure, the distribution of the log wage gaps is quite similar for the years 1999-2000 and 2004-2005 throughout the distribution, except around the sixth and seventh deciles, where gaps are higher for 2004-2005. This contributed to the widening of the average log wage gap between these two years. The dramatic fall in average gender wage differentials from 2004-2005 and 2009-2010 (as seen in Table 1) has happened mostly due to the gains by women vis-à-vis men in the middle of the distribution. Figure 2 shows that the log wage gap profiles for these two years is very similar at the extremes of the distribution, but in the middle the gaps are lower for 2009-2010 compared to 2004-2005. Thus, among the most highly paid and the very poorly paid, gender log wage gaps are virtually unchanged over the decade. The persistence of the sticky floor over the decade implies that further reductions in average gender log wage gaps could be brought about by focusing on men and women in the bottom half of the distribution.

5.1.2 Gender wage gaps in the rural and urban sectors

Table 2 provides the summary statistics for several variables that enter into the regression models we use later in the paper. The first panel of the table provides the distribution (shares) of men, women and the pooled sample over these categorical variables. The second panel provides average daily wages and the gender wage gaps as defined above. In this section, I focus on the gender wage gaps separately for the rural and urban sectors to see if there is any evidence of the sticky floor in these sectors. Women working in the regular wage urban sector in several developed or more urbanized nations experience the glass ceiling phenomenon, and it would be interesting to study whether this is the case for the urban sector in India. As expected, urban workers constitute 65 percent of all regular wage salaried workers. A test of differences in proportions reveals that the rural-urban distribution is not significantly different for men and women.

The overall average gender wage differential as given in Table 1 is about 25 percent in 2009-2010. However, this varies greatly between the rural and the urban sectors. The average rural gender wage differential is 54 percent, whereas the corresponding figure for the urban sector is about 15 percent. The fact that the average gender wage differential is much higher in rural areas than urban areas, and that rural wages are generally at the lower end of the overall wage distribution, fits in with the overall picture of the declining log wage gaps across the log wage distribution (Figure 1). But does the sticky floor phenomenon for the entire population also prevail within the urban and rural sectors? It is evident from Figure 3 that the sticky floor phenomenon is seen in both the rural and the urban sectors. As noted previously, the average gender wage differential within the rural sector was much greater than that of the urban sector, and Figure 3 confirms this. The rural gender log wage gap is higher than the urban gap at all percentiles. While in both sectors, higher gaps are seen at the bottom, the urban gaps are almost uniformly declining, whereas the rural gaps remain very high until the 8th decile. As far as the urban log wage gaps are concerned, log wage gaps fall close zero and are statistically insignificant (at the 1 percent level) near the 8th decile.

5.1.3 Gender wage gaps in the public and private sectors

We observe from Table 2 that about 33 percent of the regular wage workers work in the government or the public sector. The adjusted Wald test shows that the proportion working in the public sector is significantly different for men and women at the 1 percent level, 32 percent for men versus 40 percent for women. Table 2 also shows that average wage is much higher in the public sector than in the private sector. The average gender wage differential is higher in the public sector than in the private sector: It is about 36 percent in the public sector, while it is about 26 percent in the private sector.

To see the distribution of log wage gaps across percentiles, we see in Figure 4 that the higher average gender log wage gap in the public sector stems from the higher gaps at the bottom of the distribution. Log wage gaps are very high at the lower quantiles (at times over 1.2, for example at the second decile)⁸. Once we go beyond the median however, the gender gaps in the public sector are lower than those in the private sector, except at the very top. Though women in both sectors face a sticky floor effect, this effect is much stronger in the public sector where the drop across quantiles is sharp. Public sector wages are more regulated with fixed salary structures and seeing larger wage gaps at the bottom of the distribution as compared to the private sector is somewhat surprising: This could be an indication of the differences in the nature of jobs in terms of industry and occupation that men and women are employed in. The private sector log wage gap profile is also declining, but much flatter.

5.1.4 Gender wage gaps for different levels of education

In this section, I examine the wages of men and women according to different education levels. de la Rica et al. (2005) study the distribution of gender log wage gaps in Spain and find that the relatively flat profile of these gaps across percentiles contains a composition effect. The glass ceiling effect was prevalent for the groups with higher education, while the sticky floor effect was found in the groups with lower education levels. For my data, I have divided education into 7 categories as given in Table 2. The table shows that there are differences in education levels by gender. While illiterates comprise about 8 percent of the sample, the share of illiterates among women is much higher (15 percent) than that for men (6 percent). As we move up the education scale, we see that for all levels of schooling, the proportions are larger for men than for women. At the very top of the education spectrum (university education), however, we see that almost 30 percent of all women are graduates or have diplomas, whereas the corresponding share for men is 26 percent. The proportion of women who are postgraduates is over 12 percent, but lower for men (about 8 percent). Thus at the two extremes of the education spectrum, the illiterates and the university educated, the proportions are higher for women than men. Except primary and below, for all other six education categories in Table 2, the difference in these proportions of men and women belonging to them is significantly different at the 1 percent level.

As expected, as we move up the education scale, the average daily wages increase, for both men and women. To see how disparate average wages are for men and women, note that on average illiterate men actually get paid more than women who have completed middle school. This difference is statistically significant at the 1 percent level of significance. Average gender wage differentials are much higher for lower levels of education, and are smallest for the university educated. Combined with the fact that on average the lower educated are usually at the bottom of the overall wage distribution, and the higher educated at the top, this would fit in with the observation of the sticky floor phenomenon in Figure 2.

⁸ This corresponds to a raw wage differential (in terms of GMs) of over 230 percent.

Once we go beyond the average wage differentials and examine the distribution of log wage gaps for all these categories, we can answer the question of whether the sticky floor is only a local phenomenon for the less educated groups. The argument in de la Rica et al. (2005) hinged on differential labour market participation rates for men and women. In Spain, the labour market participation rates were found to be much lower for women than men among the less educated, while they were almost equal for those with higher education. These differences were attributed as possible explanations for the sticky floor being found the less educated and the glass ceiling among the higher educated⁹. The Indian regular wage market, however, employs a larger share of men than that of women, across educational categories, and this could be indicative of differential labour market participation rates of men and women. We would therefore, expect a sticky floor effect to be present for both these categories. I focus on the gender log wage gap across the distribution for illiterates and postgraduates as extremes of low and high education, respectively. There are two interesting points to note about Figure 5. First, and most important, we see a sticky floor effect for both categories. This is not only true for the two categories presented here, but also true for other intermediate levels of education (graphs not presented here). Nowhere do we find evidence of a glass ceiling for women. Secondly, while the log wage gaps are declining for illiterates, even at 9th decile these gaps are very high as compared to the gaps for post graduates. At the 9th decile, the gender log wage gap for postgraduates is only 0.06. Thus, even amongst the very highly educated, it is only among those earning high wages, that we see some equalization of wages between men and women.

5.1.5 Gender wage gaps across occupations

This paper uses one-digit codes according to the national classification of occupations (NCO) -2004 to create broad occupation categories. Overall, only about 0.2 of the observations are unclassified or missing and no separate category is created for such observations. Table 2 shows that the more predominant occupations for women are Professionals, Associate Professionals and Elementary occupations. Only a small percentage of women hold Managerial jobs. The three dominant occupation categories for men are the Crafts and Related Trades, Plant and Machine Operators and Services/Sales workers. Evidently, the occupational distribution of men and women is very different. Tests of differences in proportion reveal significant differences between men and women for every broad occupational category as defined here, except for Clerical workers. The differences in the wages between men and women are positive and significant for all broad occupation categories except for Managers and Legislators, where the difference turns out to be insignificant. Even though not many among women are employed in Managerial jobs, on average they get paid more than men working in similar occupations. The largest average wage gaps are found in Crafts and Related Trades, and Elementary occupations. Both these categories are low-wage occupations in general, and yet again, these results are consistent with the overall sticky floor of Figure 2.

Next, I examine whether the sticky floor effect is present within all occupations, or only in some high/low wage occupations. Figures for gender log wage gaps across quantiles for these categories separately are not presented here. For Professionals, Technicians and Associate Professionals, Clerical, Service and Sales workers, as well as those engaged in Elementary occupations; there is a clear sticky floor effect. For Plant and Machine Operators, and for those engaged as Crafts and Related Trades workers, gender log wage gaps do not exhibit any particular pattern and are mostly constant throughout the wage distribution.

⁹ Details of the link between differential labour market participation rates for men and women, and the existence of a sticky floor is discussed in Section 0

5.1.6 Gender wage gaps for different age cohorts

Age will serve as a proxy for years of experience in the log wage regressions I estimate later in the paper. This is done because the EUS does not provide a direct measure of work experience in terms of years employed. Some other papers calculate experience using a rough rule such as ‘Age – Years of education – 6’. I do not follow this approach as years of education will have to be estimated as well, since education is a categorical variable in my dataset. To avoid measurement errors arising from the estimation of years of education, many studies using the NSS-EUS (including mine) use age as a proxy for experience, especially since experience (as calculated using the rule stated above) is a linear monotonic transformation of age. The average age of men in the sample is 35.7 years while that of women is 34.72. This difference in mean age by gender is significant at the 1 percent level.

A composition effect could be one explanation for the overall sticky floor effect we see in Figure 1. Those belonging to younger cohorts, being more inexperienced are usually found at the lower end of the overall distribution. Since wages increase with tenure, the older, more experienced workers are more concentrated at the higher end of the wage distribution. The overall sticky floor phenomenon could be characterized by a composition effect if the youngest cohorts experience large gender wage gaps and the older cohorts a smaller one. To check if there is an age composition effect that drives the overall sticky floor, this paper examines the log wage gap for the distribution separately for three age cohorts, 15-34, 34- 50 and 50-59. As can be seen from Figure 6, there exists the sticky floor for all three age cohorts. It is interesting to note that for the youngest age cohort, aged 15-34, gender log wage gaps almost disappear in the top half of the distribution. On average, as compared to the youngest cohort, the wage gaps of older cohorts are higher, especially at the bottom of the distribution. Therefore, I do not find evidence of a composition effect that can explain the overall sticky floor. One reason I observe higher wage gaps for older cohorts could be because education profiles of men and women differ by age cohort. Gender differences in education levels could be most unfavourable for women who belong to the oldest cohort. For the youngest cohort on the other hand, gaps in education may be much narrower. Also, for older cohorts, there could be a larger difference in the tenure of men and women, since women are more likely to drop out of the labour force and re-enter at a later stage.

5.1.7 Gender wage gaps across industries

The NSS provides data on the industry in which a worker is employed using the National Industrial Classification (NIC) 2004 system. A five-digit industry category is given for each worker. This is extremely narrow and disaggregated to begin with. At the broader levels there are 62 divisions (two digit classification) that are clubbed into 17 sections according to NIC-2004. I further club these 17 sections into 11 broad categories presented in Table 2, which also presents summary statistics for the same. The differences in the industrial distribution for men and women are significant for all broad industry categories except Agriculture and Real estate/Renting (this category includes business activities).

On average, Financial Intermediation is the most lucrative industry to work in, with the highest average wages, both for men and women. Gender gaps in this industry are only about 8 percent, whereas for low paying industries such as Manufacturing wage gaps are close to 60 percent. Even though a very small proportion of women actually work in Electricity, Gas, Water supply and Construction, their average wage is greater than that of men. The same holds true for Transport/Communications /Storage industries. In the interest of brevity, I do not present the graphs depicting the gender log wage gaps over the wage distribution by each industry category. For the biggest employer overall, that is, Manufacturing industries

we see a relatively flat profile for gender log wage gaps across quantiles. For other industries with large employment shares such as Education and Public administration/Defence, we see a sticky floor effect. I do not find evidence of the glass ceiling phenomenon in any of the broad industry groups.

5.1.8 Gender wage gaps and caste

Table 2 shows that among the regular wage salaried workers, a greater percentage of men in the sample belong to the category of “Others” and Other Backward Castes (OBCs), whereas a greater percentage of women belong to Scheduled Castes (SCs) and Scheduled Tribes (STs). The gender difference between the proportions is significant at the 1 percent level for SCs and at the 10 percent level for Others. Thus, in the high wage social group of Others, we find a significant difference between the proportion of men and women in favour of men, whereas for the Scheduled Castes, the social group with the lowest average wages we find the reverse is true. Note that the average gender wage differential is greatest for the SCs at 66 percent, and it is less than 9 percent for Others. At the bottom of the distribution, therefore SCs, STs and OBCs face higher gender wage gaps, whereas at the top end of the distribution, dominated by Others category, gender wage gaps are lower. This fits in with the observation of the overall sticky floor in Figure 2. I find that the sticky floor phenomenon is found for all social group categories (graphs not presented here). The fall in the log wage gap is the sharper for the general category (Others) than SCs, STs or OBCs. This steep fall means that the gender gap falls to zero just above the median for Others. The implication is that within Others women in the top half of the distribution face no significant gender log wage gaps.

5.1.9 Gender wage gaps & job characteristics: unionization, contracts and permanent jobs

First we consider union membership. Note that on average about 69 percent of the regular wage population do not belong to any union (Table 2). There is no significant difference in the level of unionization between men and women. On average unionized workers earn far more than those who do not have union membership. And further, average gender wage differentials are far higher amongst those that do not belong to any union as compared to those who do. This fits the overall picture of the sticky floor. The stylized fact that unionized wages are less dispersed in general, and that gender log wage gaps among unionized workers are smaller as compared to non-unionized wages, will be confirmed later in the paper using quantile regressions. As far as share of workers in a permanent job is concerned, there is no significant difference between men and women. Given that the average gender wage differential for temporary workers is larger than for that for workers with permanent jobs, and that the latter are better paid than the former, this also fits well with the evidence of an overall sticky floor as observed in Figure 1. The employment unemployment schedule also provides information on written job contracts¹⁰. Almost 64 percent of men and 60 percent of women do not have any written job contract. We find that this difference is significant, and is a characteristic that could benefit women’s wages relative to men since workers with a contract, on average, earn more than those without one. The average gender wage differential is significantly higher within contract workers as compared to those without a contract, at the one percent level. Whether job contracts are associated with an increase in gender log wage gaps, *ceteris paribus*, is a question I examine in a later section in the paper.

¹⁰ The correlation coefficient between having a permanent job and a written job contract is 0.3. There seems to be no major overlap between these two categories, and we include them both in our wage regressions later in the paper.

5.2 The ordinary least squares log wage regression

An OLS regression is estimated on the pooled sample (using observations for both men and women). The semi-log functional form is used, (similar to Equation (1)), where the dependent variable is the natural log of (nominal) daily wage and independent variables include a set of covariates along with a gender dummy. The coefficient of the gender dummy in the pooled sample will tell us how the average gender log wage gap changes once we account for differences in various characteristics that affect wage. Note that in the pooled specification, I am constraining the returns to the labour market characteristics to be equal for men and women. Running the estimation separately for men and women (of course without the gender dummy), will give us an idea of how the returns to these characteristics differ by gender. Note that these regressions estimate reduced form equations, in the sense that they are an outcome of the forces of labour demand and supply. These refer to the standard mincerian wage regressions. I estimate two specifications for the log wage regression equation. The details of these two specifications and their results are provided in the next two sections.

5.2.1 The personal characteristics (PC) regression

The first specification is the personal characteristics regression and includes covariates for age, age squared (divided by a 100), the dummies for education, the marital status dummy, caste dummies, sector dummy (urban versus rural), dummies for different states, and of course the gender dummy. Therefore, the only covariates I include in this specification are those that pertain to the personal characteristics of the individual, as opposed to the characteristics of the job that he or she is currently working in. The independent variables in the PC specification can be thought of as exogenous to labour market outcomes, in the sense that most of them are determined at birth, and others such as education are chosen before the individual enters the labour market (expectations of labour market outcomes may affect individual's decision to accumulate human capital, but in this paper I treat education also as an exogenous variable). As in most other wage regressions, the quadratic term of age is included, since economic theory suggests that the wage-age relationship is concave. As mentioned earlier, I have created seven broad categories for education and the omitted category is that of Illiterates. In the case of marital status, the omitted category represents those Never Married. Among caste categories, Others form the omitted group. The reference group for the sector dummy is the rural sector and for state dummies is Maharashtra (the state with the largest population in our sample). The gender dummy takes a value of 1 for males, and therefore a positive coefficient on the gender dummy would be an indication of gender wage gap where men are paid more than women. Results from the PC regression are given in Table 3.

The pooled regression is based on an estimation sample of 35,337 observations and has an R-square of 0.48. The signs of all the coefficients are as expected. The coefficient for age is positive and significant, while the coefficient of the quadratic in age term is negative, but not significant. One additional year of age (experience) increases wages by about 2.5 percent, if evaluated at the mean age of 35.5 years. On average, wages in urban areas are 27 percent higher than those in rural areas. The coefficient of the marital status dummy is positive and significant implying that those married earn 14 percent more than those not currently married. Relative to the omitted category of illiterates, each of the education level dummies is positive and highly significant, and as expected, returns to education increase with the level of education. As an example, note that the coefficient of the higher secondary dummy, 0.78, means that on average, everything else being equal, completing higher secondary education is associated with a 118 percent increase in wages compared to wages of the illiterate group. The caste dummies in Table 3

indicate that as compared to the omitted Other category (social group), SCs STs and OBCs earn a significantly lower wage.

The most important coefficient to examine is the one on the gender dummy. Without any controls the coefficient of the gender dummy is 0.402. In the PC regression we find that this coefficient is virtually unchanged (0.399¹¹). This is the gender gap that remains after controlling for individual differences in covariates mentioned earlier, while keeping the returns to all these covariates same for men and women. The gender log wage gap of 0.399 that remains is a measure of unequal pay between men and women in spite of identical personal characteristics in terms of age, education, caste status, marital status, education, state and sector of residence. Given the discussion of the summary statistics in earlier sections, the wage gap not changing by much after controlling for characteristics is not a surprising result. As seen there, as far as personal characteristics are concerned, women are not at a distinct disadvantage vis-à-vis men. For instance, considering education levels, while the proportion of women who were illiterate was greater than that of men, this was also true of graduates and post-graduates. The rural-urban spread was also not very different by gender. These effects, therefore, compensated for the fact that women in our sample were, on average, younger than men and had a more unfavourable caste distribution. The pooled regression makes the assumption, however, that the earnings functions of men and women differ only with respect to the intercept term. By studying the regression results separately for men and women in the next section, we will see how valid this assumption is. The B-O decomposition (section 5.3) will also take these possible differences in returns to characteristics into account.

The second and third panel of Table 3 presents the regressions results for men and women separately. Firstly, for men the coefficient for age squared is negative and significant, whereas for women it is not significant. For men, therefore, wage profiles are concave with respect to age, while for women the curvature is much smaller. The gain in wage due to living in urban areas is greater for women, whereas return to being married is greater for men. All the education dummies are significant for both genders, but the magnitude of all the coefficients are larger for women. Within women there are no significant differences between Other category and SCs/STs as far as wages are concerned. OBC women, however, earn significantly less than women belonging to the Other category. On the other hand for men SCs, STs and OBCs earn significantly less than those who belong to the Others category, with the OBCs faring the worst.

5.2.2 The personal and job characteristics (PJC) regression

In my second specification, along with personal characteristics I also include job characteristics. The two main job characteristics I control for are industry and occupation (see Table 2 for various categories). The omitted category chosen for occupations are Elementary occupations whereas for industries, Mining and Manufacturing is the reference category. Other job characteristics I control for are union membership, public sector, job contract and permanent job dummies with the reference categories being no union membership, private sector, no written job contract and temporary job, respectively. Here, the coefficient on the gender dummy represents the average gender log wage gap “within” an occupation, industry and nature of job strata. As discussed earlier, the occupational and industrial distributions are quite different for men and women (Table 2). Some other job characteristics (such as job contracts, public/private sector) also differ for men and women. The coefficient on the gender dummy in the PC regression also captured the differences in wages between men and women attributable to their entering into different occupations

¹¹ This corresponds to a raw wage gap of 49 percent i.e. the “T” in section 4.1.1 is 0.49

and industries. In the PJC regression, however, having controlled for broad industry and occupation groups one eliminates part of the average log wage gap arising from these sources. Thus, the remaining gender log wage gap in the PJC regression reflects discrimination that corresponds to the concept of unequal pay between men and women for roughly equal (or similar) work. In the PC specification, on the other hand, the gender log wage gap represented unequal pay for equal personal characteristics (and unequal work). Since we are controlling for 1 digit NCO occupations and broad industry categories, however, the PJC does reflect job and pay discrimination within each of these categories.

The results for the PJC are given in Table 4. The focus is on the differences between the PC regression and the PJC regression. One notable difference is that the coefficient for the marital status dummy is insignificant for women in the full scale (PJC) regression. Also the PJC reveals that SC women are paid significantly lower than Other category women, while OBC women are not underpaid significantly. This is in contrast to the PC regression where the reverse was true. The coefficients of the education dummies fall in the PJC as compared to the PC regression. This indicates that attaining greater levels of education also leads to upward mobility in terms of working in higher paying industries and occupations. In the PC regression, the returns to education were larger because they included the positive returns to education in the form of better access to jobs in high paying industries and occupations. In the PJC regression, however, the returns denote the benefits of attaining higher education *within* an occupation, industry and job type. As before, the main coefficient of interest is the coefficient of the gender dummy, 0.368. This is the gender gap that remains after controlling for all the characteristics that we have available in the dataset that may affect wage, including job characteristics. Once again, the returns to these characteristics are constrained to be equal for both men and women in the regression using the pooled sample. As compared to the PC regression, gender dummy coefficient is now lower. The fact that controlling for personal characteristics alone did not reduce the log wage gap by much (0.402 to 0.399), and controlling for job characteristics reduces it somewhat (0.391 to 0.368) has some implications¹². This means that personal characteristics of women do not place them at a great disadvantage vis-à-vis men, whereas the types of jobs they enter do contribute somewhat more to the gender log wage gaps that we see. Therefore the industry and occupations into which women typically enter, along with the nature of the job has some role in explaining gender differences in wages. However, even after controlling for all these characteristics, 94 percent¹³ of the original gap of seen in the PJC regression still persists. This is indicative of high levels of gender discrimination in the regular wage labour market as this gap exists even after comparing women and men who do broadly similar work.

For men, being educated to the primary level or below does not yield higher wages than those who are illiterate. Among the general job characteristics, note that unionization, on average, benefits women more than men, as does having a permanent job. Relative to non-unionized workers, union membership increases wages of men by 20 percent but increases wages of women by 46 percent. Belonging to the public sector, benefits both men and women, but benefits men more. The coefficients for all the occupation dummies are positive for the pooled sample as well as for men. These coefficients should be interpreted as being relative to elementary occupations. For men, all occupations pay significantly more

¹² The initial “uncontrolled” average log wage gap is 0.402 in the PC regression and 0.391 in the PJC regression. This difference simply reflects the difference in size of estimation samples of the two regressions, with the PJC regression run on a marginally smaller sample size having lost some observations on account of inclusion of several additional variables with some observations missing.

¹³ $(0.368/0.391)*100$

than Elementary occupations, whereas for women this holds true only for Managers/Legislators, Professionals and crafts and related trades workers.

5.3 The Blinder-Oaxaca decomposition results

The first panel of Table 5 presents the B-O decomposition results for the PC regression while the second presents the results for the full scale PJC regression. The first column in both panels refers to the case which assumes that the non-discriminatory wage structure is the one faced by men (Equation (6)). Column two refers to the case where the coefficients of women are used, as in Equation (7). Column three, the pooled wage structure, uses the coefficients of the pooled sample (including a gender dummy as control). The first row of the tables gives the average gender log wage gap as given by the coefficient of the gender dummy in a log wage regression (pooled regression) without any controls. This is the overall difference between the average log wage of men and average log wage of women (Term “ $\overline{\log(w_m)} - \overline{\log(w_w)}$ ” of Equation (4)) and is the value to be decomposed. The second row gives us the explained component (“E”) of the gap, while the third row gives the discrimination component (“D”). The fourth row gives the percentage of the average gender log wage gap that is unexplained (discrimination). The overall gap to be decomposed in the PC regression is 0.4 which corresponds to a 50 percent wage differential between men and women. Irrespective of the choice of the non-discriminatory wage structure, the explained component is very close to zero, and is insignificant. The discrimination component is almost entirely responsible for the overall log wage gap between men and women. In the case where the male wage structure is assumed as the non-discriminatory one, the percentage of the gap due to discrimination is over a hundred percent. This means that in the absence of discrimination women would actually earn wages that are higher than that of men (though negligibly and insignificantly so). Thus, the personal characteristics of women are not altogether unfavourable as compared to men, but it is the returns to these characteristics that lead to the yawning gender gap in wages. One could conclude, therefore, that at least within the class of regular wage workers, pre-market discrimination is not an important factor that leads to log wage gaps.

Note that for the PC regression the unexplained gap in Table 5 assuming pooled coefficients is 0.399. This is the same as the coefficient of the gender dummy we encountered in the first panel of Table 3. This decomposition assumes that the non-discriminatory wage structure is based on coefficients of the pooled sample. This assumption is equivalent to the restriction implicit in the results of Table 3 which was that the returns paid to men and women are equal to the coefficients on the pooled sample (and hence equal to each other). The coefficient of the gender dummy, therefore, represented the gender gap that remained “unexplained” after controlling for personal characteristics, which is exactly what “D” represents.

Turning to the second panel of Table 5, which shows results of the B-O decomposition for the PJC regression, note that the discrimination component is once again very large. In the case of using men’s coefficients as the non-discriminatory wage structure, the discrimination component is once again more than the total, while the explained component is insignificant. If the wage structure of women is assumed as the non-discriminatory one, about 80 percent of the gap is found to be due to differences in coefficients, while the remaining is due to differences in characteristics. In this case, we find the explained component as statistically significant at 5 percent. Once again, the unexplained component of 0.368 in the pooled coefficient wage structure is the same as the coefficient of the gender dummy in the PJC regression.

Looking at both panels, the contribution of discrimination to the gender log wage gap is at least over 80 percent (PJC regression, using coefficients of women). Thus, looking at both specifications (PC and PJC) and various assumptions about the non-discriminatory wage structure, the smallest value of the gender log wage gap due to discrimination is 0.32, or a 38 percent gap in raw wage terms, and the maximum contribution of endowments is 0.073, or just over 7 percent. Throughout the paper we have seen that as far as differences in endowments are concerned, regular wage women do not lag significantly behind men, and in fact in some respects have better characteristics. Most of the gap, therefore, comes from the differences in coefficients, and is unexplained even when we control for job characteristics. This is taken as strong evidence of wage discrimination against women who are otherwise as qualified as men and are doing similar jobs as men.

5.4 The quantile regression results

5.4.1 Quantile regression results for the pooled sample

In the interest of brevity, I present the tables for the PC and the full scale PJC quantile regressions, only for the pooled sample (Table 6 and Table 7, respectively). The analysis of separate quantile regressions by gender is carried out through examples using graphs alone. Tables for the same can be provided by the author on request. The tables present the value of the coefficients and the t-statistics from log wage regressions for the 10th, 30th, 50th, 70th and 90th percentiles, along with the OLS coefficients for comparison.

The coefficient of interest is the coefficient on the gender dummy. Table 6 reveals that controlling for personal characteristics does not change the sticky floor effect we saw in the “uncontrolled” log wage gaps in Figure 1. After adding these controls gender log wage gap is as high as 0.68 at the 10th percentile and declines uniformly as one moves up the wage distribution to higher percentiles. The coefficient of 0.68 corresponds to a 100 percent wage differential in terms of raw wages at the 10th percentile. At the 90th percentile the log wage gap is close to 0.21, corresponding to a 23 percent wage differential in raw wages. Therefore controlling for personal characteristics does not alter the fact that log wage gaps are far greater at the bottom of the distribution than at the top of the distribution. Once again, it is important to keep in mind that this regression assumes that the returns to the various covariates (coefficients) are the same for men and women across all quantiles. The next section reveals that this is actually not the case. The decomposition that will follow (MMM decomposition) will allow for the fact that men and women have different returns to the same covariates, and is therefore a more robust indicator of the extent of discrimination across quantiles.

The coefficient of the gender dummy in Table 7 reflects the gender log wage gap at various percentiles after controlling for job characteristics in addition to personal characteristics. Once again, the sticky floor effect prevails. Figure 7 gives a better idea of how the gender log wage gap changes across quantiles once I control for these characteristics. The solid black line in the figure depicts the gender log wage gap without any controls, this is the same as shown in Figure 1. The reason this line looks smoother is that in Figure 7 I have plotted the gaps only for the nine deciles instead of each percentile. The dotted line in the figure shows the gender gap that remains after controlling for all characteristics. This is essentially a plot of all the quantile regression coefficients of the gender dummy from Table 7. As can be seen from the figure, the sticky floor still persists even after controlling for individual differences in labour market characteristics of men and women, while assuming returns to these are the same across gender. The sticky floor effect is, however, flatter. Note that the average log wage gap from the OLS regression reduced

from 0.4 (no controls) to 0.368 for the PJC regression. The QR technique tells us that the changes to the gap however are not the same at the bottom and the top of the distribution. At the bottom of the distribution, say at the 10th percentile the gender log wage gap reduces from 0.74 to 0.56. This reduction in gender gaps is seen up till the 6th decile. The gender gap remains unchanged at the seventh decile, and at the top two deciles the gap with controls is actually higher. At the 9th decile the gap with controls is 0.21 whereas the gap without controls is 0.07. By considering only the decline in average log wage gap from OLS regressions, one ignores the possibility that the gap could change differently at different points of the distribution. Thus, when controlling for individual differences in levels of covariates (while keeping returns to covariates to be the same for men and women), the gap reduces at lower quantiles and rises at higher ones. This means that at the bottom of the distribution, log wage gaps were overestimated, and were partly explained by the unfavourable distribution of covariates such as education, occupation, industry and other job characteristics that the low wage earning women face. The fact that more women are illiterate, and more women enter into Elementary occupations or low paying industries can explain this effect. Conversely, at the top of the distribution, wage gaps were “underestimated”. Controlling for the characteristics of women actually increases wage gaps at the top. This could be because at the top of the distribution women’s endowments were quite favourable. For example as seen in Table 2, a larger share of women are graduates and postgraduates, and work as professionals, technicians and associate professionals, compared to men. The declining coefficient of the gender dummy across quantiles as seen in Figure 7 would mean that the wage premium of being a regular wage male is more at the bottom of the distribution, than at the top. Therefore, a greater proportion of men in the population would reduce overall wage inequality.

I briefly turn to the quantile regression coefficients of the covariates other than the gender dummy in the pooled PJC regression (Table 7). Though this is not the focus of my paper, the examples and interpretations here will be useful for understanding the analysis in the next section, where I analyse returns to labour market characteristics separately for men and women. First, we examine the quantile regression coefficients as well as the ordinary least squares coefficient of the contract dummy which gets activated if the worker has a written job contract. This is a good example of a location shift effect. The 95 percent confidence interval of the OLS coefficient (0.14 to 0.20) contains all the quantile regression coefficients. In this case, therefore the OLS coefficient of 0.172¹⁴ is a good approximation of how belonging to a job with a contract affects wage at all points of the conditional wage distribution. Intuitively, this means that the gain due to having a job contract is almost the same for low, middle and high wage earners.

Figure 8 shows the impact of union membership on wages. Union membership affects wages at the top and bottom of the distribution quite differently. In Figure 8, the average impact on log wage as given by the OLS coefficient is about 0.219, which translates to a 23 percent premium in (raw) union wages. The quantile regression coefficients are all positive which means union membership increases wage throughout the distribution. The value of the coefficients, however, is not uniform across the distribution. From the figure it is evident that low wage earners benefit more than high wage earners from being union members. At the bottom of the distribution the gains are as high as 35 percent at the second decile, while at the 9th decile the gains are only about 14 percent. Union membership, therefore, has an inequality decreasing effect for overall wage inequality in the economy, since it benefits low wage earners

¹⁴ The OLS coefficient of 0.172 means that workers who have a written job contract, on average, earn 19 percent more than those who do not.

disproportionately more than high wage earners¹⁵. The dispersion of the wage distribution will decrease as unionization takes place among the regular wage workers.

5.4.2 Quantile regression by gender

In this section I look at quantile regression results by gender (Tables can be provided by the author on request). My analysis here will be based on graphs that plot these coefficients across quantiles, separately for men and women. Such an analysis is instructive because the same covariate may have very different effects on the wages of men and women, on average, and at different points of the distribution. From the previous section we know that a declining trend in quantile regression coefficients means that the covariate tends to decrease wage dispersion, whereas an increasing trend increases wage dispersion. In this section I examine how a particular characteristic affects wages within a gender group, and also make deductions about its impact on the gender wage gap by comparing the magnitude of the coefficients across gender.

Consider the marriage dummy which gets activated if the individual is currently married. We know that being married has very different effects for men and women because in the OLS PJC regression it was found that being married increases male wages significantly, whereas it had no significant effect on female wages. Figure 9 reveals a more detailed picture. The marriage dummy coefficient is positive for men throughout the distribution, but it decreases in magnitude as we move to higher deciles. This means that having more married men in the sample will reduce wage inequality within men. For women, however, the coefficient of the marriage dummy is positive for the lower half of the distribution and negative for the top half. This suggests that being married actually has a wage reducing effect for high wage earning women. The decreasing coefficients still mean that the wage dispersion within women will fall as the proportion of married women increases, as low wage earners will see an increase in wages and the high wage earner will see a reduction. What happens to gender wage gap as the proportion of married individual increases (keeping the ratio of proportion of married men to proportion of married women the same)? Marriage benefits men more than women: At almost all quantiles the coefficients of men are higher than that of women. This means marriage is associated with an increase in the average gender log wage gap.

As seen in Figure 10, union membership also affects men and women differently. It benefits both men and women, and that these benefits are more at the bottom of the distribution than at the top for both groups. It is interesting to note that union membership benefits women far more than it benefits men at every quantile. This means that if all men and women were to unionize there would be a convergence between male and female wages as the relative gains made by women would be greater than men. As a consequence, the average gender wage log gap would decrease. Also, since we see a downward trend for both, as noted in the previous section, overall wage dispersion also reduces.

As an example of the differential returns to education, I examine the returns to graduation¹⁶ in Figure 11. The coefficient for women is higher everywhere than that for men. This holds true for all other education categories. It is important to keep in mind that the reference category is that of illiterates. As one moves up the wage distribution, the pattern of a declining returns for women and rising returns to men in Figure

¹⁵ The pattern of declining union wage premium seen here was also noted in Chamberlain (1994). He used 1987 data for U.S manufacturing industries and found that union wage premium declines monotonically as one moves up the wage distribution from 28 percent at the bottom of the distribution to less than 1 percent at the top.

¹⁶ A typical graduate is one who has completed about 15 years of education in total

11, is also seen for secondary, higher secondary and postgraduates. The range of coefficients for women is 1.4 at the first decile of the distribution and 1 at the ninth decile. In raw wage terms these correspond to a 300 percent wage premium at the bottom and a 170 percent wage premium at the top. This declining trend means that as more illiterate women gain higher education, wage dispersion among women would decrease. In contrast, the coefficient for men rises from 0.4 at the bottom decile to about 0.6. This corresponds to a raw wage premium of about 50 percent at the bottom of the distribution and 80 percent at the top. Acquiring other higher education degrees also has this inequality increasing effect within men. The implication of the coefficients of women being higher than that of men for each education category is that the relative wage gains to illiterate women as they acquire greater levels of education are higher than the gains that illiterate men would enjoy. Therefore, if all illiterate men and women were to acquire these higher education degrees, the average gender wage gap would decrease.

Finally, I look at QR coefficients for high paying occupation dummies. These represent the wage premium vis-à-vis the omitted category of Elementary occupations. The overall wage inequality effects of attaining these positions are clearly in the direction of greater wage dispersion, for men and women alike. As Figure 12 shows, in the case of Managers and Senior Officials the increase in wage spread is more among women than men, given the steeper slope of the increase in quantile regression coefficients for women. Thus while low wage women benefit a lot less than low wage men, when it comes to acquiring managerial positions, high wage women gain almost exactly the same as high wage men. This is because at the ninth decile returns to managerial positions are almost the same for men and women. In the case of professionals, however, gains are insignificant for women at the bottom of the distribution, while they are quite large for men (graph not presented here). Though the coefficients for both men and women are increasing, the gap between the returns of men and women almost remains constant across the distribution. The trend seen of the quantile regression coefficients of men being (everywhere) higher than that of women is seen for all occupation categories. This implies that if women and men in Elementary occupations acquired managerial, professional and technical positions, average gender wage gap would increase as men would gain more than women at every quantile.

I find similar impacts on average gender wage gap when I look at the returns to various industry dummies, where the omitted category was the Mining and Manufacturing industry. Just as the OLS results indicated (Table 3 and Table 4), these results point to the fact that for certain characteristics such as higher education (illiterates as omitted category), union membership, and urban sector residence returns are higher for women than for men and therefore, have a gender gap reducing effect. Characteristics such as returns to high paying occupations, industries, working in the public sector and being married benefit (base category) men more than (base category) women and leads to an increase in the average gender gap. The quantile regressions are useful in telling us how these coefficients change across quantiles, and have different implications for low paid and high paid men and women. How inequality within gender gets affected can often differ by gender for the same covariate and the QR analysis helps us study questions such as this in more detail.

5.5 The Machado-Mata-Melly decomposition results

In this section the MMM decomposition is applied to the full scale PJC model¹⁷. Table 8 provides the decomposition results. The first column provides the deciles at which the MMM decomposition is computed. Column two provides the fitted values of log wages at different quantiles for men. These are estimated values using the unconditional distribution of men. Column three provides the fitted values for women. The difference between column two and three is the log wage gap to be decomposed and this is provided in column four. Note that this total differential is based on the two estimated unconditional distributions (rather than the empirical unconditional distributions), and hence the values of this total differential are slightly different from the value of the gender gap at each quantile as given by Figure 1. The decomposition in columns five to seven of Table 8 relies on constructing a counterfactual distribution using the returns of women and the covariates of men (CF_2). The log wages at different deciles of this counterfactual distribution constructed are given in column five. Column six is the difference between column five and column three, i.e. it is the difference in log wage between the counterfactual distribution and the estimated distribution for women at each decile. Due to the fact that the counterfactual is constructed using covariates of men and returns of women this difference is the portion of the log wage gap attributable to characteristics (explained gap). The remaining component of the gender log wage gap is given in column seven, which represents the contribution of differences in coefficients. Column seven therefore is calculated as the difference between column two and column five. This is interpreted as the discrimination component at each decile. Figure 13 plots these MMM decomposition results based on the counterfactual constructed using the returns of women and covariates of men. The solid black line represents the total differential to be decomposed (column four). The sticky floor effect is clearly seen as the gender log wage gaps fall from over 0.7 at the first decile to 0.1 at the ninth. The decomposition can tell us which factor is responsible for the sticky floor. The contribution of discrimination (or coefficients) is plotted as the dotted black line (column seven). Clearly, the trend of declining gender log wage gaps across quantiles is driven by that of declining discrimination across deciles. The discrimination component is about 0.6 at the first decile. In raw wage terms, this means that male wages are 82 percent higher than female wages *purely* due to discrimination at the first decile. At the median this log wage gap is about 0.30 which corresponds to a 35 percent raw gender wage gap. This is not very dissimilar to the contribution of discrimination at the mean which was 0.32, given in the PJC panel of Table 5 on the B-O decomposition results (using coefficients of women). Turning to the high wage earners, gender wage gaps due to discrimination are about 22 percent at the seventh decile and 10 percent at the ninth decile. Thus, the sticky floor effect persists even when I control for various characteristics and allow for the returns to these labour market characteristics to differ for men and women. In absolute magnitudes, the wage gaps caused due to discrimination are much higher for low wage earners than for high wage earners. The explained component in contrast is relatively small in magnitude and forms a very small part of the total gap throughout (column six). The maximum contribution of covariates is at the second decile where gender gaps in wages due to characteristics are 12 percent. This means that in the absence of discrimination, we would observe a gender gap of 12 percent at the second decile. Even at this decile the gender gap due to differences in coefficients was about 65 percent¹⁸. Therefore, in a relative sense

¹⁷Conclusions based on the decomposition results for the PC model are quite similar. Tables and figures for the same can be provided by the author on request.

¹⁸ Note here that I am not speaking of contribution to the total log wage gap. The gender gap in raw wages due to discrimination was 65 percent and 12 percent was due to differences in covariates. This adds up to the total gender gap of about 87 percent.

covariates did not contribute much even at the bottom of the distribution. The absolute contribution of characteristics falls to about only about 1 percent at the ninth decile. Thus, for high wage earners the contribution of differences in characteristics is close to zero. Even though the overall log wage gap is lower at the top, still most of that 10 percent gender gap in wages remains unexplained, i.e., is attributable to discrimination.

Just as in the B-O decomposition, the MMM decomposition is also subject to an index number problem and could also be done based on the alternative specification of the counterfactual distribution. Columns eight to ten of Table 8 correspond to the MMM decomposition results that use the counterfactual distribution based on the returns of men combined with the covariate distribution of women (CF_1). The difference between column two and column eight now gives us the contribution of differences in endowments of men and women (column nine). The difference between the log wage at a given decile of the counterfactual distribution and the estimated distribution for women is entered in column ten as the contribution of differences coefficients (or discrimination). Overall, once again we see from the table that the discrimination component dominates and is responsible for the sticky floor effect we see in the total log wage gaps. The distribution of characteristics is not very unfavourable to women. In fact, these results show that at the top the distribution women are actually at an advantage with respect to men in terms of characteristics. Everywhere beyond the fourth decile the discrimination component is actually larger than the total differential and the explained component is negative. This means that in the absence of discrimination, in the top half of the distribution women would actually end up earning more than men. At the 80th percentile, for instance women would have earned 4 percent more than men had there been no discrimination. Once again, therefore, discrimination is almost entirely responsible for the sticky floor effect, which persists even after controlling for all available personal and job characteristics.

6 Conclusion: A summary of results and their interpretation

6.1 Summary of results

In this paper I used data from the 66th round (2009-10) of the Employment-Unemployment Survey (EUS) of the National Sample Survey Organization and examined gender wage inequality for the regular wage workers in India in considerable detail. The focal point of this analysis was to look at log wage gaps at different points of the distribution and not just at the mean. The main finding was that of a sticky floor effect, that is, the phenomenon of declining gender log wage gaps across the quantiles. At the first decile, “uncontrolled” log wage gap was 0.74 and at the ninth decile it was only 0.07. These correspond to a raw wage gap of 109 percent and 7 percent, respectively. Summary statistics revealed some factors that may be contributing to the sticky floor effect: a) Log wage gaps were lower for the university educated and higher among illiterates. b) Across occupations, higher log wage gaps were found in Elementary occupations as opposed to high paying occupations such as Managers, professionals and associate professionals. c) I also find smaller wage gaps in urban areas. d) Gender gaps are highest among lower castes, especially SCs and STs, lower for OBCs and the lowest among Others. These fit with the finding of an overall sticky floor. I also examine the distribution of log wage gaps *within* different education levels, occupations, industries and sectors, and find that a sticky floor exists within these categories as well.

Gender wage gaps exist and are a reality in virtually every economy, though the degree varies across countries. Given an average log wage gap, I now make normative judgments about which of these phenomena (glass ceiling or sticky floor) I find more unsettling. In India, while minimum wage differs

across states and industries, in order to bring some uniformity to the minimum wage structure, the National Floor Level Minimum Wage (NFLMW) was introduced in the early nineties by the Ministry of Labour and Employment. The NFLMW is non-statutory. It can be thought of as a floor for minimum wages across the country, in the sense that all states are encouraged to ensure that their minimum wages are set above this level. The NFLMW is frequently revised in order to account for rising price levels¹⁹. Figure 14 shows the percentiles at which men (solid vertical) and women (dashed vertical) attain the NFLMW of Rs. 100 per day. While for men 22 percent of the regular wage working population is below the NFLMW, almost 42 percent of such women are below it. This disparity between men and women conforms to the observation of the sticky floor documented in Figure 1. The presence of a glass ceiling effect, on the other hand, would mean that the percentile at which men and women achieve the minimum wage would be very close to each other as wage gaps are lower for low wage earners. Not only is the disparity between men and women important, but more important is the fact that 42 percent of women get paid wages that are lower than the NFLMW even though they work as regular wage workers. If minimum wage laws were strictly adhered to these figures should have been close to zero. This data provides clear evidence that minimum wage laws are not implemented in practice. In most developed nations average gender gaps are lower and only a small proportion of the working population is below minimum wage. In the United States the prevailing federal minimum wage is 7.25 \$ per hour²⁰. For hourly wage workers²¹ in the year 2010 only 6 percent of workers worked at or below the minimum wage. The percentage of men who were paid either the minimum wage or lower was 4.5 percent, whereas for women this figure was 7.3 percent. This is in sharp contrast to the Indian case in terms of both the overall proportions and the gap between proportions of men and women who are not paid the minimum wage. Evidence of a sticky floor implies that women at the lower end of the overall wage distribution experience larger wage gaps compared to women at the upper end. Large shares of these low wage earners do not even enjoy minimum wages required to sustain a decent standard of living. Using estimations from the EUS we know that in 2009-2010, over 43 lakh women working in the regular wage sector earn wages below the NFLMW. Given these facts, evidence of a sticky floor as opposed to that of a glass ceiling is more worrying in my view. The Central and state governments need to ensure that minimum wage laws are not violated. This may result in a disappearance of the sticky floor from the bottom end of the wage distribution, as currently more women earn wages below the minimum wage level compared to men.

The log wage gaps that I document could be due to differences in the characteristics of men and women or they could be due to gender differences in the rates of return to these characteristics. I therefore looked at some average levels of various characteristics to examine whether there were significant gender differences. I found that on average, men were older than women; the proportions of women at the two ends of the educational and occupational spectrum (those with university degrees and in Professional jobs on the one hand, and those who are illiterate or in Elementary jobs on the other) were higher than corresponding proportions for men. I estimated standard OLS Mincerian wage equations separately for men and women and found that returns to education (for illiterates) and union membership are higher for women than men. The gain by attaining managerial and professional positions was greater for men than

¹⁹ In its most recent revision, the NFLMW was increased to Rs. 115 per day in April 2011. The previous revision was undertaken in November 2009 from Rs. 80 per day to Rs. 100 per day. The 66th round data that this paper relies on was conducted from July 2009 to June 2010 and therefore the relevant NFLMW for this paper is Rs. 100/ Day.

²⁰ Source: <http://www.bls.gov/cps/minwage2010tbls.htm> accessed on 7/5/2012

²¹ Workers paid at hourly rates comprise about three- fifths of wage and salary workers in the United States

women (with base category as Elementary occupations). This preliminary analysis pointed toward varying characteristics and returns across gender.

To quantify the contribution of differences in characteristics and the differences in returns toward average log wage gap, I applied the popular mean wage decomposition method first proposed by Blinder and Oaxaca (1973). I find that the discrimination (differences in returns) is dominant in explaining the average gender log wage gap. Across different specifications, discrimination, at the minimum, accounted for over 80 percent of the average log wage gap and at the maximum explained more than the total. “Discrimination” accounting for over 100 percent of the gap means that if women had enjoyed the same returns to characteristics as men, then, on average, they would earn more than men because of their superior characteristics.

Next I estimated earnings functions using quantile regressions. The benefit of using quantile regressions over OLS wage regressions is that they allow returns of covariates to vary at different points of the wage distribution. In a pooled regression of men and women, the declining coefficient of the gender dummy across percentiles, even after adding several controls was the first indication that the sticky floor effect persists even when I control for differences between labour market endowments of men and women. Drawing on the inequality literature I also looked at the effects of some covariates on overall wage inequality. For example, I find that union membership benefits the low wage earners more than the high wage earners, and therefore conclude that move towards unionization reduces overall wage inequality.

Moving from pooled quantile regressions, I then estimated the quantile regressions separately for men and women. This highlighted how the same covariate had different implications for inequality within a particular gender and gender inequality between groups. Returns to higher education, occupation, unionization and marriage were analysed for men and women as select examples. For instance, as illiterate men acquire higher education inequality increases, as returns to those at the bottom of the distribution are smaller. The reverse is true of women, where benefits of higher education to the low wage workers are greater as compared to high wage earners. Since the returns to higher education that illiterate women enjoy are higher than those for illiterate men, this covariate also has the effect of reducing the gender gap as these men and women acquire higher education. In contrast, returns to top jobs as measured by coefficients of the high wage occupation dummies were higher for men than for women working in Elementary occupations. Therefore, even if men and women were promoted at the same rates, this would lead to an increase gender gap. I also find that an increase in the proportion of married individuals would result in an increase in the average gender log wage gap.

Finally, I applied the Machado-Mata-Melly (MMM) decomposition to decompose gender wage gaps at each quantile. The MMM decomposition is far more informative than the B-O decomposition and is essential for a study of wage gaps across the distribution. It performs the decomposition at different points of the distribution and therefore enables the quantification of endowment and discrimination component in a manner that allows returns to vary at different points of the distribution. The MMM decomposition shows that the contribution of the explained component (i.e. gender wage gap due to differences in characteristics/ endowments) is quite low at all quantiles. This means that for both low wage earners and high wage earners, the distribution of characteristics is not very unfavourable to women. We conclude that it is the discrimination component that drives the sticky floor result. For low wage earners log wage gaps due to discrimination are very large (over 0.6 at the first decile) and for high wage earners these are smaller (about 0.1 at the 9th decile).

6.2 Comparisons with the literature

The sticky floor effect I find is in sharp contrast to other studies on developed countries where there is evidence of increasing gender wage gap across quantiles, better known as the glass ceiling effect. Albrecht et al. (2003) was one of the first papers to identify this phenomenon for Sweden. Arulampalam et al. (2007) examine 11 European nations²² in their study and find some evidence of a sticky floor only in Italy and Spain, and a glass ceiling in all other nations. de la Rica (2005) finds evidence for a glass ceiling in Spain among workers with high education and a sticky floor effect for less educated. Kee (2005) finds evidence of a strong glass ceiling effect in the Australian private sector labour market. Recent studies on China (Chi and Li, 2007), Thailand (Fang and Sakellariou, 2010), Sri Lanka (Gunewardena, 2008), Vietnam (Pham and Reilly, 2006) and Philippines (Sakellariou, 2004) find a sticky floor effect for all these countries. These studies provide evidence for an emerging trend of finding a sticky floor in developing economies, and a glass ceiling in the developed nations. To the best of my knowledge, this paper is the first to examine the trend in log wage gaps across percentiles for the Indian labour market and my findings are similar to those found for other South Asian countries – that of a sticky floor for regular wage workers in India.

Another interesting point to note is that the magnitudes of log wage gaps, on average and across quantiles are much larger for India as compared to other European nations. Consider the average wage gaps for the 24 countries examined in Christofides et al. (2010). Only three of those European nations²³ had average gender log wage gaps greater than those found in India. Among the 11 nations studies in Arulampalam et al. (2007) the largest average gender log wage gaps was found in Britain (0.25) and the lowest in Italy (0.063). These are in sharp contrast to what I find in the case of the Indian regular wage market, where the average gender log wage gap is 0.4.

The decline in wage gaps as one moves from the bottom end to the top of the wage distribution is also quite drastic in the Indian case. According to Arulampalam et al. (2007) a sticky floor is defined to exist if the 10th percentile wage gap is higher than the 25th percentile wage gap by 2 percentage points. An alternate weaker definition would be to maintain the 2 percentage points criterion but compare the 10th and the 50th percentile instead. These definitions have become fairly popular in the literature and are used in several papers as a rough rule to establish the presence of a sticky floor. In this paper the unconditional log wage gap at the 10th percentile is 0.74, whereas the gap at the 25th percentile is 0.54. This is a 20 percentage point difference, far greater than the 2 percentage point difference usually referred to. The percentage point difference between log wage gaps at the 10th percentile and the 50th percentile is even greater (30 percentage points). If we look at the gender gaps due to the discrimination component alone in the MMM results of Table 8, again we find a very steep sticky floor that more than satisfies the Arulampalam et al. (2007) criteria. Therefore, the sticky floor effect in India is particularly “strong” when compared to other nations that find a similar effect, especially in Europe.

²² Austria, Belgium, Britain, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands and Spain

²³ Cyprus, Estonia and the Czech Republic

6.3 Possible reasons for the sticky floor effect

One explanation for the sticky floor finding that is often cited in the literature is related to the statistical discrimination by employers due to low labour force participation rates of women²⁴. In India, for instance, traditionally women bear the primary household responsibilities. Societal and religious norms place the burden of family duties disproportionately on women. Because of this, men are perceived to be more stable in jobs vis-à-vis women. Given the higher probability of dropping out of the labour market, employers discriminate against women as they enter into the labour market because they expect future career interruptions. As women move up the occupation structure and gain job experience, employers become aware of their reliability and therefore, discriminate less. Men usually have more work experience or tenure than women on average. In addition, we know from Table 4 that an additional year working for women yields greater gains (especially since we find that the age profiles of men are more concave than women). Each year of experience that women gain therefore narrows the gender gap. Studies using panel data that follow the gender gap as women gain more experience within the firm also confirm this explanation. On the other hand, women who have high levels of education and are at the top end of the distribution are perceived to have high levels of commitment and due to their past investments in education are thought to be stable employees.

At the higher end of the wage distribution the nature of jobs are very different from those at the bottom. The women working in these jobs are more likely to be the urban educated elite working in managerial or other professional positions. Firstly, the scope to discriminate against these workers is lesser given their attributes and backgrounds. Even if strong anti-discriminatory legislations are in place, these high wage earning women are more likely to be aware of their rights and may be in a better position to take action against perceived discrimination. According to Arulampalam et al. (2007), “only the more articulate and better educated are willing to take legal action against breaches of the law” (p. 176). One way to do so would be through formal mechanisms within the organisation. In addition to this, it is possible that these women would have options to counter such discriminatory practices by moving to other jobs. Employers would be aware of these possibilities themselves and hence, may not be able to discriminate a great deal between similarly qualified men and women. Secondly, the payment mechanism in these jobs would be far more structured and rigidly defined. Whether in the public sector or the private sector, most high paying jobs will have written contracts with predefined clauses for basic increases in salaries, year on year.

Contrast this to a situation where an employer is paying a regular wage to a woman with no education working in an Elementary occupation, which is a typical example of a worker at the bottom of the wage distribution in the Indian context. It is easier for the employer to discriminate in this case, even if the woman possesses identical personal characteristics to a man in terms of education (or its lack thereof), age, caste status etcetera, and, even if their job characteristics are the same in terms of sector, industry and occupation. This is because these jobs are in the informal sector and outside the jurisdiction of labour laws. Women at the bottom have less bargaining power compared to men due to family commitments or social custom and are more likely to be subject to the firms’ market power. Chi and Li (2007) noted that compensation practices for jobs at the low end of the distribution in China are much less regulated and that unskilled and that illiterate women workers are treated particularly unfavourably due to an abundant

²⁴ For Spain, de la Rica et al. (2005) explained the sticky floor effect for workers with low education using a similar argument

supply of unskilled male labour. Thus, a sticky floor could arise because anti-discriminatory policies are more effective at the top of the distribution.

Article 39 of the Indian constitution envisaged equal pay for equal work for both men and women. To this end legislations such as the equal remunerations act (1976) were enacted after the equal remuneration ordinance was introduced in the year 1975. Given the fact that 42 percent of women in the population under study were below even the NFLMW, it is clear that legal obligations set by minimum wage legislations are rarely binding for such jobs. Absence of strong minimum wage legislations means that wage gaps can be larger at the bottom of the distribution.

Job segregation is also a known contributor to wider gaps at the bottom as men and women only enter into exclusively 'male' and 'female' jobs. Low skilled jobs for women may pay less than other jobs that require intense physical labour, which men typically do. Our model specification controls for broad industry and occupation groups. This means that within certain low paying broad industrial categories men and women could be doing different kinds of jobs and that could be picked up as the discrimination component. Chi and Li (2007) find that the sticky floor in China is associated with a particularly low-paid group of production workers.

A limitation of a cross-section dataset of one country is that country-specific policies that affect wage distributions are unobservable and are subsumed in the intercept terms²⁵. Christofides et al. (2010) and Arulampalam et al. (2007) are two notable studies that examine how country specific policies and institutions can affect gender wage gaps on average, and over entire distribution. The OECD work-family reconciliation index is a composite index based on indicators of childcare provisions, maternity leave, voluntary part-time work, etc. Their results show that there is a negative correlation between the work-family index and the sticky floor²⁶ and a positive correlation between the index and a glass ceiling. The implication is that countries where the index is higher, that is, countries with greater concessions for working women to balance work with household responsibilities, exhibit a glass ceiling effect. Countries where the index is smaller are more likely to be characterized by large gaps at the bottom of the distribution and small gaps at the top. In India, childcare provisions, maternity leave, voluntary leave and other concessions such as these are certainly not available to the women at the lower end of the wage distribution. However, these provisions would be available to the high-wage earners. By enabling women to preserve their ties with the organisation they are working with and therefore incentivising investment in education and training, these women would earn relatively higher wages as compared to those who do not benefit from these policies. Women at the bottom of the distribution who do not get the benefits of such provisions would have a higher probability of dropping out of the labour force or participate less intensely in the labour market and employers could discriminate against them for this reason. This could be another possible reason for the sticky floor effect.

²⁵ The detailed B-O decompositions (not presented here) reveal that a substantial part of the discrimination component stems from the differences in the intercept terms for men and women.

²⁶ The Glass ceiling is defined based on the 90-50 pay difference and the sticky floor defined on the 10-50 pay difference

7 References

- Albrecht, James, Björklund, Anders and Vroman, Susan (2003), "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics* 21: 145-178.
- Arulampalam, Wiji, Booth, Alison L., Bryan, Mark L. (2007), "Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution," *Industrial and Labor Relations Review*, ILR Review, Cornell University, ILR School, vol. 60(2), pages 163-186, January.
- Azam, Mehtabul, Prakash, Nishith (2010), "A Distributional Analysis of Public Private Wage Differential in India", IZA Discussion Paper No. 5132
- Banerjee, Biswajit & Knight, J.B. (1983), "Caste discrimination in the Indian urban labour market", *Journal of Development Economics* 17 (1985) 277-307
- Blinder, Alan (1973), "Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources* 8:436-455
- Brown, Randall S., Moon, Marilyn & Zoloth, Barbara S. (1980), "Occupational Attainment and Segregation by Sex", *Industrial and Labor Relations Review*, Vol. 33, No. 4 (Jul., 1980), pp. 506-517
- Chamberlain, Gary (1994), "Quantile Regression, Censoring and the Structure of Wages", *Advances in Econometrics*. Christopher Sims, Ed. New York: Elsevier, pp171-209
- Chi, Wei & Li, Bo (2008), "Glass Ceiling or Sticky Floor? Examining the Gender Pay Gap across the Wage Distribution in Urban China, 1987-2004", *Journal of Comparative Economics*, 36 243–263
- Christofides, Louis N. & Polycarpou, Alexandros & Vrachimis, Konstantinos (2010), "The Gender Wage Gaps, 'Sticky Floors' and 'Glass Ceilings' of the European Union," IZA Discussion Papers 5044, Institute for the Study of Labor (IZA)
- Cotton, Jeremiah (1988), "On the Decomposition of Wage Differentials", *The Review of Economics and Statistics*, Vol. 70, No. 2 (May, 1988), pp. 236-243
- de la Rica, Sara, Dolado, Juan J. and Llorens, Vanesa (2005), "Ceiling and Floors: Gender Wage Gaps by Education in Spain", IZA Discussion Paper No. 1483. IZA Discussion Papers 1483, Institute for the Study of Labor (IZA)
- DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica* 64: 1001-1044.
- Duraisamy, M and Duraisamy, P (1999), "Women in the Professional and Technical Labour Market in India: Gender Discrimination in Education and Employment and Earnings", *Indian Journal of Labour Economics*, Vol. 42, No.4, pp. 599-612.
- Fairlie, Robert W. (2005), "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models", *Journal of Economic and Social Measurement*, 30: 305-316.
- Fang, Z. and Sakellariou, C. (2011), "A Case of Sticky Floors: Gender Wage Differentials in Thailand", *Asian Economic Journal*, 25: 35–54. doi: 10.1111/j.1467-8381.2011.02047.x
- Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux (2007), "Decomposing Wage Distributions using Recentered Influence Functions Regressions", mimeo, University of British Columbia.

- Fortin, Nicole M., Lemieux, Thomas, Firpo, Sergio (2011), *Decomposition Methods in Economics* (edited by Ashenfelter and Card.), Chapter 1, *Handbook of Labour Economics* (Volume 4A)
- Gomulka, Joanna, Stern, Nicholas (1990), "The employment of married women in the United Kingdom, 1970-1983", *Economica* 57: 171-199
- Gunewardena, Dileni, Abeyrathna, Darshi, Ellagala, Amalie, Rajakaruna, Kamani and Rajendran, Shobana, (March 2008) "Glass Ceilings, Sticky Floors or Sticky Doors? A Quantile Regression Approach to Exploring Gender Wage Gaps in Sri Lanka", PMMA Working Paper No. 2008-04. Available at SSRN: <http://ssrn.com/abstract=1124158> or <http://dx.doi.org/10.2139/ssrn.1124158>
- Hnatkovska, Viktoria , Lahiri, Amartya, Paul, Sourabh B., forthcoming, "Castes and Labor Mobility," *American Economic Journal: Applied Economics*
- Hung, Reilly, Pham, Barry, T., (2006), "The Gender Pay Gap In Vietnam, 1993-2002: A Quantile Regression Approach", No 34, PRUS Working Papers, Poverty Research Unit at Sussex, University of Sussex.
- Jenkins, Stephen P., (1994), "Earnings discrimination measurement: A distributional approach", *Journal of Econometrics*, 61: 81-102
- Kee, H. J. (2006), "Glass Ceiling or Sticky Floor? Exploring the Australian Gender Pay Gap", *Economic Record*, 82: 408–427. doi: 10.1111/j.1475-4932.2006.00356.x
- Kingdon, G. G. and Unni, J. (2001). "Education and Women's Labour Market Outcomes in India." *Education Economics* 9(2): 173-195
- Koenker, Roger, Bassett, G., (1978), "Regression Quantiles", *Econometrica*, 46, 33- 50.
- Machado, José F., Mata, José (2005), "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression", *Journal of Applied Econometrics* 20: 445–465.
- Madheswaran, S. & Khasnobis, B.G (2007) " Gender Discrimination in the Labour Market: Evidence from the NSS", WIDER research project on "Gender wage Gap and its Impact on poverty: Evidence from India"
- Melly, Blaise (2005), "Decomposition of differences in distribution using quantile regression," *Labour economics* 12: 577-590.
- Melly, Blaise (2006), "Estimation of counterfactual distributions using quantile regression," University of St. Gallen, Discussion Paper.
- Mincer, Jacob A. (1974), "Schooling, Experience, and Earnings," NBER Books, National Bureau of Economic Research, Inc, number minc74-1.
- Neumark, David (1988), "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination", *The Journal of Human Resources*, Vol. 23, No. 3 (Summer, 1988), pp. 279-295
- Oaxaca, Ronald L. (1973), "Male-Female Wage Differentials in Urban Labour Markets", *International Economic Review* 14: 693-709.
- Oaxaca, Ronald L. & Ransom, Michael R. (1994) "On discrimination and the decomposition of wage differentials," *Journal of Econometrics*, Elsevier, vol. 61(1), pages 5-21, March.
- Reilly, B. and Dutta, P.V. (2005), "The Gender Pay Gap and Trade Liberalization: Evidence for India", Poverty Research Unit at Sussex Working Paper, 3
- Sakellariou, Chris (2004), "The use of quantile regressions in estimating gender wage differentials: a case study of the Philippines," *Applied Economics*, Taylor and Francis Journals, vol. 36(9), pages 1001-1007

8 Appendix

8.1 Tables

Table 1: Nominal average daily wage by gender and average wage differential, across NSS rounds

Statistic/Year	1999-2000	2004-2005	2009-2010
Daily wage men (Rs./Day)	151.42	176.90	325.00
Daily wage women (Rs./Day)	116.93	131.89	260.31
Avg. gender wage differential (%)	29.50	34.13	24.85
N (Sample Size)	38,649	41,277	35,363
Population Size	46,636,357	54,068,431	58,976,514

Table 2: Summary statistics – Shares (distribution) and average daily wages, by gender, 2009-2010

Dependent Variables	Share (%)			Avg. Daily wages (Rs.)			Average gender wage differential (%)
	Men	Women	Pooled	Men	Women	Pooled	
Sector - Rural/Urban							
Rural	34.8	36.2	35	251	162.78	234.99	54.19
Urban	65.2	63.8	65	364.49	315.74	356.09	15.44
Sector - Public/ Private							
Private	68.14	60	66.72	248.61	196.84	240.49	26.3
Public	31.86	40	33.28	508.21	373.61	480.02	36.03
Education Level							
Illiterates	6.23	14.69	7.72	146.51	87.12	126.67	68.16
Primary & Below	14.07	12.22	13.74	157.71	96.65	148.18	63.17
Middle	16.48	10.71	15.47	188.02	108.79	178.39	72.84
Secondary	18.24	10.58	16.9	260.1	179.29	251.22	45.08
Higher Secondary	11.66	9.38	11.26	324.12	225.1	309.64	43.99
Graduate/Diploma	25.58	29.97	26.35	496.23	395.29	476.08	25.54
Postgraduate	7.74	12.45	8.57	652.84	526.03	620.48	24.11
Occupations							
Managers/Legislators	4.52	1.57	4	734.7	764.87	736.78	-4.1
Professionals	9.78	16.57	10.97	617.72	450.32	573.36	37.2
Technicians/Assoc. Pr	11.61	27.43	14.39	467.54	284.37	406.26	64.4
Clerical	11.55	10.45	11.36	392.32	366.78	388.2	7
Service/Sales	16.59	13.03	15.96	222.05	190.96	217.6	16.3
Craft/Skilled Agr.	18.42	8.36	16.66	212.59	102.88	202.93	106.6
Plant/Machine Opera	15	2.09	12.73	228.52	165.7	226.7	37.9
Elementary	12.53	20.52	13.94	172.99	100.38	154.23	72.3

Dependent Variables	Share (%)			Avg. Daily wages (Rs.)			Average gender wage differential (%)
	Men	Women	Pooled	Men	Women	Pooled	
Industry							
Agriculture	2.99	2.99	2.99	137.05	76.23	126.39	79.78
Manufacturing/Mining	27.12	13.56	24.74	254.23	160.82	245.24	58.08
Elec., Gas, Water/Constr.	4.72	1.62	4.18	362.79	468.04	369.96	-29.01
Retail/Hotels	13.77	4.83	12.2	169.46	142.95	167.61	18.54
Transport/Comm.	13.46	2.23	11.49	287.22	404.87	291.24	-40.96
Fin. Intermediation	4.17	3.04	3.97	572.48	531.12	566.92	7.79
Real Estate/Renting	4.33	3.83	4.24	494.69	440.45	486.1	12.31
Public Admin. /Defence	13.69	10.4	13.11	477.66	360.3	461.32	32.57
Education	10.36	33.35	14.39	446.32	304.36	388.58	46.64
Health & Social Work	2.28	9.75	3.59	425.13	234.33	334.23	81.42
Other Services	3.11	14.39	5.09	200.4	124.01	162.5	61.59
Caste Profiles							
SC	16.41	19.7	16.99	262.19	157.93	240.96	66.02
ST	4.77	5.15	4.83	289.64	213.64	275.44	35.57
OBC	35.76	34.63	35.56	288.42	217.87	276.36	32.38
Others	43.06	40.52	42.62	383.31	352.26	378.13	8.82
Union Membership							
No Union	68.44	69.12	68.56	253.29	185.75	241.36	36.36
Union Members	31.56	30.88	31.44	479.98	426.63	470.8	12.5
Permanent Job							
Temporary	32.07	31.1	31.9	173.82	126.69	165.77	37.2
Permanent	67.93	68.9	68.1	396.37	321.74	383.14	23.2
Written job contract							
No Contract	63.75	60.1	63.11	245.2	195.39	236.94	25.49
Contract	36.25	39.9	36.89	484.18	375.5	463.71	28.94
TOTAL							
Total	100	100	100	325	260.31	313.65	24.85

Table 3: OLS log wage Personal Characteristics regression, pooled sample and separately by gender, 2009-2010

Independent Variables	Pooled			Men			Women		
	Coefficient	S.E	Significance	Coefficient	S.E	Significance	Coefficient	S.E	Significance
Constant	3.258	0.093	***	3.678	0.098	***	2.955	0.219	***
Age	0.030	0.005	***	0.035	0.006	***	0.034	0.012	***
Age Squared /100	-0.006	0.006		-0.012	0.007	*	-0.010	0.015	
Gender	0.399	0.022	***	-			-		
Sector	0.241	0.019	***	0.214	0.021	***	0.360	0.041	***
Marital Status	0.132	0.022	***	0.122	0.026	***	0.082	0.040	**
Primary & Below	0.191	0.029	***	0.101	0.029	***	0.269	0.064	***
Middle	0.342	0.029	***	0.245	0.029	***	0.458	0.072	***
Secondary	0.557	0.029	***	0.444	0.029	***	0.804	0.075	***
Higher Sec.	0.781	0.031	***	0.654	0.032	***	1.083	0.074	***
Graduate/Diploma	1.208	0.032	***	1.047	0.035	***	1.614	0.059	***
Postgraduate	1.499	0.036	***	1.347	0.038	***	1.855	0.074	***
ST	-0.065	0.034	*	-0.067	0.035	*	-0.065	0.088	
SC	-0.092	0.021	***	-0.084	0.021	***	-0.083	0.055	
OBC	-0.101	0.019	***	-0.103	0.021	***	-0.077	0.045	*
State Dummies	Yes			Yes			Yes		
Regression Summary									
Number of observations	35,337			28,977			6,360		
Population size	58,946,386			48,597,848			10,348,538		
F(48, 35289)	230.87			196.19			55.12		
Prob> F	0			0			0		
R-squared	0.475			0.4602			0.5071		

Table 4: OLS log wage Personal Job Characteristics regression, pooled sample and separately by gender, 2009-2010

Dep Variable: Ln(wage)	Pooled			Men			Women		
	Coefficient	S.E	Significance	Coefficient	S.E	Significance	Coefficient	S.E	Significance
Constant	3.415	0.087	***	3.782	0.092	***	3.318	0.201	***
Age	0.026	0.005	***	0.031	0.005	***	0.030	0.010	***
Age Square/100	-0.017	0.006	***	-0.023	0.006	***	-0.021	0.013	
Gender	0.368	0.022	***						
Sector	0.256	0.020	***	0.218	0.021	***	0.369	0.040	***
Marital Status	0.079	0.020	***	0.082	0.023	***	0.026	0.047	
ST	-0.099	0.031	***	-0.108	0.032	***	-0.107	0.078	
SC	-0.124	0.019	***	-0.131	0.019	***	-0.092	0.051	*
OBC	-0.097	0.017	***	-0.100	0.019	***	-0.088	0.040	
Primary & Below	0.077	0.028	***	0.003	0.029		0.175	0.063	***
Middle	0.183	0.027	***	0.113	0.028	***	0.290	0.066	***
Secondary	0.301	0.028	***	0.226	0.028	***	0.544	0.082	***
Higher Sec.	0.403	0.032	***	0.301	0.032	***	0.788	0.090	***
Graduate/Diploma	0.651	0.040	***	0.496	0.041	***	1.209	0.093	***
Postgraduate	0.837	0.051	***	0.646	0.046	***	1.445	0.132	***
Job Characteristics (General)									
Public	0.305	0.026	***	0.333	0.030	***	0.280	0.047	***
Union	0.219	0.016	***	0.179	0.016	***	0.378	0.041	***
Contract	0.173	0.017	***	0.174	0.017	***	0.169	0.042	***
Permanent	0.228	0.016	***	0.211	0.017	***	0.266	0.039	***
Job Characteristics (Occupation)									
Manager/Legislators	0.556	0.047	***	0.628	0.051	***	0.345	0.129	***
Professionals	0.514	0.035	***	0.555	0.037	***	0.182	0.086	**
Technicians/Assoc. Prof.	0.294	0.031	***	0.367	0.032	***	-0.052	0.077	
Clerical	0.138	0.027	***	0.156	0.028	***	-0.115	0.078	
Service/Sales	0.052	0.028	*	0.079	0.025	***	-0.075	0.083	

Table 4 Continued...

Dep Variable: Ln(wage)	Pooled			Men			Women		
	Coefficient	S.E	Significance	Coefficient	S.E	Significance	Coefficient	S.E	Significance
Crafts and other trades	0.076	0.024	***	0.102	0.024	***	-0.150	0.081	*
Plant/Machine operator	0.093	0.034	***	0.108	0.034	***	-0.054	0.098	
Job Characteristics (Industry)									
Agriculture	0.001	0.086		-0.045	0.092		0.193	0.281	
Elec/Gas/Water	0.136	0.042	***	0.127	0.047	***	0.100	0.086	
Retail/ Hotels	-0.119	0.024	***	-0.130	0.023	***	-0.265	0.086	***
Transport / Comm.	0.064	0.034	*	0.054	0.036		-0.013	0.103	
Financial Intermediation	0.113	0.037	***	0.115	0.039	***	0.023	0.095	
Real estate	0.235	0.037	***	0.238	0.040	***	0.150	0.101	
Public ad. / Defence	-0.006	0.040		0.006	0.044		-0.251	0.082	***
Education	-0.269	0.036	***	-0.217	0.041	***	-0.481	0.076	***
Health, Social work	-0.173	0.042	***	-0.095	0.044	**	-0.356	0.082	***
Other Services	-0.192	0.039	***	-0.161	0.036	***	-0.323	0.085	***
State Dummies	Yes			Yes			Yes		
Regression Summary									
Number of observations	33,720			27,685			6,035		
Population size	56,044,047			46,305,777			9,738,270		
Design df	33,719			27,684			6,034		
F(48, 35289)	266.3			221.63			70.7		
Prob> F	0			0			0		
R-squared	0.5824			0.5784			0.6118		

Table 5: Aggregate Blinder-Oaxaca decomposition results PC and PJC model, using men, women and pooled wage structures (WS)

Terms of the decomposition	PC Regression			PJC Regression		
	Men WS	Women WS	Pooled WS	Men WS	Women WS	Pooled WS
Difference ($\log(w_m) - \log(w_w)$)	0.402	0.402	0.402	0.391	0.391	0.391
Explained (E)	-0.006	0.005	0.003	-0.009	0.073	0.023
Unexplained (D)	0.408	0.397	0.399	0.4	0.318	0.368
Percentage Unexplained	101.45	98.82	99.18	102.25	81.35	94.23

Table 6: Quantile regressions for the PC specification (Pooled sample), 2009-2010

DepVar: ln(daily wage)	Percentiles											
Independent Vars	10 th		30 th		50 th		70 th		90 th		OLS	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Constant	2.28	23.95	2.99	28.28	3.40	47.58	3.73	52.6	3.93	43.23	3.258	35.08
Age	0.043	8.15	0.025	4.36	0.020	5.18	0.025	6.61	0.051	11.12	0.030	5.88
Age Sq./100	-0.040	-6.03	-0.004	-0.6	0.009	1.84	0.007	1.46	-0.029	-5.09	-0.006	-0.95
Gender	0.683	36.19	0.530	25.23	0.417	28.62	0.288	19.54	0.214	10.18	0.399	18.48
Sector	0.277	13.39	0.252	13.45	0.208	17.18	0.197	17.25	0.173	11.86	0.241	12.61
Marital Status	0.223	10.12	0.172	7.68	0.125	8.09	0.080	5.4	0.039	1.78	0.132	5.89
Prim & Below	0.176	5.88	0.174	5.02	0.220	9.03	0.168	7.05	0.145	4.65	0.191	6.66
Middle	0.265	8.7	0.291	8.33	0.350	14.43	0.322	13.68	0.319	10.82	0.342	11.83
Secondary	0.416	13.9	0.482	14.31	0.564	24.13	0.561	24.66	0.553	19.32	0.557	19.46
Higher Secondary	0.537	17.5	0.642	18.18	0.790	31.87	0.865	35.54	0.810	27.37	0.781	25.12
Graduate/Diploma	0.821	25.43	1.123	34.28	1.301	57.16	1.330	60.12	1.259	44.17	1.208	38.1
Postgraduate	1.186	36.85	1.491	40.35	1.621	61.51	1.605	58.96	1.472	37.79	1.499	41.9
ST	-0.144	-4.07	-0.076	-2.05	-0.099	-3.71	-0.085	-3.34	-0.059	-1.88	-0.065	-1.91
SC	-0.112	-5.31	-0.137	-5.88	-0.085	-5.38	-0.077	-5	-0.062	-2.93	-0.092	-4.42
OBC	-0.142	-7.21	-0.107	-5.42	-0.110	-8.16	-0.111	-8.52	-0.083	-4.44	-0.101	-5.21
State Dummies	Yes											
N	35,337											

Table 7: Quantile regression for the PJC specification (Pooled sample), 2009-2010

DepVar : ln(Daily wage)	Percentiles											
Independent Variables	10 th		30 th		50 th		70 th		90 th		OLS	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Constant	2.352	25.26	3.139	48.46	3.507	60.77	3.849	75.01	4.339	72	3.415	39.31
Age	0.039	7.52	0.023	6.62	0.026	8.56	0.025	9.64	0.023	7.65	0.026	5.68
Age Square/100	-0.040	-6.19	-0.014	-3.25	-0.017	-4.45	-0.014	-4.28	-0.011	-2.79	-0.017	-2.93
Gender	0.563	35.45	0.426	31.97	0.341	27.67	0.275	24.24	0.210	14.62	0.368	17.09
Sector	0.300	15.69	0.255	21.64	0.222	22.4	0.199	23.33	0.178	18.28	0.256	12.92
Marital Status	0.139	7.1	0.123	9.26	0.072	6.02	0.015	1.39	0.028	2.1	0.079	4.04
ST	-0.114	-4.31	-0.128	-5.71	-0.118	-5.95	-0.076	-4.07	-0.128	-5.96	-0.099	-3.2
SC	-0.082	-4.74	-0.125	-9.17	-0.123	-9.97	-0.134	-12.24	-0.109	-8.19	-0.124	-6.59
OBC	-0.089	-5.41	-0.089	-7.62	-0.091	-8.67	-0.093	-9.71	-0.103	-9.7	-0.097	-5.54
Primary & Below	0.149	5.8	0.071	3.4	0.048	2.5	0.063	3.66	0.072	3.58	0.077	2.79
Middle	0.222	8.56	0.202	9.6	0.144	7.38	0.156	9.03	0.174	8.81	0.183	6.66
Secondary	0.323	12.18	0.302	14.64	0.257	13.38	0.268	15.7	0.325	16.06	0.301	10.56
Higher Secondary	0.415	12.81	0.384	16.81	0.306	14.51	0.384	20.34	0.448	19.35	0.403	12.59
Grad Diploma	0.589	16.22	0.639	26.69	0.576	26.9	0.618	31.76	0.683	29.49	0.651	16.19
Postgraduate	0.792	20.18	0.808	29.95	0.750	30.22	0.792	34.13	0.810	25.09	0.837	16.49
Job Characteristics (General)												
Public	0.284	12.23	0.359	23.96	0.435	33.22	0.373	31.1	0.271	18.05	0.305	11.67
Union	0.260	17.73	0.260	23.1	0.198	19.52	0.163	17.57	0.128	11.18	0.219	13.83
Contract	0.147	10.59	0.175	16.69	0.165	16.42	0.179	18.87	0.161	13.23	0.173	10.42
Permanent	0.189	11.8	0.180	15.24	0.206	20.09	0.238	26.35	0.266	25.05	0.228	14.34
Job Characteristics (Occupation)												
Manager/Legislators	0.497	11.88	0.579	21.18	0.632	24.64	0.639	26.58	0.647	19.36	0.556	11.78
Professionals	0.437	12.64	0.525	22.35	0.597	27.83	0.584	30.87	0.582	26.26	0.514	14.54

Table 7 Continued...

DepVar : ln(Daily wage)	Percentiles											
Independent Variables	10 th		30 th		50 th		70 th		90 th		OLS	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Technicians/Assoc. Profes.	0.221	7.63	0.284	13.59	0.328	17.21	0.362	20.6	0.355	17.68	0.294	9.64
Clerical	0.158	5.73	0.148	7.4	0.157	8.55	0.146	8.79	0.138	7.3	0.138	5.14
Service/Sales	-0.001	-0.03	-0.007	-0.36	0.083	4.88	0.117	7.43	0.136	6.73	0.052	1.84
Crafts and other trades	0.064	2.69	0.074	3.85	0.103	6	0.098	6.29	0.083	4.67	0.076	3.14
Plant & Machine operator	0.108	3.3	0.148	7.22	0.132	7.26	0.127	7.83	0.117	5.94	0.093	2.75
Job Characteristics (Industry)												
Agriculture	-0.072	-0.88	0.018	0.33	-0.008	-0.16	-0.022	-0.51	-0.009	-0.14	0.001	0.01
Elec/Gas/Water	0.120	3.64	0.101	3.49	0.152	5.48	0.139	4.63	0.110	4.69	0.136	3.21
Retail/ Hotels	-0.110	-4.53	-0.041	-2.37	-0.121	-7.56	-0.152	-10.54	-0.180	-9.97	-0.119	-4.91
Transport Communication	0.063	2.01	0.024	1.22	0.026	1.53	0.021	1.44	0.064	3.29	0.064	1.92
Financial Intermediation	0.119	3.51	0.089	3.61	0.098	4.22	0.107	5	0.098	3.39	0.113	3.03
Real estate	0.117	3.47	0.177	7.15	0.230	9.87	0.304	14.16	0.287	11.74	0.235	6.27
Public ad. / Defense	-0.019	-0.53	0.006	0.28	-0.043	-2.28	-0.024	-1.43	-0.018	-0.91	-0.006	-0.15
Education	-0.356	-11.14	-0.306	-14.73	-0.267	-14.49	-0.224	-13.53	-0.168	-8.04	-0.269	-7.39
Health, Social work	-0.158	-4.54	-0.173	-6.36	-0.195	-8.18	-0.200	-9.14	-0.089	-3.18	-0.173	-4.15
Other Services	-0.204	-7.24	-0.157	-6.98	-0.218	-9.75	-0.259	-12.25	-0.190	-6.48	-0.192	-4.88
State Dummies	Yes											
N	33,720											

Table 8: The MMM decomposition results, PJC Regression (all figures expressed in log daily wage), 2009-2010

Deciles	Men	Women	Diff.	Returns of Women			Returns of Men		
				CF ₂	Char.	Coeff.	CF ₁	Char.	Coeff.
1	4.368	3.662	0.706	3.769	0.107	0.599	4.329	0.039	0.668
2	4.697	4.081	0.617	4.2	0.12	0.497	4.676	0.021	0.595
3	4.947	4.41	0.538	4.523	0.114	0.424	4.943	0.004	0.533
4	5.174	4.72	0.453	4.815	0.095	0.358	5.191	-0.018	0.471
5	5.401	5.018	0.383	5.097	0.079	0.304	5.432	-0.03	0.414
6	5.641	5.333	0.309	5.388	0.055	0.253	5.678	-0.037	0.345
7	5.901	5.659	0.242	5.7	0.041	0.201	5.936	-0.035	0.278
8	6.195	6.021	0.174	6.047	0.026	0.147	6.232	-0.038	0.211
9	6.575	6.473	0.102	6.483	0.011	0.091	6.604	-0.03	0.131

8.2 Figures

Figure 1: Gender log wage gap across quantiles and the average gap, 2009-2010

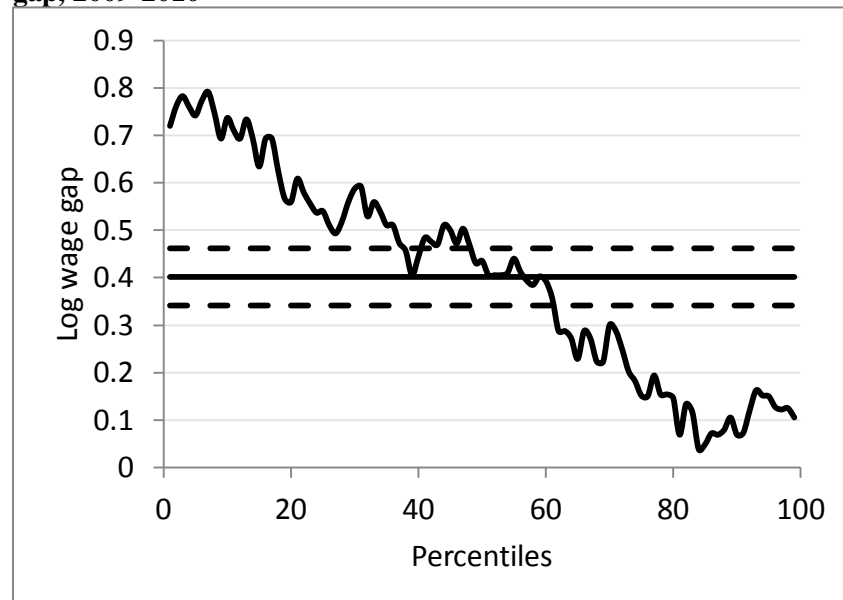


Figure 2: Gender (log) wage gap across quantiles, 1999-2000 to 2009-2010

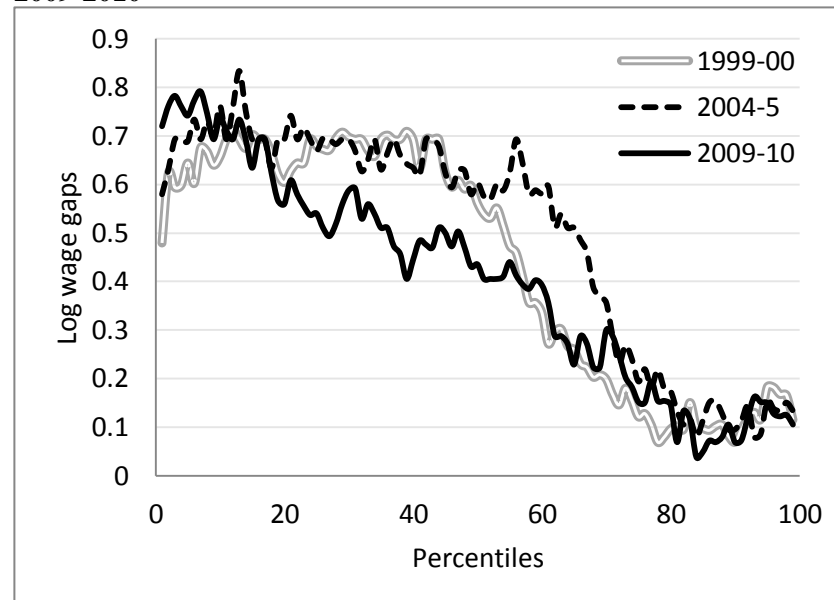


Figure 3: Gender log wage gap across quantiles and OLS log wage gap, by sector, 2009-2010

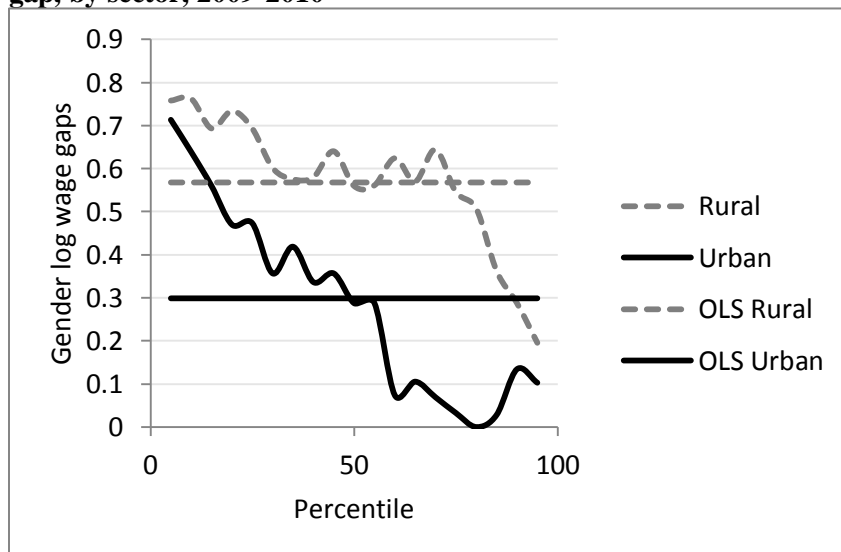


Figure 4: Gender log wage gaps across quantiles for the public and private sectors, 2009-2010

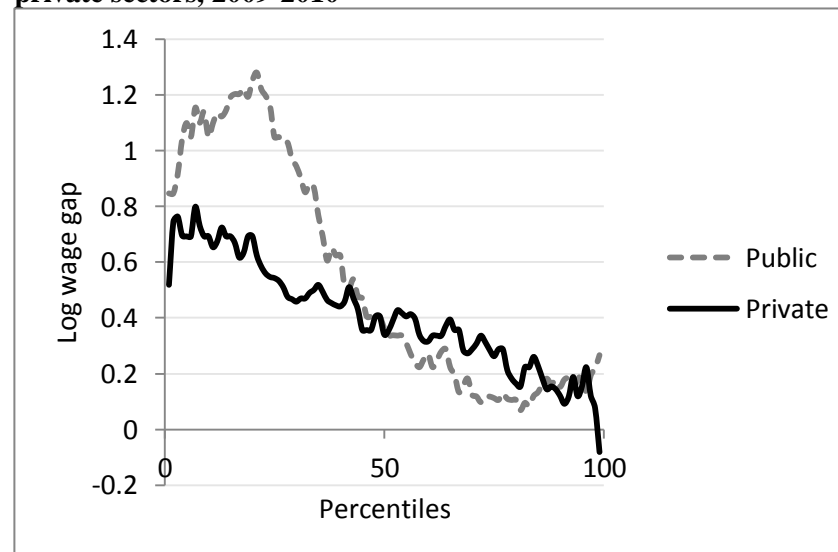


Figure 5: Gender log wage gaps by deciles, illiterates and postgraduates, 2009-2010

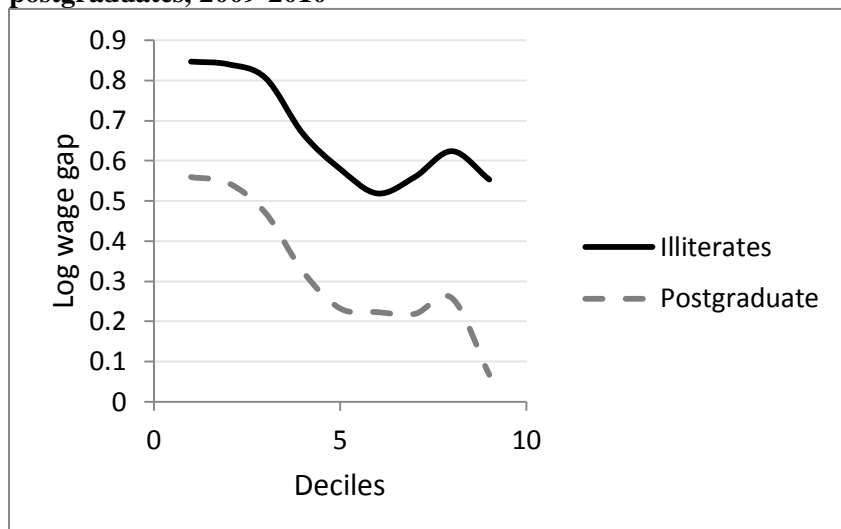


Figure 6: Gender log wage gaps for different age cohorts, 2009-2010

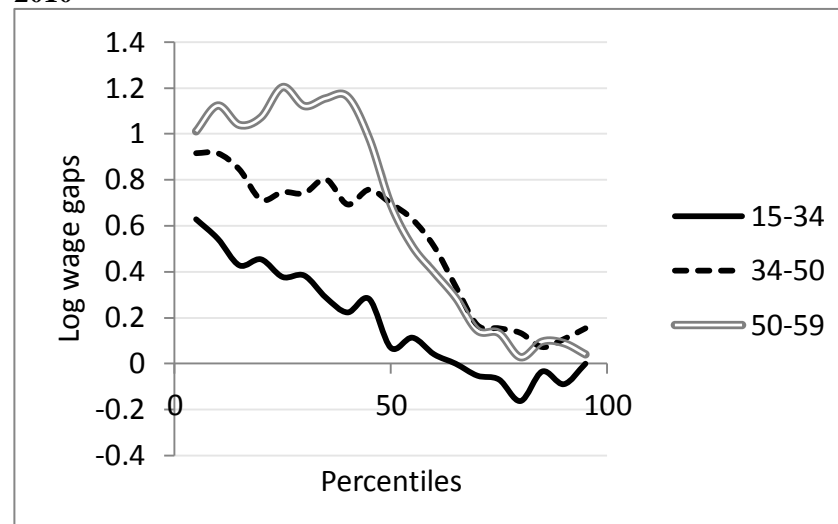


Figure 7: Gender wage gap, no controls and all controls (PJC), 2009-2010

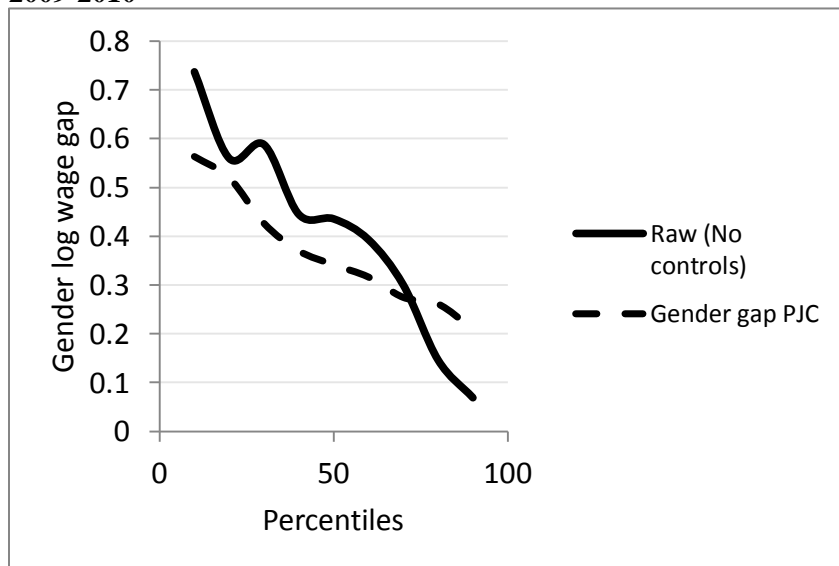


Figure 8: Coefficients of Union dummy (QR and OLS) PJC regression, pooled sample, 2009-2010

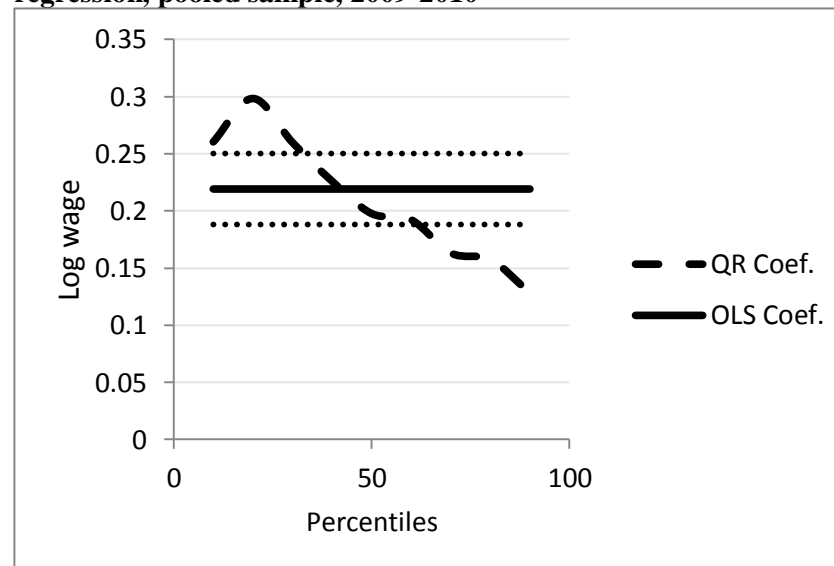


Figure 9: Coefficients of the marriage dummy (PJC regression), by gender, 2009-2010

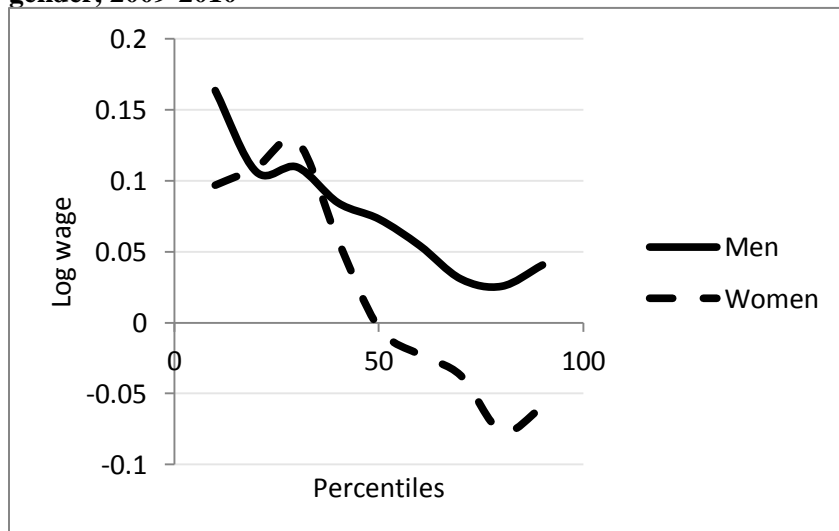


Figure 10: Coefficients of the union dummy (PJC regression), by gender, 2009-2010

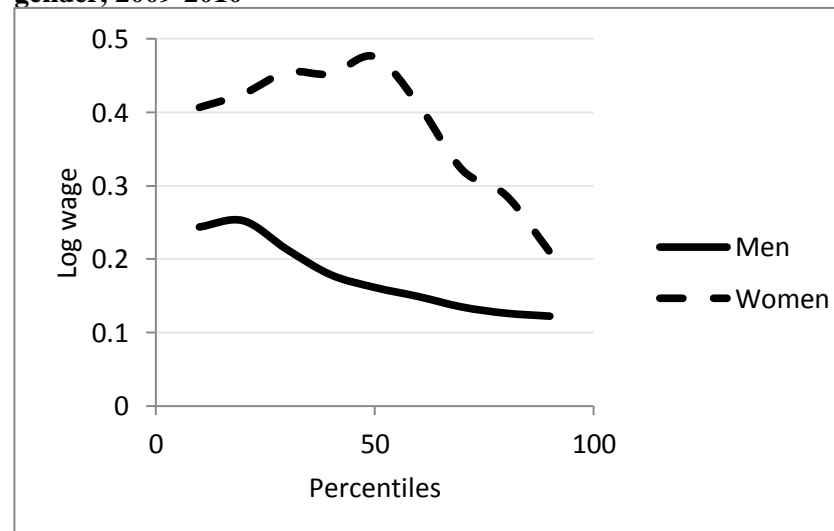


Figure 11: Coefficients of the graduate dummy (PJC regression), by gender, 2009-2010

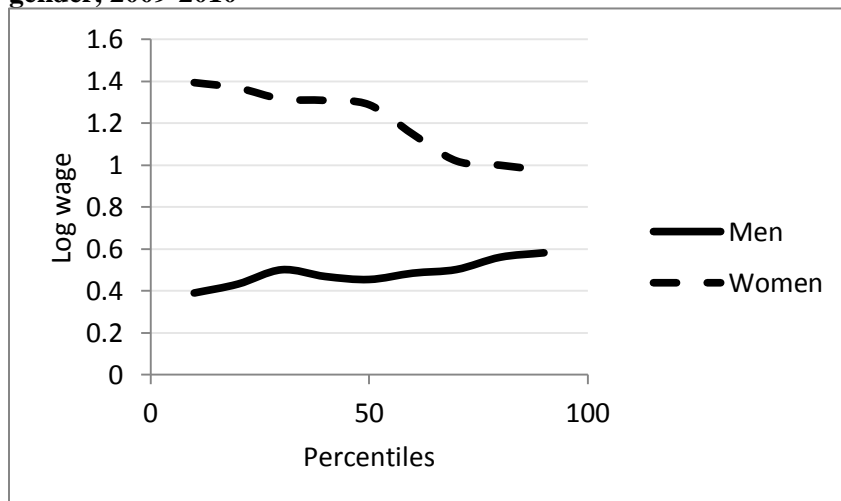


Figure 12: High paying Occupation dummy QR coefficients, 2009-2010

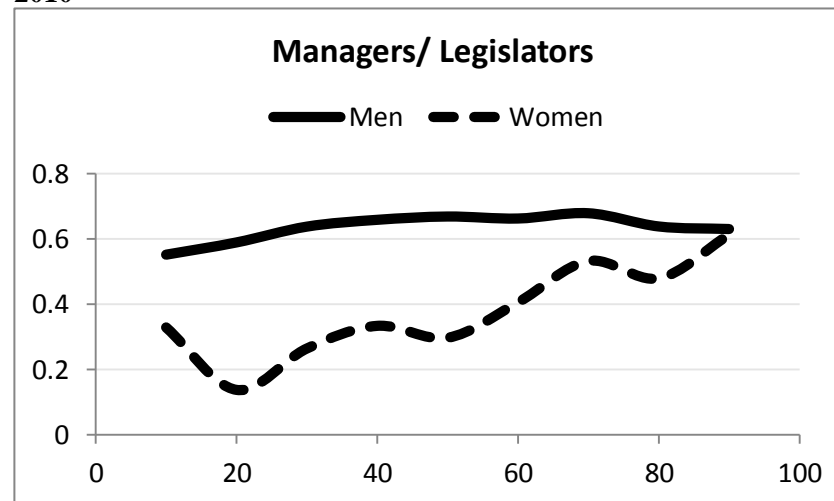


Figure 13: The MMM decomposition results, counterfactual using returns of women, 2009-2010

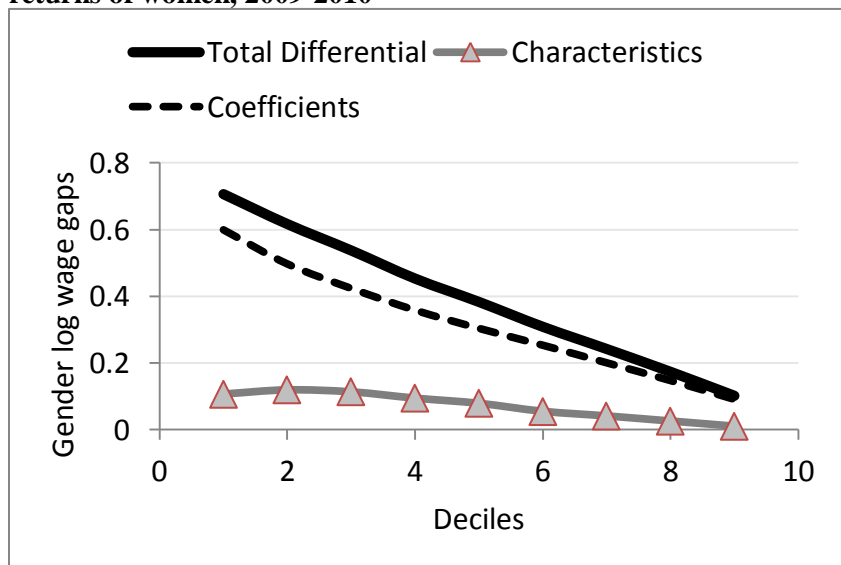


Figure 14: The sticky floor and the minimum wage, 2009-2010

