Formal But Less Equal.
Gender Wage Gaps in Formal and
Informal Jobs in Brazil

Sarra Ben Yahmed

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Formal But Less Equal.
Gender Wage Gaps in Formal and
Informal Jobs in Brazil

Sarra Ben Yahmed

Formal but less equal.
Gender wage gaps in formal and informal jobs in Brazil.

Sarra Ben Yahmed*

Centre for European Economic Research (ZEW)

December 2016

Abstract

In developing countries, a large share of employees work informally and are not covered by employment protection legislation. I study here how gender wage inequality differs across formal and informal jobs in Brazil. The raw gender wage gap is higher in informal jobs (13%) compared to formal jobs (5%), but I show that this difference is an artefact of different male and female selection processes. First, women have better observable characteristics than men and the female advantage is stronger among formal employees. Second, men and women entering formal and informal jobs have different unobservable characteristics. Controlling for endogenous selection into formal vs. informal jobs, I find that the gender gap in wage offers is high and increases with education in formal jobs. In informal jobs, however, estimated wage offers are the same for men and women. I discuss the potential implications of these findings regarding the effect of labour market regulation on gender wage gaps.

Keywords: Gender wage gaps, informality, selection into work statuses, Brazil.
JEL codes: J31, O17

1 Introduction

A striking characteristic of labour markets in developing countries is the existence of a large informal sector where labour regulations are inexistent. Labour regulations protect workers against several risks, for example against unexpected job loss by defining the conditions under which firms are allowed to lay off employees. Men and women do not depend equally on labour regulations as, for example, women need to stop working around childbirth. Parental leave and benefits, and job protection during the leave, are entitled to employees as long as their contract has been officially registered; they are almost exclusively used by women. This asymmetry affects how men and women value formal jobs. It also influences the incidence of statistical discrimination against women.[1] While labour protection and labour costs are lower in the informal segment of the labour market, informal jobs may offer other features valuable to workers such as greater flexibility. Those aspects are likely to influence female and male wages differently. It is thus important to understand how men and women sort across types of jobs and to analyse how the gender wage gaps in formal and informal jobs differ. Moreover, it is useful for policy makers to learn about the gender wage differences in formal and informal jobs separately as it sheds some light on how regulation affects women's prospects relative to men's in the labour market.

The aim of the paper is to examine whether there exist significant gender wage gaps in both formal and informal jobs and whether they are different. I do not attempt to estimate the causal effect of labour market legislation on employers' discrimination against women, but I do show that the gender gap in potential wage offers is much higher in formal jobs and discuss the scenarios that are consistent with this finding. To recover the gender gaps in potential wage offers I control for both observable characteristics of workers and non-random selection into formal and informal jobs. Non-random selection into jobs means that, even after controlling for observable characteristics, formal workers are systematically different from informal workers in a way that affect their wages.[2] Failing to deal with non-random selection into formal and informal jobs is a major concern as it would lead to misleading estimates of gender wage gaps for two reasons. First, selection based on unobservables would bias the coefficient estimates in the wage equation. Second, if selection is not random the observed wage gap does not reflect the gap in wage *offers*. It is thus important to recover the differences in wage offers to conduct the decomposition on the appropriate total wage difference. I am then able to show what share of the total gender wage gap remains truly unexplained by workers' productive characteristics. To do so, I first study the sorting of men and women into different employment statuses using a multinomial logit model. It is here relevant to describe the potential work statuses with more than two outcomes, informal and formal employment, as inactivity and unemployment are common situations, especially among women. I then investigate how selection into work status affects the estimation of the gender wage gaps using the control function approach. Wage equations are estimated for formal employees

---

[1]Discrimination implies that two individuals who are equally productive on the job are treated unequally because they belong to different groups, defined by gender here. Statistical discrimination suggests that employers have imperfect information on the behaviour of their employees, regarding for example labour market attachment as defined by quit probability or leave duration. If employment discontinuity and job protection generate higher labour costs for the firm because of re-organization of personnel, vacancy and replacement costs, employers have incentive to reduce expected costs and treat unequally men and women because women take *on average* more or longer leave.

[2]This is also referred to as 'selection based on unobservables'.

and informal employees separately. In the last step, I use the Oaxaca-Blinder-Ransom decomposition to compute the gender wage gaps in both types of jobs.

The empirical analysis uses the Brazilian household survey, the Pesquisa Nacional por Amostra de Domicilios (PNAD), for the year 2009. In Brazil, all employees have a labour card (*carteira de trabalho*) which is a booklet where their employment biography is recorded. In this document employers provide some of the contract conditions like a short job description, the remuneration and the starting date. By signing it, both the employer and the employee commit to abide by labour regulations. The PNAD provides information on whether the worker's labour card is signed by the employer so that I am able to adopt a definition of informality based on employers' compliance with labour market regulations. A formal worker is an employee with a registered labour contract, hence entitled to labour rights and benefits, while an informal worker is employed without having a legal contract declared by his/her employer.

Looking at the raw data, I find that women are more often unemployed than men and that the informality rate is higher among working women compared to working men. The raw gender wage gap is higher among informal workers and increases with education in both segments of the labour market. This pattern is in line with recent evidence on the heterogeneity of informal labour markets (Gunther and Launov, 2012) and points to different labour market selection processes across formal and informal sectors.

Controlling for observable characteristics, I find no significant difference between the gender wage gap among formal workers and the gender wage gap among informal workers. This result would indicate that discrimination and/or compensating differentials, for protection in formal jobs and for flexibility in informal jobs, have the same impact on the average female wage relative to the average male wage in formal and informal jobs. These result and interpretation however ignore the non-random selection of men and women into a type of job because of preferences or characteristics that cannot be directly controlled for.

I then proceed with controlling for selection based on unobservable characteristics. I show that men and women differ in their selection in formal and informal jobs. Controlling for selection into work status, I find that the gender gap in potential wage *offers* is no longer significant in the informal segment of the labour market but remains significantly positive among formal employees.

Because labour market decisions and observed gender wage gaps differ across the schooling distribution, I conduct the analysis for three different education groups. The gender gap in potential wage offers is the highest among highly educated people in formal jobs but it remains insignificant across all educational level in informal jobs. These findings are consistent with greater incidence of statistical discrimination against women and glass ceiling effect in formal jobs because employment protection is more binding and costly for employers with female employees. A positive wage gap in formal jobs is also consistent with compensating wage differentials for job protection when protection is more valued by women.

This paper contributes to a rather small strand of literature that studies labour market outcomes for men and women when a large share of employment is informal. My approach is most closely related to the papers that study the gender wage gap among informal workers and formal workers separately.

Pagán and Ullibarri (2000) use the Jenkins index to compute gender wage gaps for different groups of workers in Mexico. They find that the part of the gender wage gap that cannot be explained by observable characteristics is larger in unregistered firms compared to registered firms. Tansel (2001) defines informality as the absence of social security protection and estimates the gender wage gaps among covered and uncovered wage earners in Turkey. She finds that, in 1994, the adjusted wage gap is strong and positive among covered workers but not significant among uncovered workers. Deininger et al. (2013) look at the gender wage gap in India in casual work in either agricultural or non-agricultural sectors. They control for selection into labour market participation using the Heckman's (1979) methodology. They find that the gender wage gap due to different returns to characteristics is particularly important for casual workers working in the agriculture; however, gender discrimination is much lower or inexistent in non-agricultural sectors. I depart from these papers in two ways. First, in the definition of informality as I focus on employers' compliance with labour regulation rather than on social security coverage or temporary work. Having information on the worker's contract rather than on firm registration enables me to consider informal workers working in both registered and unregistered firms. Second, in the empirical methodology, I compare two approaches to deal with non-random selection into multiple employment outcomes and estimate the wage function and selection bias for different education groups.

This paper is also related to the vast literature on the segmentation of the labour market and the formal wage premium. Several authors question the dualistic view of the labour market and find that some workers are actually better off choosing the informal segment of the labour market (Magnac, 1991; Maloney, 1999; Carneiro and Henley, 2002; Gunther and Launov, 2012; Pradhan and Van Soest, 1995). Interview data for Brazil reported in Maloney (2004) suggests that about 30% of both male and female informal employees are voluntarily working informally. However, the motives behind this choice differ across men and women. Out of the 30% of women preferring informal employment, about 13% say the reason is competing household chores. Almost no men chose to work informally to balance paid employment and household production. More than 10% of men say that they do not wish a formal job because they earn more in their current informal job, but less than 4% of women say so. This suggests that there might be different selection rules into informal vs formal jobs for men and women. I attempt to uncover these selection rules and to present estimates of the gender wage gaps adjusted for the selection biases.

The present paper also contributes to the literature on gender wage gaps and selection into employment. Arabsheibani et al. (2003) study the gender wage differentials in Brazil over the period 1988-1998. They find that the gender wage gap, especially the part due to different returns of identical characteristics, has fallen over the period but remains positive. Madalozzo (2010) confirms the fall in the wage gap until the end of the 90s for Brazil and finds no further decrease in the 2000s. Santos and Ribeiro (2006) find evidence of a glass ceiling in Brazil. These three papers study the gender wage gap in Brazil but they do not distinguish between formal and informal employees, nor do they investigate the impact of different selection biases between men and women on the gender wage gap.

A vast literature, starting in the late 1970s, has studied the effect of selection bias on the gender wage gap, mostly focusing on the United States. Among recent papers, Blau and Kahn (2006) show that the decline in the gender wage gap during the 1980s was overstated, as it is largely explained by sample

selection. They also show that selection has contributed to the slower reduction in the gender wage gap during the 1990s. Looking at European countries and the United States, Olivetti and Petrongolo (2008) point out that non-random selection explains why gender employment gaps are negatively correlated with gender wage gaps across countries. Women are on average positively selected into employment. Countries with particularly high gender employment gaps, as in southern Europe, are characterized by a strong positive female selection which in turn reduces the observed gender wage gap. This small observed gender wage gap is actually an artefact of the selection process: women who are employed have better abilities than non-employed women, which overestimates female wage offers. These studies find that correction for selection has important consequences for the assessment of gender wage gaps. In this paper, I assess the effect of selection bias on gender wage gap estimations in a labour market where the co-existence of formal and informal jobs modifies the selection process.

The present paper is also linked to the empirical research on the heterogeneity of the wage gaps across groups with different skill levels. Albrecht et al. (2003) show that the gender wage gap is increasing along the wage distribution in Sweden. For Spain, de la Rica et al. (2008) find that the gender wage gap is high and increases with the wage (glass-ceiling effect) among highly educated workers while it is lower and decreases with wage among less educated workers (floor effect). The innovation of this paper is to explore how the wage gap differs by education groups for informal wage-earners and formal wage-earners separately.

The remainder of the paper is organized as follows. I start by discussing the impact of informality on gender employment and wage inequality while reviewing the related literature. In section 3 I describe the data and provide descriptive statistics on gender inequalities in the Brazilian labour market. Section 4 sets up the empirical model. In section 5 I discuss the results, looking at the selection into potential outcomes for men and women before moving on to the comparison of the gender wage gaps in the formal and informal sectors. The last section concludes.

2 Why should informality matter?

Before considering the empirical evidence, I first ask why the gender wage gap should differ across the formal and the informal segments of the labour market. The existing theoretical explanations have focused on understanding the formal wage premium, but they have not provided any explanation for gender differences in the formal wage premium nor have they explained gender wage differences *within* each sector. I use the existing literature to postulate hypotheses about the mechanisms that can explain why the formal gender wage gap differs from the informal gender wage gap.

Individual characteristics across types of jobs

According to the dualistic view of the labour market, the informal segment is characterized by lower wages. Empirical evidence confirms that formal jobs offer on average higher wages than informal jobs.[3] Comparisons of raw average wage gaps are informative about the *accepted* wage offers but

---

[3]See Magnac (1991) who analyses female wages in Colombia, Gasparini and Tornarolli (2009) who focus on different Latin American countries and Almeida and Carneiro (2007) who find that the formal raw wage premium is positive in

conceal heterogeneity in workers' observable and unobservable attributes. Empirical papers show that the formal-informal wage difference varies with workers' skill levels. Studying the urban labour market in Mexico, Gong and Van Soest (2002) find a significant wage premium in formal jobs for educated men but not for men with low education who earn more on average in informal jobs. For women, the differences between formal and informal wages is small. It is thus important to take into account men's and women's productive characteristics, such as education levels and compare individuals with similar observable productivity within the two segments. Moreover, men and women may also sort into the different types of jobs because of unobservable characteristics that may correlate with wages. To infer how informality affects gender gaps in potential wage offers, one needs to account for this endogenous sorting of men and women into formal and informal jobs.

Compensating wage differentials

What should explain different gender wage gaps in formal and informal jobs once we have controlled for individual characteristics? From the labour supply side, if individuals have different preferences for the type of jobs, the theory of compensating wage differentials can give an explanation for formal-informal wage differences within groups and it can also help understand the gender wage gaps in formal jobs and in informal jobs. Formal jobs offer non-wage benefits that are not available in informal jobs such as job severance contribution, maternity leave, unemployment benefits, and social security. In a frictionless market, workers with identical productivity should earn a higher wage in the informal segment to compensate for the absence of these benefits. If women value job protection more than men, for reasons linked to maternity for example, women should be ready to accept lower wages compared to men in the formal sector but not in the informal sector. This would lead to a gender wage gap among formal employees only. However, if women value the flexibility of informal jobs more than men, to better combine paid work and household production for example, one should also observe a gender wage gap among informal employees. There are reasons to value the amenities of both sectors and workers' preferences over formal or informal employment hinge on the balance of the advantages and disadvantages of both statuses. Gender differences in preferences are not a priori clear cut, which makes it difficult to draw theoretical predictions on the overall effect of preferences on gender wage gaps across sectors.

Labour regulation and statistical discrimination

From the labour demand side, job offers stem from both registered and unregistered firms. While firms operating informally cannot offer legal contracts to their employees, firms operating formally might decide to hire workers formally or informally. Why would employers set different wages to a man and a woman with similar observable characteristics and employed under the same type of contract? Employers compare the costs and benefits of labour contract registration for both men and women and set their hiring decisions and wage setting rules accordingly. Under imperfect information on the behaviour of their employees, regarding for example labour market attachment as defined by quit probability or leave duration, employers have incentive to use information on usual behaviour in the

---

Brazil and decreases with regulation enforcement.

demographic group to which the employees belongs to. In other words, employers have incentive to treat a man and a woman differently because women have *on average* lower labour market attachment, even if this man and this woman are equally productive on the job. Employers may expect a higher quit rate among women because of, for example, permanent or temporary leave due to maternity. Lazear and Rosen (1990) provide a theoretical explanation where stronger domestic responsibilities generate higher female quit rates and lower female wages due to statistical discrimination.[4] Employment discontinuity generates higher labour costs for the firm because of vacancy and replacement costs; it can also generate forgone profits if no one can replace the employee on leave or if the time out of the job causes a loss of (general or specific) skills. Employers may want to compensate for the higher female quit rate by paying them lower wages. This argument applies especially to formal jobs where employers abide by labour regulations such as the protection of the job during maternity leave. Put differently, statistical discrimination against women would lower female wages in formal jobs where employers' costs associated with their employees' job discontinuity are larger. The impact of labour regulation on gender wage gaps has been in part addressed in papers that investigate the effects of family policies on gender outcomes. Ruhm (1998) exploits variation in maternity leave coverage across European countries between 1969 and 1993. He concludes that more generous leave policies are associated with increased employment but with lower wages of women relative to men. Gruber (1994) uses variation in the timing of a reform in health insurance laws across states of the US. The reform increased employers' relative costs of insuring women of childbearing age and his findings indicate that much of these costs were shifted to female wages. Using data for several developed economies, Mandel and Semyonov (2005) find that longer maternity leaves corresponds to wider gender gaps.

In the case of Brazil, women in formal jobs are entitled to 120 days of paid maternity leave while men are entitled to up to 5 days of paid paternity leave. Moreover, job protection implies that women cannot be made redundant in the 5 months following delivery. In informal jobs, women do not benefit from these rights. In this context, the expectation is that the gender wage gap is bigger in formal jobs because of statistical discrimination and/or unequal compensating wage differentials. I am not aware, however, of any study that is able to exploit a labour or familly policy reform in a context of dual labour market where a significant share of the workforce is not covered by mandated benefits. In the absence of exogenous change in regulation, this paper does not provide any direct evidence on the effect of employment protection legislation on the wages of women relative to men. Instead I compare the gender wage gaps across jobs with and without mandated benefits and employment protection, and discuss how the findings are consistent with the arguments discussed above. The next section presents the econometric methods used to explore whether the gender wage gap is higher in the formal segment of the labour market.

---

[4] See also Bertrand et al. (2010)

3 The Econometric model

To compare the gender wage gaps among formal and informal employees, I investigate how selection shapes the gender wage gaps in these two different segments of the labour market. I first compute the raw wage gaps and the wage gaps adjusted for observable characteristics in both segments. By comparing the raw and the adjusted wage gaps, one can see how gender differences in observable characteristics shape gender wage inequality. Next, I compute the wage gaps controlling for both observable characteristics and the selection into the different labour statuses.

3.1 The raw and the adjusted wage gaps

The raw wage gap in sector j is estimated from an equation where $\ln w_{ij}$ the hourly log wage is regressed on a constant and a female dummy only:

$$\ln w_{ij} = \beta_0 + \alpha_j F_{ij} + u_{ij}, \tag{1}$$

where $F_{ij} = 1$ if employee i working in sector j is a woman. The raw wage gap in sector j is:
$E(\ln w_j | female) - E(\ln w_j | male) = \widehat{\alpha}_j$.

Different methods are used in the literature to compute the adjusted wage gap. One method is to estimate a mincerian wage equation on a pooled sample with a female dummy to capture the gender wage gap. The problem with this method is twofold. First, it might suffer from misspecification if the differences in returns to specific characteristics matter for the estimation of the wage gap. Second, I cannot estimate the selection rule for men and women separately using one wage equation on a pooled sample.

Instead, I use a version of the wage gap decomposition developed by Oaxaca (1973) and Blinder (1973) that avoids important methodological problems discussed in Oaxaca and Ransom (1994) and Oaxaca and Ransom (1999). The decomposition methodology that follows has been presented in Fortin (2008) and is not sensitive to the choice of the reference wage structure. The reference wage structure is taken from the estimation of a common wage regression on the pooled sample of both men and women where the male advantage equals the female disadvantage with respect to the reference. I estimate three equations, two separate wage equations for men and women and a pooled wage equation with gender dummies and an identification restriction. Each equation is estimated separately for the formal and the informal segments denoted with the subscript $j = 2, 3$.

$$\ln w_{ipj} = \beta_{0pj} + \alpha_{pfj} F_i + \alpha_{pmj} M_i + \mathbf{X_i} \beta_{pj} + u_{ij} \qquad \text{with } \alpha_{pfj} = -\alpha_{pmj}, \tag{2a}$$

$$\ln w_{ifj} = \beta_{0fj} + \mathbf{X_i} \beta_{fj} + u_{ifj}, \tag{2b}$$

$$\ln w_{imj} = \beta_{0mj} + \mathbf{X_i} \beta_{mj} + u_{imj}, \tag{2c}$$

where X is a set of control variables that includes the number of years of education, the age and the age squared, the tenure in the current job and its square, whether the person is black, whether the

person lives in an urban area, and indicators for regions and sectors of activity. To capture demand side effects, I use regional unemployment rates that characterize the state of the local labour market. I construct the regional unemployment for different education groups in order to identify the impact of lower labour demand even when controlling for regional dummies. The assumption is that labour markets are skill-specific, at least to some extent. Even if workers may accept a job for which they are overqualified, the unemployment rate among people of the same (generally defined) skill level will impact their decision to participate, their job finding rate and their wages.

The zero conditional mean assumption $E(u_m|x_m) = E(u_f|x_f) = 0$ ensures that the error is uncorrelated with the regressors so that the OLS estimates are unbiased. The zero conditional mean assumption also ensures that the total average wage gap can be exactly decomposed into terms based on observables and their returns. For the wage decomposition to be exact though, only a weaker ignorability assumption is sufficient; what is needed is that the distribution of u given X is the same for the two groups. In other terms, the decomposition allows for selection on unobservables as long as they are the same for both men and women and yield identical selection biases. See Fortin et al. (2011) for a discussion of the assumptions required for identification in wage decompositions. Under the ignorability assumption, the total wage gap in each segment can be decomposed into three terms:

$$\overline{\ln W}_{mj} - \overline{\ln W}_{fj} = (\overline{\mathbf{X}}'_m - \overline{\mathbf{X}}'_f)\widehat{\beta_{pj}} + \overline{\mathbf{X}}'_m(\widehat{\beta_{mj}} - \widehat{\beta_{pj}}) + \overline{\mathbf{X}}'_f(\widehat{\beta_{pj}} - \widehat{\beta_{fj}}).$$

The first term on the right hand side accounts for gender differences in characteristics, it is the endowment term. The rest accounts for gender differences in the prices associated with given characteristics, it is also called the coefficient term and is here decomposed into the male advantage with respect to the reference prices and the female disadvantage with respect to the reference prices. The adjusted wage gap is the sum of the male advantage and the female disadvantage in the treatment of the characteristics :

$$\overline{WG}_j = \overline{\mathbf{X}}'_m(\widehat{\beta_{mj}} - \widehat{\beta_{pj}}) + \overline{\mathbf{X}}'_f(\widehat{\beta_{pj}} - \widehat{\beta_{fj}}). \tag{3}$$

The adjusted wage gap takes into account the observable differences in characteristics between men and women, however it does not account for the selection of men and women into formal or informal jobs because of unobserved characteristics.

This can be problematic given that the conditional independence assumption is strong and that even the ignorability assumption may not hold in the case studied here. Women have a much lower labour market participation rate than men and the selection of men and women into different types of jobs is certainly not random. What is more, selection into employment may follow different processes for men and women. The descriptive statistics (see below) show that the female unemployment rate is higher than the male unemployment rate and that the informality rate is higher among active women compared to active men. If $E(u|X) \neq 0$ in equation (2), the coefficients of the wage equation are biased. If the ignorability assumption does not hold, men employed in a given type of job are different in observables and in unobservables from women who are employed in the same type of job. In that case, the selection biases differ for men and women and the estimations of the wage gaps are thus biased

8

too. To eliminate the selection biases I adopt a control function approach that is presented in the next sub-section.

3.2 Treatment for selection into multiple employment statuses

Labour market outcome

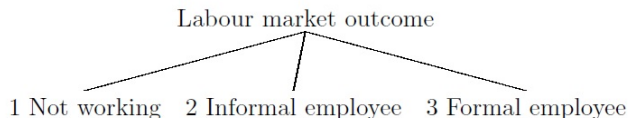1 Not working   2 Informal employee   3 Formal employee

Figure 1: Work status, simultaneous decisions

I use a multinomial model to estimate the probability to be in formal employment and in informal employment taking into account inactivity and unemployment as an alternative status (Figure 1). In this setting, individuals have different probabilities to be in a given work status depending on their characteristics and preferences, as well as on demand constraints and employers' behaviours that may cause job rationing and segregation.

To model individuals' choices (constrained or unconstrained), I consider here a one-step process where selection into inactivity or unemployment and the two employment categories happens simultaneously. The model has three mutually-exclusive outcomes denoted j: inactivity or unemployment ($Y_i = 1$), informal employment ($Y_i = 2$), and formal employment ($Y_i = 3$).

The latent value (or utility) associated with being in state j is denoted V_{ij}. State j is observed $Y_i = j$ if the value associated with this state is higher than the value of the other states, or in other words, if status j is the best available option for individual i:

$$Y_i = j \text{ if } V_{ij} > \max_{k \neq j}(V_{ik}).$$

I assume that the utility associated with work status j follows a linear function: $V_{ij} = \mathbf{X}_{1i}\lambda_j + \mathbf{Z}_i\alpha_j + \mu_{ij}$, for $j = 1, ..., 3$. If I further assume that the errors are independent and identically distributed following a type I extreme value distribution, the probability of being in status j for individual i is defined by the multinomial logit model (McFadden, 1973) as in equation 4.[5]

$$P_{ij} = Pr(Y_i = j) = \frac{exp(\mathbf{X}_{1i}\lambda_j + \mathbf{Z}_i\alpha_j)}{\sum_j^N exp(\mathbf{Z}_i\alpha_j)}. \tag{4}$$

---

[5]The independence from irrelevant alternatives is a strong assumption of the logit model. It implies that the odds ratio between two alternatives are independent of other alternatives. In the case under study here, as an illustration, it implies that should all non-working women start working, the repartition of women in formal and informal jobs would remain unchanged. Although not completely implausible, this assumption is clearly restrictive. Bourguignon et al. (2007) evaluate selection bias correction techniques based on the multinomial logit model and show with Monte Carlo experiments that those methods provide good bias correction in the outcome equation even when the IIA hypothesis is violated.

The full model of selection and wage determination can be written as follows

$$\ln w_{ij} = \mathbf{X}_{1ij}\beta_{1j} + \mathbf{X}_{2ij}\beta_{2j} + u_{ij}, \text{ if } V_{ij} > \max_{k \neq j}(V_{ik}) \text{ for } j = 2, 3 \quad ,$$

$$V_{ij} = \mathbf{X}_{1i}\lambda_j + \mathbf{Z}_i\alpha_j + \mu_{ij}, \qquad j = 1, ..., 3 \quad ,$$

where individual i earns a wage w_{ij} if she is a formal worker $j = 2$ or an informal worker $j = 3$ and j is the observed outcome if the value associated with state j is the highest. A selection bias arises if the unobserved characteristics that influence wages u_{ij} are correlated with the unobserved determinants of the selection process μ_{ij}, if $E(u|x,)$.

The vector $\mathbf{X}_{1i}$ includes productive characteristics of individual i namely years of education, age and age squared, whether the person is black, whether the person lives in an urban area, a macroeconomic demand side variable to capture rationing: regional unemployment rate by education group, and regional dummies. The vector $\mathbf{X}_{2ij}$ includes wage determinants that are only observable if individual i works: tenure in the current job, tenure squared, and indicators for the sector of activity. In the selection equation, $\mathbf{X}_{1i}$ includes wage determinants as they influence the choice of the work status. The vector $\mathbf{Z}$ additionally includes variables that do not directly affect wages but are relevant to the work status determination in the selection equation. I discuss the set of excluded variables in detail below.

To control for selection in the wage equation, I introduce a correction term that I denote $h(P_1, ..., P_3)$ where P_j denotes the probability to be in state j. The control function $h(.)$ is equal to the conditional mean of the error term $E(u_j|Y = j, \mathbf{X}_1, \mathbf{Z})$. Adopting Lee's (1983) approach, I assume that the joint distribution of u_j and a transformation[6] of μ_j does not depend on the other μ_k for $k \neq j$. Under this assumption and a additional linearity assumption the expected value of u_j, conditional on category j being observed is: [7]

$$E(u_j|Y = j, \mathbf{X}_1, \mathbf{Z}) = \sigma_j \rho_j \left(-\frac{\phi(\Phi^{-1}(P_j))}{P_j} \right),$$

with ϕ and Φ the standard normal density and cumulative distribution functions, σ_j is the standard deviation of the wage errors and ρ_j is the correlation coefficient between the errors of the outcome equation and the errors of the wage equation in sector j. The control function is $h = -\frac{\phi(\Phi^{-1}(P_j))}{P_j}$ and $\sigma_j \rho_j$ is estimated by least squares. Only one correlation parameter ρ_j is estimated per wage equation under this method.

I also follow Dubin and McFadden (1984) who make less restrictive assumptions on the correlation between u_j and the $(\mu_k - \mu_j)$. In particular, these allow the selection bias to originate in the correlation of u_j not only with μ_j but also with μ_k for $k \neq j$. The linearity assumption on the conditional mean of the wage equation residuals is as follows:

$$E(u_j|Y = j, \mathbf{X}_1, \mathbf{Z}) = \sigma_j \frac{\sqrt{6}}{\pi} \sum_k \rho_{jk}(\mu_k - E(\mu_k)),$$

---

[6]The transform is defined as $\Phi^{-1}(F(\mu_j|\mathbf{X}_1, \mathbf{Z}))$ with Φ the standard normal cumulative distribution function and F the cumulative distribution function of μ_j.

[7]See Lee (1983) for a proof.

where j is the final outcome and $k = 1, ...3$ all the potential outcomes. ρ_{jk} is the correlation coefficient between u_j and μ_k and Dubin and McFadden (1984) make the restriction that the correlation coefficients sum up to zero $\sum_k \rho_{jk} = 0$. Given the multinomial logit formulas it follows that:[8]

$$E(\mu_j - E(\mu_j)|V_j > \max_{s \neq j}(V_s), Z) = -\ln(P_j),$$

$$E(\mu_k - E(\mu_k)|V_j > \max_{s \neq j}(V_s), Z) = \frac{P_k \ln(P_k)}{1 - P_k}, \text{ for } k \neq j.$$

The following wage equations corrected for selection are then estimated by least squares:

$$\ln w_{ipj} = \alpha_{pfj} F_i + \alpha_{pmj} M_i + \mathbf{X}_i \beta_{pj} + \theta_{pj} h_{pj}(P_1, ..., P_3) + u_{ij} \text{ with } \alpha_{pfj} = -\alpha_{pmj}, \quad (5a)$$

$$\log w_{ifj} = \mathbf{X}_{ij} \beta_{fj} + \theta_{fj} h_{fj}(P_1, ..., P_3) + \epsilon_{ifj}, \quad (5b)$$

$$\log w_{imj} = \mathbf{X}_{ij} \beta_{mj} + \theta_{mj} h_{mj}(P_1, ..., P_3) + \epsilon_{imj}, \quad (5c)$$

where $\theta_j h_j(P_1, ..., P_3) = E(u_j|X, Y = j)$ and depends on the model assumptions. The estimation of equations (5) allows us to recover ρ_j, the correlation between u_j and μ_j, when Lee's model is adopted, and the correlation between u_j and all the μ_k for $k = \{1...j...3\}$ if the Dubin-Mac Fadden approach is used.

The total decomposition with the additional term that captures the difference in average selection bias is expressed as follows:

$$\overline{\ln W}_{mj} - \overline{\ln W}_{fj} = (\overline{\mathbf{X}}'_m - \overline{\mathbf{X}}'_f)\widehat{\beta_{pj}} + \overline{\mathbf{X}}'_m(\widehat{\beta}_{mj} - \widehat{\beta}_{pj}) + \overline{\mathbf{X}}'_f(\widehat{\beta}_{pj} - \widehat{\beta}_{fj})$$
$$+ \theta_{mj} \mathbf{h}_{mj}(P_1, ..., P_3) - \theta_{fj} \mathbf{h}_{fj}(P_1, ..., P_3).$$

The last term capturing the selection effect has been treated in different ways in the literature on wage gap decompositions. Neuman and Oaxaca (2004) present different variations of the decomposition when selection is controlled for and show how the selection term can be included in the endowment term and/or in the coefficient term. I follow Yun (2007) who advocates treating selection as a separate term in the decomposition. In that way, the selection term provides a measure of the difference between the observed wage gap and the gap in wage offers.[9]

The wage gap due to different returns to observable characteristics in sector j is:

$$\overline{WG}_{Sj} = \overline{\mathbf{X}}'_m(\widehat{\beta}_{mj} - \widehat{\beta}_{pj}) + \overline{\mathbf{X}}'_f(\widehat{\beta}_{pj} - \widehat{\beta}_{fj}). \quad (6)$$

---

[8]See Dubin and McFadden (1984) for a proof.

[9]This approach has been adopted by Reimers (1983) for the analysis of the ethnic wage gap in the U.S., by Wright and Ermisch (1991), Ogloblin (1999), Appleton et al. (1999) among others for gender wage gap decomposition, and by Ermisch and Wright (1993) for the estimation of wage offers in part-time and full-time jobs among women.

The adjusted wage gap in equation (6) differs from the one in (3). First, the coefficients are now consistently estimated following the treatment for selection. Second, instead of explaining part of the total observed wage gap, the difference in returns now explains the gap in wage offers $\overline{\ln W}_{mj} - \overline{\ln W}_{fj} - (\theta_{mj} h_{mj}(P_1, ..., P_3) - \theta_{fj} h_{fj}(P_1, ..., P_3))$.

Equations (5) are also estimated for various education groups separately to explore how the selection rules and the gender gaps in wage offers differ across groups.

3.3 Identification

To identify the effect of selection and purge the wage estimates from the selection bias without relying on the functional forms, I make exclusion restrictions. The excluded variables I chose are the share of household members holding formal jobs, as well as various demographic characteristics: the presence of children, the marital status, an indicator for lone mothers and the share of household members who need care (children under the age of 10 and elderly people above 85 years old). The latter variable may be more appropriate than just the number of children as, first, older children may take care of younger children, and second, older family members may live in the same household and receive care. It may be argued that children, or in general relatives who need care, can affect the productivity of women on the job (through fatigue and lower availability) and thus may not be an appropriate excluded variable. The share of household members holding a formal job can determine the probability to be in formal vs. informal employment through a network effect. Arguably, an individual with household members who are working in the formal segment can receive more information about job openings in the formal sector, how and where to search for formal jobs and/or might be recommended for a job. This networking effect may also be thought of as a cultural norms or social capital channel.[10] Additionally, conditioning on having found a formal job, the share of household members holding formal jobs has a priori no effect on wages.

The empirical approach will hence consist of computing the formal (informal) gender wage gaps, firstly controlling for observable characteristics only, and secondly controlling additionally for endogenous selection into formal (informal) employment. The latter specification implies to estimate the probability to be in each outcome in a first step using a multinomial logit. This empirical strategy will be applied to the whole sample, as well as separately for different education groups to capture potential heterogeneity in the selection patterns and wage gaps along the skill distribution.

4 Empirical evidence

4.1 The data

Individual information is taken from the 2009 Brazilian household survey, the Pesquisa Nacional por Amostras de Domicilio (PNAD). The work here focuses on urban areas. The PNAD provides information about the individuals of roughly 121,000 urban households. In 2009, around 215,000

---

[10]Rauch (1991) mentions this as the 'inheritance' of formal and informal sector status due to 'connections' made by formal sector relatives, and stresses that it reenforces inequality persistence across families.

working-age people (18-65) were interviewed, among whom 53% were women. Sample weights ensure the representativeness of the survey. The different employment categories are the following: employees (wage-earners of the public and private sectors) which include domestic workers employed by private households; self-employed; employer; unpaid and family workers. In this paper, I focus on gender differences among informal *wage-earners* only, including domestic workers but excluding self-employed, employers, unpaid and family workers.

The survey provides direct and reliable information that enables us to classify employees into formal and informal wage-earners. Individuals are asked if their labour card is signed by their employer; if it is not, they are not registered and are not entitled to any labour rights or benefits. The labour card is used in the private sector; workers in the public sector have other types of contracts and are considered as formal employees in this study.

Table 1 gives the demographic, household and educational characteristics of men and women holding formal and informal jobs. Informal employees are on average younger than formal employees. Men and women working formally are of the same age on average but in informal jobs women are slightly older than men. Women who hold informal jobs are more often the head of the household and live less often in couples compared to women in formal jobs. A larger share of women have young children in the informal sector: 48% of women working informally have children under 14 years of age against 43% for women in the formal sector.

The PNAD provides information on the composition of the household. A household can be made of several families, e.g. two families sharing a dwelling or one family hiring a domestic employee with or without his/her family. Formal workers live in households with a higher share of formal workers, which goes in the direction of the network argument. The share of formal household members is about 4 percentage points higher among formal female workers compared to informal female workers. This difference is about 6 percentage points among male workers. Women tend to live in households/families with a higher share of formal wage-earners; this differential can be explained by the higher male participation rate and lower male informality rate compared to the corresponding female rates, a difference that is discussed below.

Turning to educational and job characteristics, both men and women in formal employment are better educated than those in informal employment and women are more educated than men in both segments of the labour market. As documented in the literature on informality, wages are higher on average in formal jobs. Women earn lower hourly wages than men in both segments of the labour market. Full-time work is less common among women and among informal workers. There are no major differences across gender or sector in the distribution of age at first job nor in the average tenure in the current job, which is somewhat surprising as I would have expected higher job turnover and thus lower tenure in informal jobs. Among formal wage workers, women work more often in the public sectors (22% of women compared to 11% of men) and are more often civil servants.

Table 2 describes in more detail the educational attainment for different employment statuses. It reveals that the female distribution of school attainment dominates the male distribution. There are fewer low-educated women and more high-educated women in the overall population and especially among working individuals. The gender differences in educational attainment are the strongest among

Table 1: Descriptive statistics by gender

	Formal				Informal			
	(1)		(2)		(3)		(4)	
	Men		Women		Men		Women	
Demographics								
Age (mean)	35.47	(10.22)	35.78	(9.98)	31.43	(10.11)	34.07	(10.03)
Head of household	0.62	(0.49)	0.28	(0.45)	0.47	(0.50)	0.32	(0.47)
Living in couple	0.81	(0.39)	0.72	(0.45)	0.75	(0.43)	0.65	(0.48)
Children under 14	0.43	(0.50)	0.43	(0.49)	0.42	(0.49)	0.48	(0.50)
Household composition								
Number of people	3.70	(1.54)	3.57	(1.44)	4.05	(1.81)	3.82	(1.62)
Family members	3.55	(1.37)	3.42	(1.27)	3.83	(1.59)	3.64	(1.45)
among the working-age household members								
Share of the household members								
with a formal job[a]	0.17	(0.28)	0.22	(0.31)	0.11	(0.23)	0.18	(0.28)
Share of the family members								
with a formal job[a]	0.79	(0.33)	0.75	(0.34)	0.85	(0.28)	0.79	(0.31)
Mother lives in	0.24	(0.43)	0.22	(0.41)	0.37	(0.48)	0.21	(0.41)
the household								
Education								
Illiterate	0.04	(0.20)	0.02	(0.14)	0.08	(0.28)	0.06	(0.23)
Years of schooling (mean)	8.95	(3.97)	10.37	(3.77)	7.12	(3.96)	7.74	(3.88)
Job related variables								
Hourly Wage	8.73	(23.20)	7.60	(21.77)	4.38	(6.54)	3.81	(6.41)
Hours of work	44.30	(10.98)	38.49	(12.79)	43.78	(12.15)	35.52	(15.29)
Full time	0.91	(0.28)	0.75	(0.44)	0.88	(0.33)	0.61	(0.49)
Several jobs	0.05	(0.21)	0.06	(0.24)	0.04	(0.18)	0.04	(0.19)
Union membership	0.22	(0.41)	0.21	(0.41)	0.07	(0.25)	0.04	(0.20)
Public sector	0.11	(0.33)	0.22	(0.41)	.	.	.	.
Civil servant	0.07	(0.25)	0.14	(0.34)	.	.	.	.
Age at first job								
Under 10	0.09	(0.29)	0.06	(0.24)	0.13	(0.34)	0.08	(0.27)
10-14	0.37	(0.48)	0.26	(0.44)	0.39	(0.49)	0.35	(0.48)
15-17	0.29	(0.46)	0.28	(0.45)	0.31	(0.46)	0.28	(0.45)
17-19	0.16	(0.37)	0.21	(0.41)	0.13	(0.34)	0.15	(0.36)
20-24	0.07	(0.26)	0.14	(0.35)	0.05	(0.22)	0.09	(0.29)
25-29	0.01	(0.10)	0.03	(0.17)	0.01	(0.08)	0.02	(0.15)
More than 30	0.00	(0.04)	0.02	(0.12)	0.00	(0.04)	0.02	(0.14)
Tenure	2.85	(3.07)	2.89	(3.15)	2.66	(2.82)	2.84	(2.97)
(mean number of years)								
Night work	0.02	(0.14)	0.01	(0.08)	0.02	(0.13)	0.01	(0.08)
N	43483		31136		8397		8655	

Source: Author's calculation based on the PNAD 2009, IBGE, Brazil. Standard deviations in parentheses.
The columns give the shares among male formal wage-earners (1), female formal wage-earners (2), male informal wage-earners (3) and female informal wage-earners (4). [a] The share of working-age household/family members holding formal jobs excludes the respondent.

formal employees. This is consistent with a sorting of men and women where gender is a signal for labour market attachment or quit probability and working women compensate for a lower average labour market attachment in their group by having a higher education level (see Lazear and Rosen (1990) for a theoretical model and de la Rica et al. (2008) for an empirical analysis where they explain the distribution of the wage gap in Spain with a similar rationale). The table also shows that the informal population is diverse. 35% of women, against 38% of men, have primary education or less, and at the same time 11% of unregistered women and 10% of unregistered men have tertiary education.

Table 3 highlights differences across gender and educational level in participation rates, unemployment rates and informality rates in 2009. The participation rate is lower for women than men. It is on average 68% for women and 93% for men. The participation gap decreases with the education level. Among people with primary education or less, only 53% of women decide to participate in the labour market while 89% of men do so, which corresponds to a gap of 36 percentage points. The participation rate increases with education and more rapidly for women. Among people with tertiary education, the participation gap is of 10 percentage points.

Table 2: Share of educated people among inactive, formal and informal workers

	All		Not working		Formal		Informal workers	
	Men	Women	Men	Women	Men	Women	Men	Women
Primary or less	29	27	37	37	19	12	38	35
Secondary	53	52	48	51	59	51	51	54
Tertiary	18	21	15	13	22	37	10	11

Source: Author's calculation based on the PNAD 2009, IBGE, Brazil. All numbers are in percentage.

Table 3: Descriptive statistics by education groups and gender

Level of education	Participation rate		Unemployment rate		Informality rate among working individuals	
	Men	Women	Men	Women	Men	Women
Total	93	68	6	8	15	22
Primary or less	89	53	5	6	22	38
Secondary	95	71	7	9	14	22
Tertiary	97	87	4	5	7	7

Source: Author's calculation based on the PNAD 2009, IBGE, Brazil.

Columns 2 and 3 in Table 3 report the unemployment rate by gender and educational level. The female unemployment rate is higher than the male unemployment rate in all education groups ; the difference is larger for people with medium level of education.

The last two columns show the informality rate which measures the share of wage-earners without a registered labour contract. Informality is higher for women than for men, the difference being larger for employees with secondary education or less. The informality rate decreases with education. Among female wage-earners, 38% of women with primary education or less are employed without a contract; this share is 22% among women with secondary education and 7% of women with tertiary education.

The gender gap in informality rates decreases with education as well. It is huge among low-skill people with 16 percentage points difference. It is about 7 percentage points higher for women among workers with secondary education or less. The informality rate is the same for men and women with tertiary education.

Table 4: Employment shares and informality rate by gender and sectors in urban areas

Sector	Employment share		Informality rate	
	Women	Men	Women	Men
Agriculture	3	7	12	34
Construction	1	16	11	21
Manufacturing	13	17	12	12
Services	83	60	25	13

Source: Author's calculation based on the PNAD, 2009, IBGE, Brazil. Urban population only.

Table 5: Employment shares and informality rate by gender and occupations

Occupation	Employment share		Informality rate	
	Women	Men	Women	Men
Managers and Officials	5	7	5	4
Professionals	12	6	6	9
Technicians and associate professionals	7	8	10	12
Clerical support workers	14	7	10	9
Service and sale workers	49	24	37	14
Skilled agricultural and forestry workers	3	7	13	34
Craft and related trades workers	8	34	14	18
Operators, and assemblers	2	7	12	17

Source: Author's calculation based on the PNAD, 2009, IBGE, Brazil.

I now turn to the distribution of female and male employment across sectors and to the informality rate within sectors. As reported in Table 4, 83% of female employees work in the service sector where the informality rate for women is 25%. Only 60% of male employees work in this sector and they have a lower informality rate, 13%. The highest informality rate is in the construction activities. Only 1% of working women are employed in the construction sector but 11% of them hold informal jobs. The manufacturing industry employs 13% of the female labour force and 17% of the male labour force and the informality rate for both men and women 12%. In agriculture, a sector that employs respectively 3% and 7% of the female and male urban labour force, the female informality rate is lower than the male informality rate by 20 percentage points.

Table 5 gives the distribution of female and male employment across occupations and the informality rate within each occupation. The informality rate is the lowest among managers and officials for whom 5% of female managers and 4% of male managers report not having a registered contract. It is the highest for women in the service and sale occupation where 37% of women are informal. This occupation

employs almost the majority of female employees. In contrast, only 24% of men work as service and sale workers and 14% of them are informally employed.

4.2 Wage distributions across genders

To complete the preliminary description of the gender differences across formal and informal jobs, I compute raw wage differences. Table 1 showed the average raw hourly wages in formal and informal jobs for men and women. Figure 2 displays here the hourly wage distributions.

Figure 2: Wage Distributions by gender
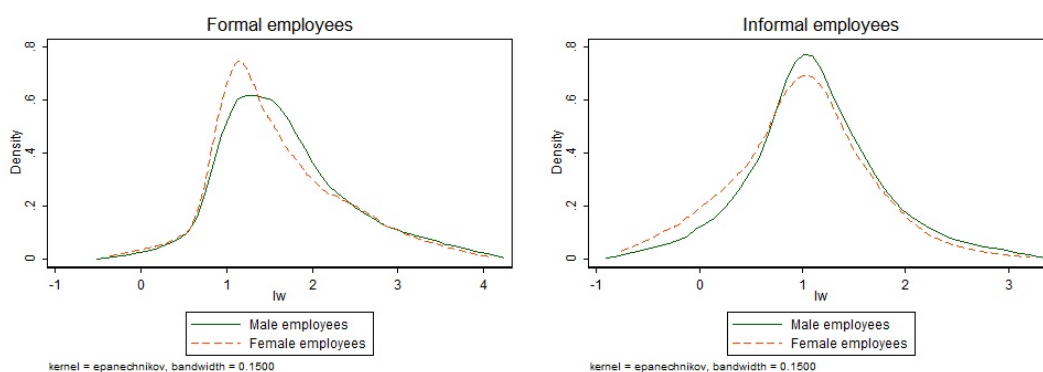
Source: PNAD, 2009, IBGE, Brazil.

Among formal workers, the female wage distribution is shifted farther to the left compared to the male wage distribution which indicates that the raw difference between male and female wages is positive especially in the middle of the wage distribution. This is true at all education levels (Figure 3), for formal employees in the service sector and in the manufacturing sector (Figure 4).

In the informal segment of the labour market, the male and female wage distributions are almost identical except at the bottom where the lower tail of the female distribution is fatter. This pattern is however not observed among men and women with tertiary education. For this group, the male distribution dominates the female distribution which indicates that the raw gender wage gap in the informal segment is higher among high-skill workers.

17

Figure 3: Wage Distributions by gender and education level
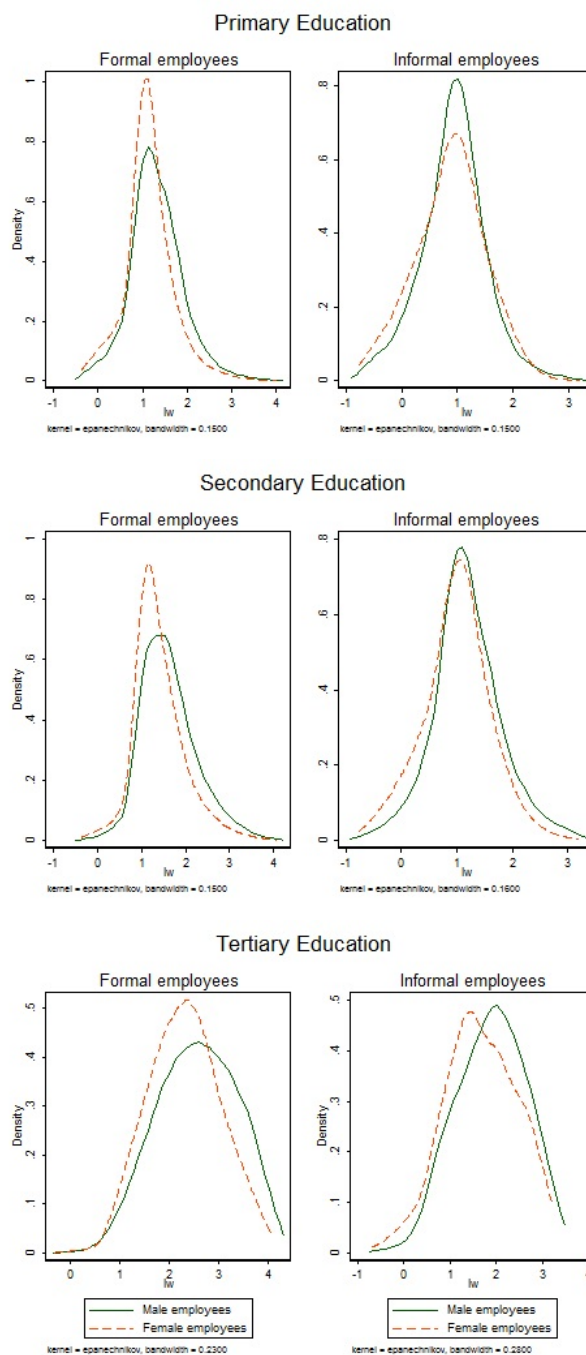
Figure 4: Wage Distributions by gender and sector
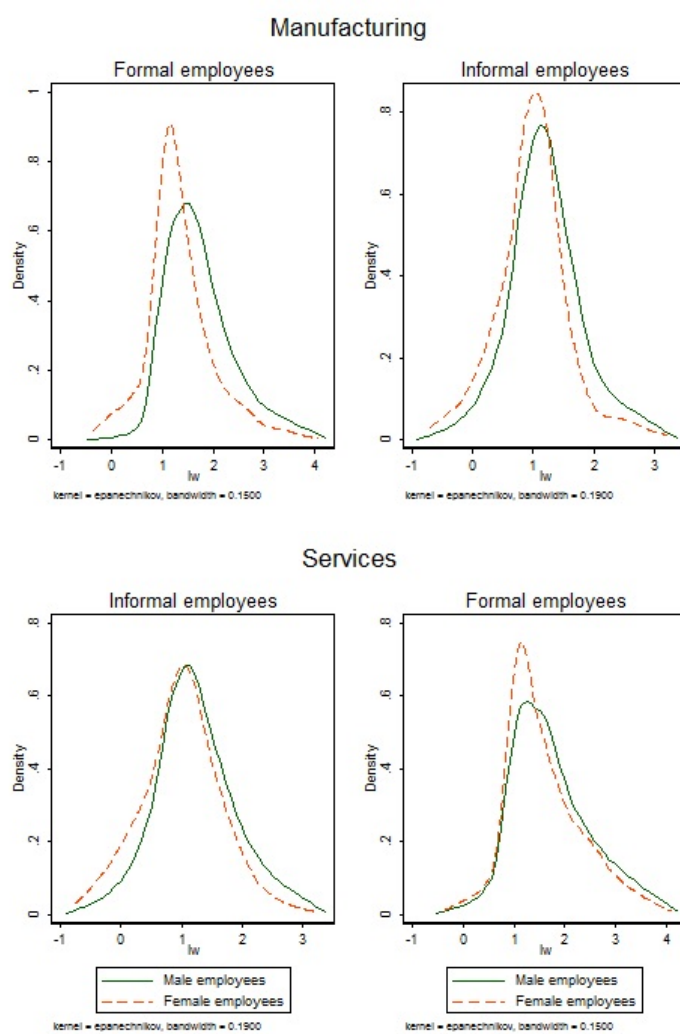
4.3 Selection into multiple potential employment statuses

Table 6: Labour market status, marginal effects for women.

Women	Not working	Informal employee	Formal employee
Age	-0.039***	0.008***	0.031***
	(0.001)	(0.001)	(0.001)
Age^2	0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Years of education	-0.028***	-0.011***	0.039***
	(0.001)	(0.000)	(0.002)
Non-white	-0.026***	0.021***	0.005
	(0.005)	(0.002)	(0.005)
Having children	0.022***	-0.009	-0.013***
	(0.006)	(0.005)	(0.004)
...under 14	0.011	0.000	-0.011
	(0.008)	(0.005)	(0.007)
Head of household	-0.044***	0.018*	0.026**
	(0.013)	(0.007)	(0.009)
Living in couple	0.078***	-0.032***	-0.046***
	(0.011)	(0.003)	(0.010)
Lone mother	-0.066***	0.033***	0.033***
	(0.010)	(0.007)	(0.009)
Share of formal workers in the household^a	-0.044***	-0.033***	0.077***
	(0.007)	(0.004)	(0.005)
Share of individuals who need care	0.234***	-0.075***	-0.159***
	(0.014)	(0.010)	(0.017)
Unemployment rate (regional, by education group)	1.184***	0.233*	-1.418***
	(0.125)	(0.099)	(0.118)
N 72,982			

Notes: Marginal effects, standard errors in parenthesis. The marginal effects of each explanatory variables on the probability to be in the three different outcomes are computed based on a multinomial logit estimation.^{a}The share of working-age household members holding formal jobs excludes the respondent.

I start the empirical analysis by estimating the multinomial logit equation (4) to understand the impact of supply side and demand side variables on the probability of being a formal employee, an informal employee or not working. I estimate the multinomial logit model for men and women separately. The marginal effects are reported in Tables 6 and 7. The tables provide an estimate of the effect of a marginal change in each continuous variable, or a discrete change for indicators, for an individual with average characteristics in the male sample and in the female sample. The estimated coefficients of the multinomial logit estimation are provided in the appendix.

The probability to work as formal employee increases with age for both men and women. An additional year of education reduces the probability of being out of the labour force much more for women; it also increases the chances to be formally employed, the latter effect being stronger for women again.

Table 7: Labour market status, marginal effects for men.

Men	Not working	Informal employee	Formal employee
Age	-0.022***	-0.010***	0.032***
	(0.001)	(0.001)	(0.001)
Age^2	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Years of education	-0.013***	-0.014***	0.027***
	(0.001)	(0.001)	(0.001)
Non-white	-0.003	0.006	-0.003
	(0.003)	(0.003)	(0.003)
Having children	0.021***	-0.017***	-0.005
	(0.004)	(0.005)	(0.007)
...under 14	-0.043***	0.004	0.039***
	(0.005)	(0.005)	(0.006)
Head of household	-0.117***	-0.032***	0.150***
	(0.006)	(0.005)	(0.006)
Living in couple	-0.045***	-0.006	0.050***
	(0.003)	(0.005)	(0.006)
Share of formal workers	-0.047***	-0.114***	0.161***
in the household^a	(0.005)	(0.005)	(0.007)
Share of individuals	-0.022	0.030**	-0.007
who need care	(0.014)	(0.010)	(0.016)
Unemployment rate	0.093	-0.292**	0.199
(regional, by education group)	(0.090)	(0.110)	(0.143)
N 55,992			

Notes: Marginal effects, standard errors in parenthesis. The marginal effects of each explanatory variables on the probability to be in the three different outcomes are computed based on a multinomial logit estimation.^{a}The share of working-age household members holding formal jobs excludes the respondent.

Other variables such as the family structure have opposite effects on men and women. The presence of young children and living in couples reduce the probability of inactivity for men while it increases it for women. This does not hold for lone mothers who have a greater probability to hold a formal or informal job and a lower probability to remain inactive compared to women without children (the effect of being a lone mother overcompensates the effect of having children on each outcome probability). Contrary to women with young children living in couples, men with young children have a lower probability to be inactive but a higher probability to hold a formal job. Those results are consistent with the traditional division of roles within the household.

People in a household with a higher share of formal workers are more likely to hold a formal jobs and less likely to hold an informal job. The effect is stronger for men. This finding points towards the network effect discussed in the previous section.

Looking at demand side factors, I find that a higher regional unemployment rate, calculated for each skill and gender group, increases non-participation for women but not for men. It reduces the probability to find a formal job and increases the probability to hold an informal job for women. The opposite holds for men. Higher unemployment rates reduces the probability to have an informal jobs relative to the two other outcomes. This may reveal an insurance effect: as it becomes tougher to find a job, men are less willing to work informally and tend to search more intensively for formal jobs that

are more secure and provide unemployment benefits in case of lay off.

4.4 Wages

Table 8 presents the estimates of the female wage equations and Table 9 the estimates of the male wage equations. In each table, columns (1) and (2) give the OLS estimates of a simple wage equation and columns (3) to (6) give the ones of a wage equation augmented with the control function.[11]

The return to education is stronger in the formal sector for both women (Table 8) and men (Table 9), this pattern is robust to the introduction of the selection control function. Wages also increase significantly with age, the effect is of the same magnitude in both formal and informal jobs but it is stronger for men than for women. I do not see here evidence of a concave effect of age on wages.

Tenure in current firm increases male wages in formal jobs once I control for selection but not in informal jobs. As for women, the effect is not significant in formal jobs while it has a negative effect in informal jobs. As I control for age, this result does not mean that actual wages are declining along women's careers. Negative returns to tenure in informal jobs means that women who keep on working informally for the same employer have lower wages compared to women who have changed job more recently. The negative returns of tenure with the same employer may be due to low female job mobility along with monopsony power of employers. Kambourov and Manovskii (2009) and Sullivan (2009) argue that specific skills are occupation rather than firm specific. They find that tenure with the same employer has zero or negative effects on wages when tenure in occupation and in the industry are controlled for. Schmiedern (2007) also finds negative returns to tenure with the employer for women in Germany.

The regional unemployment rate affects negatively wages; this effect is also robust to selection treatment. For both women and men, unemployment reduces more the formal employees' wages; according to this result, the formal segment of the labour market seems to be competitive.

In columns (3) to (6) of Tables 8 and 9, the control function is included as an additional regressor. The selection bias is significant in both the formal and the informal sectors, for both men and women. Tables 8 and 9 report the correlations between the errors of the wage equation and the errors of the selection equation when all education groups are pooled together. The correlation coefficient gives us the direction of the average selection rules for men and for women. Note that when Lee's approach is adopted, a negative $\sigma_j \rho_j$ implies a positive selection bias as $\sigma_j \rho_j \left(-\frac{\phi(\Phi^{-1}(P_j))}{P_j} \right)$ is strictly positive.[12] Men are positively selected in informal employment and negatively selected in formal employment according to both Lee's and DMF's method. This is consistent with the conclusions of papers that find that some workers have an individual comparative advantage in the informal sector and would not do any better in the formal sector (Gindling, 1991; Maloney, 1999, 2004). For given values of observable characteristics, men holding informal jobs have unobserved characteristics that are on average most valued in this sector. Consequently, observed wages overestimate male wage offers in informal jobs. By contrast, men are on average negatively selected into the formal sector. Those with the highest

---

[11] For this step, I use the Stata command selmlog that implements the methods described and evaluated in Bourguignon et al. (2007). See http://www.parisschoolofeconomics.com/gurgand-marc/selmlog/selmlog13.html

[12] See Lee (2007).

wage potentials in formal jobs do not self-select into those jobs and choose other work statuses (note that I do not consider self-employed and employers). Negative selection occurs when the reservation wage is increasing with the wage offer. This selection pattern remains the same when I estimate the selection term for different education groups.[13] As for women, they are positively selected in both informal and formal jobs. The average selection rule hides heterogeneity across education groups. The positive selectivity bias in formal jobs holds for women with secondary education or less but not for women with tertiary education who are negatively selected in formal jobs as men are. Highly educated women working in the formal sector are those with lower wage potential compared to highly educated women in other work statuses. The implications of the different selection patterns across gender for the estimated gender wage gaps in formal and in informal jobs are discussed in the next section.

---

[13]Results available upon request.

Table 8: Female hourly wages in the informal and formal sectors.

	OLS		Selection			
Control function	Informal	Formal	Informal		Formal	
			Lee	DMF	Lee	DMF
	(1)	(2)	(3)	(4)	(5)	(6)
Years of education	0.042**	0.072**	0.033**	0.015**	0.085**	0.112**
	(0.003)	(0.002)	(0.003)	(0.004)	(0.006)	(0.004)
Age	0.037**	0.026**	0.044**	0.037**	0.035**	0.039**
	(0.003)	(0.002)	(0.004)	(0.005)	(0.002)	(0.001)
Age2	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure (in years)	-0.022**	0.004	-0.022**	-0.022**	0.004	0.004**
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.002)
Tenure^2	0.002**	-0.001*	0.002**	0.002**	-0.001	-0.001**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Black	-0.045**	-0.086**	-0.028**	-0.022*	-0.085**	-0.105**
	(0.013)	(0.008)	(0.010)	(0.009)	(0.002)	(0.011)
Unemployment rate	-4.407**	-7.251**	-3.919**	-3.056**	-7.538**	-7.364**
(regional, by education group)	(0.483)	(0.778)	(0.430)	(0.110)	(0.032)	(0.273)
Constant	-0.020	0.763**	-0.383*	-0.247	0.330*	-0.192**
	(0.082)	(0.075)	(0.156)	(0.169)	(0.160)	(0.029)
Sector dummies	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes
σ^2			0.453**	0.383**	0.248**	0.487**
			(0.056)	(0.039)	(0.016)	(0.099)
ρ_1				0.185		0.471**
				(0.154)		(0.091)
ρ_2			-0.292**			-1.006**
			(0.047)			(0.081)
ρ_3				-0.531**	-0.311**	
				(0.132)	(0.075)	
R^2	0.27	0.53				
N	12,569	29,464	12,569	29,464	12,569	29,464

Notes: * $p < 0.05$; ** $p < 0.01$ Standard errors in parenthesis. Clustered s.e. in OLS regressions. Bootstrap estimates of the s.e. when controlling for selection to account for the two-step procedure. In Lee's approach, a negative $\sigma_j\rho_j$ implies a positive selection bias as $\sigma_j\rho_j h(P_j)$ is then strictly positive. The DMF approach imposes the restriction $\sum_k \rho_{jk} = 0$ so that ρ_2 in column (4) and ρ_3 in column (6) are both positive. The selection bias is positive in both formal and informal jobs using both approaches.

Table 9: Male hourly wages in the informal and formal sectors.

Control function	OLS		Selection			
	Informal	Formal	Informal		Formal	
			Lee	DMF	Lee	DMF
	(1)	(2)	(3)	(4)	(5)	(6)
Years of education	0.053**	0.078**	0.031**	0.026**	0.068**	0.080**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Age	0.053**	0.046**	0.041**	0.027**	0.031**	0.031**
	(0.004)	(0.004)	(0.006)	(0.002)	(0.001)	(0.002)
Age2	-0.001**	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure (in years)	-0.002	0.005	-0.003	-0.004	0.004**	0.004**
	(0.007)	(0.003)	(0.010)	(0.003)	(0.001)	(0.001)
Tenure^2	-0.000	-0.001*	-0.000	0.000	-0.001**	-0.001*
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Black	-0.101**	-0.090**	-0.090**	-0.090**	-0.088**	-0.096**
	(0.014)	(0.007)	(0.003)	(0.006)	(0.003)	(0.002)
Unemployment rate	-4.125**	-7.166**	-4.199**	-4.157**	-7.397**	-7.015**
(regional, by education group)	(0.435)	(0.386)	(0.606)	(0.025)	(0.003)	(0.136)
Constant	-0.072	0.341**	-0.195*	0.182**	0.877**	0.530**
	(0.077)	(0.075)	(0.097)	(0.020)	(0.002)	(0.036)
Sector dummies	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes
σ^2			0.866**	0.396**	0.247**	0.508**
			(0.133)	(0.014)	(0.004)	(0.086)
ρ_1				0.210**		0.798**
				(0.028)		(0.048)
ρ_2			-0.467**			-0.783**
			(0.020)			(0.058)
ρ_3				-0.629**	0.451**	
				(0.044)	(0.022)	
R^2	0.33	0.48				
N	11,056	37,851	11,056	37,851	11,056	37,851

Notes: * $p < 0.05$; ** $p < 0.01$ Standard errors in parenthesis. Clustered s.e. in OLS regressions. Bootstrap estimates of the s.e. when controlling for selection to account for the two-step procedure. In Lee's approach, a negative $\sigma_j \rho_j$ implies a positive selection bias as $\sigma_j \rho_j h(P_j)$ is then strictly positive. The DMF approach imposes the restriction $\sum_k \rho_{jk} = 0$ so that ρ_2 in column (4) is positive and ρ_3 in column (6) is negative. The selection bias is positive in informal jobs and negative in formal jobs using both approaches.

4.5 The gender wage gap in informal and formal jobs

Table 10 displays the estimated gender wage gaps in formal and in informal jobs for the whole population as well as for different groups of education. Bold characters indicate that the difference between the formal and the informal wage gaps is significant at 10% ($|t| > 1.64$) according to the Welch's test.

Panel 1 gives the raw (or total) gender wage gaps. They are positive and significant in both formal and informal jobs. The average raw gap is significantly higher among informal employees, 13% compared to 5% among formal employees.[14] However, this conceals different compositions of the male and female labour force. When I estimate the gaps for different education groups, which already accounts for some differences in productive characteristics, I see that the gender wage gap in informal jobs is higher than the gap in formal jobs but not significantly so only for the most educated employees (29% vs. 28%) . For workers with primary education or no education, the wage gap is stronger among employees with a legal contract which may seem at odds with the intuition that labour market regulation, in particular minimum wages, should reduce the scope for wage gaps at the bottom of the wage distribution. Note however that the minimum wage also serves as a reference for employers deciding over informal employees' wages. This so called "lighthouse" effect implies that the minimum wage is actually most binding in informal jobs where the lowest wages are observed. Another striking pattern is the increase in the gender wage gap with the education level which can be interpreted as a form of "glass ceiling" in both the formal and the informal segments. That being said, the raw wage gap does not account properly for the labour force heterogeneity. As men and women might have different characteristics in both types of jobs, the next panels provide a more detailed analysis.

Panel 2 of Table 10 shows that controlling for observable characteristics such as the exact number of years of education, age, tenure within the firm, sector of activity and location, increases the wage gap in both formal and informal sectors. This is an expected result as women are more educated than men and working women present overall better characteristics on average than working men in this sample. Since the female advantage in observables is stronger in the formal sector, the formal adjusted wage gap increases more than the informal adjusted wage gap (from panel 1 to panel 2). As a result, the wage gaps in formal and informal jobs are not statically different from one another. Although skills receive lower returns in informal jobs, the gender difference in returns is about the same in percentage terms in formal and informal jobs: the average wage gap is about 0.2 log points which amounts to a difference of 22 percent. If I look at the gaps for different education groups, adjusting women's returns to the returns obtained by men would increase women's wages by 14 (21) percent among formal (informal) female employees with primary education. It would increase women's wages by 23 (27) percent among formal (informal) female employees with secondary education. For high-skilled workers however, the adjusted wage gaps are not statistically different in the two segments. The small sample size for workers with tertiary education might be responsible for the lower precision in the estimates. Looking at the numbers, the gap is higher among formal employees at 28 percent compared to 24 percent among informal employees. The so-called "glass ceiling effect" may be stronger in jobs with

---

[14]The results are expressed on the logarithmic scale. To obtain the difference in percentage points: $(exp(WG)-1) \times 100$.

legal labour contracts which goes in the direction of the hypothesis formulated in section 2.

I now turn to the effect of the selection bias. The data give information on observed wages only, for workers working in a given sector. To infer the magnitude of the wage gap correctly though, I want to compare wage *offers* (that would be) made to all men and women. If selection into formal vs. informal jobs is non-random, observed wages either overstate or understate wage offers. If the selection bias differs by gender, the observed raw wage gap, given in the first panel of Table 10, will not reflect the raw difference in wage offers. Controlling for sorting across work statuses enables me to recover the average wage offers within each segment of the labour market, providing the control function captures the selection bias properly. Panel 3 shows how selection changes the average gender wage gaps differently in the informal and formal sectors.

In the informal sector, the observed wage gap overstates the wage gap in wage offers (from panel 1 to panel 3.1). This is because observed informal wages overestimate informal wage *offers* for both men and women but male wage offers are more strongly overestimated than female wage offers. Controlling for selection into informal jobs reduces the average informal wage offer more for men than for women, and thus reduces the average wage gap that remains to be explained by differences in characteristics and in the returns to those characteristics. Among informal employees, differences in returns have no role in explaining the gap anymore (panel 3.2). Put differently, after purging the estimates from the selection effect, I cannot reject the hypothesis that men and women receive equal treatment for their skills in the informal sector.

In the formal sector, on the other hand, the gender gap in wage offers is offset by the selection bias and is underestimated by the observed gender wage gap. This is because observed male wages underestimate male wage offers while observed female wage overestimate female wage offers (panel 3.1). Observable characteristics are better among working women which makes the part of the gender wage gap due to different returns even bigger than the total wage difference adjusted for the selection bias (panel 3.2). The increase in the wage gap with education is robust to the treatment of selection.

These results highlight that labour regulation may impact gender wage inequality in the urban labour market in Brazil. The finding that wage gaps are positive and significant only in formal jobs is consistent with the following explanation. If employers believe that women have a higher quit probability, statistical discrimination induces employers to pay lower wages to women because they expect higher average female labour cost. I argue that the gap in expected labour cost because of gender differences in labour market attachment is higher in jobs where employment protection is binding. When an employee takes a temporary leave, his/her job must remain available to him/her, generating costs due to vacancy or temporary replacement. This effect is expected to be weaker in informal jobs because the job of the employee on leave can be allocated to another worker. I also find that the wage gap is higher among high-skilled workers in formal jobs, a result that is commonly found in the literature on gender wage gaps. This finding is often explained by statistical discrimination that produces higher gender gaps in high-wage jobs. Higher gender differences in pay among formal workers and the increase in the pay gap with the education level are thus consistent.

Table 10: Gender wage gap decomposition. Informal and formal sectors, urban areas.

Level of Education	All	Primary or less	Secondary	Tertiary
1-Total wage gap: $\overline{TWG}_j = \overline{\ln W}_{mj} - \overline{\ln W}_{fj}$				
Informal	**0.123****	**0.046***	0.190**	0.252**
	(0.016)	(0.028)	(0.016)	(0.050)
Formal	**0.051***	**0.162****	0.195**	0.246**
	(0.022)	(0.019)	(0.016)	(0.016)
Welch's t-statistics	-2.66	3.94	0.21	-0.19
2-Controlling for observables only				
Part due to differences in returns: $\overline{WG}_j = \overline{\mathbf{X}}'_m(\widehat{\beta}_{mj} - \widehat{\beta}_{pj}) + \overline{\mathbf{X}}'_f(\widehat{\beta}_{pj} - \widehat{\beta}_{fj})$				
Informal	0.214**	**0.190****	**0.236****	0.213**
	(0.021)	(0.032)	(0.019)	(0.054)
Formal	0.205**	**0.135****	**0.203****	0.246**
	(0.011)	(0.014)	(0.012)	(0.016)
Welch's t-statistics	-0.05	-2.22	-1.67	0.99
3-Gap in wage offers. Controlling for observables and self-selection				
3.1-Total gap in wage offers : $\overline{TWG}_{Sj} = \overline{\ln W}_{mj} - \overline{\ln W}_{fj} - (\theta_{mj}h_{mj} - \theta_{fj}h_{fj})$				

		All	Primary or less	Secondary	Tertiary
Lee	Informal	**-0.179**	**-0.199***	**-0.044**	-0.663
		(0.104)	(0.080)	(0.105)	(0.812)
	Formal	**0.357****	**0.241****	**0.420****	0.828**
		(0.024)	(0.045)	(0.023)	(0.07)
	Welch's t-statistics	4.44	4.21	3.82	1.41
DMF	Informal	**-0.021**	**-0.07**	**0.129**	-0.238
		(0.089)	(0.096)	(0.085)	(0.666)
	Formal	**0.332****	**0.234****	**0.331****	0.511**
		(0.021)	(0.055)	(0.028)	(0.074)
	Welch's t-statistics	3.60	2.46	2.07	0.86

3.2-Part due to difference in returns : $\overline{WG}_{Sj} = \overline{\mathbf{X}}'_m(\widehat{\beta}_{mj} - \widehat{\beta_{pj}}) + \overline{\mathbf{X}}'_f(\widehat{\beta_{pj}} - \widehat{\beta_{fj}})$

		All	Primary or less	Secondary	Tertiary
Lee	Informal	**-0.101**	**-0.055**	**-0.010**	-0.701
		(0.103)	(0.080)	(0.106)	(0.812)
	Formal	**0.446****	**0.158****	**0.371****	0.479**
		(0.018)	(0.048)	(0.025)	(0.048)
	Welch's t-statistics	6.86	1.61	3.07	1.47
DMF	Informal	**0.054**	0.060	0.160	-0.271
		(0.090)	(0.097)	(0.083)	(0.669)
	Formal	**0.469****	0.161**	0.321**	0.332**
		(0.020)	(0.055)	(0.026)	(0.074)
	Welch's t-statistics	5.56	1.02	1.40	0.74
Number of informal workers		23,625	9,316	12,886	1,423
Share of women		53%	51%	55%	57%
Number of formal workers		67,315	12,019	40,935	14,361
Share of women		44%	32%	42%	60%

Notes: * $p < 0.05$; ** $p < 0.01$ s.e. in parenthesis. Panel 1: equation (1). Panel 2: equation (3). Panel 3.2: equation (6). The results are expressed on the logarithmic scale. To obtain the difference in percentage points: $(exp(WG) - 1) \times 100$. The Welch's test is applied to test the difference between the formal and the informal gaps with different population sizes and variances. Bold characters indicate that the difference between the formal and the informal wage gaps is significant at 10% when $|t| > 1.64$, the difference is significant at 5% if $|t| > 1.96$.

5 Conclusion

This paper investigates gender wage inequality in formal and informal jobs in Brazil. The data shows that the average raw gender wage gap is positive and significant in both sectors. The informal sector features the highest total average gender wage gap but this conceals differences in male and female characteristics. When I ignore the selection bias, the differences in returns to productive characteristics are the same in formal and informal jobs and are responsible for about 22% of the pay gap. The paper additionally shows that the similarity between the formal and informal wage gaps is artificially generated by different selection of men and women into formal and informal jobs. In the informal sector, both male and female observed average wages overestimate their respective average wage offers but the more so for men. As a result the observed wage gap overestimates the gap in wage offers. I find that the difference in average wage offers faced by men and women is actually entirely explained by differences in selection bias. The gender gap due to different returns to characteristics is not significant in the informal sector. The opposite happens in the formal sector. The gender difference in selectivity bias narrows the gender gap in observed wages. This is because observed female wages overestimate female wage offers while observed male wages underestimate male wage offers. As a result, even after controlling for selection, the gender wage gap due to different returns to characteristics is strongly positive in formal jobs. Moreover, the gender wage gap increases with education in the formal sector.

Bigger gender differences in returns to productive characteristics in the formal sector certainly cannot lead to the conclusion that employment protection legislation is detrimental to women. First, the formal segment of the labour market provides higher wages to women, even if the formal wage premium is lower for women than for men. Additionally, given that women face a higher unemployment rate and need to take maternity leave, the flow of earnings of women relative to men can be higher in the formal sector because unemployment benefits and maternity leave benefits compensate for wage losses in the formal sector while this does not happen in the informal sector. Further work is needed, first to really identify the impact of labour regulation on discriminatory behavior and, second, to investigate how participating in the informal sector affects gender differences in earnings over the life cycle.

References

Albrecht, J., Bjorklund, A., and Vroman, S. (2003). Is there a glass ceiling in sweden? *Journal of Labor Economics*, 21(1):145–177.

Almeida, R. and Carneiro, P. (2007). Inequality and employment in a dual economy: Enforcement of labor regulation in brazil.

Appleton, S., Hoddinott, J., and Krishnan, P. (1999). The gender wage gap in three african countries. *Economic Development and Cultural Change*, 47(2):289–312.

Arabsheibani, G. R., Carneiro, F. G., and Henley, A. (2003). *Gender wage differentials in Brazil: trends over a turbulent era*, volume 3148. World Bank.

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

Blau, F. D. and Kahn, L. M. (2006). The us gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review*, pages 45–66.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455.

Bourguignon, F., Fournier, M., and Gurgand, M. (2007). Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys*, 21(1):174–205.

Carneiro, F. G. and Henley, A. (2002). Modelling formal vs. informal employment and earnings: Micro-econometric evidence for brazil.

de la Rica, S., Dolado, J., and Llorens, V. (2008). Ceilings or floors? gender wage gaps by education in spain. *Journal of Population Economics*, 21:751–776.

Deininger, K., Jin, S., and Nagarajan, H. (2013). Wage discrimination in india's informal labor markets: Exploring the impact of caste and gender. *Review of Development Economics*, 17(1):130–147.

Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, pages 345–362.

Ermisch, J. F. and Wright, R. E. (1993). Wage offers and full-time and part-time employment by british women. *The Journal of Human Resources*, 28(1):111–133.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. *Handbook of labor economics*, 4:1–102.

Fortin, N. M. (2008). The gender wage gap among young adults in the united states the importance of money versus people. *Journal of Human Resources*, 43(4):884–918.

Gasparini, L. and Tornarolli, L. (2009). Labor informality in latin america and the caribbean: patterns and trends from household survey microdata. *Desarrollo y Sociedad*, (63):13–80.

Gindling, T. H. (1991). Labor market segmentation and the determination of wages in the public, private-formal, and informal sectors in san jose, costa rica. *Economic Development and Cultural Change*, 39(3):585–605.

Gong, X. and Van Soest, A. (2002). Wage differentials and mobility in the urban labour market: a panel data analysis for mexico. *Labour Economics*, 9(4):513–529.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American economic review*, pages 622–641.

Gunther, I. and Launov, A. (2012). Informal employment in developing countries: opportunity or last resort? *Journal of Development Economics*, 97(1):88–98.

Kambourov, G. and Manovskii, I. (2009). Occupational mobility and wage inequality. *The Review of Economic Studies*, 76(2):731–759.

Lazear, E. P. and Rosen, S. (1990). Male-female wage differentials in job ladders. *Journal of Labor Economics*, 8(1):S106–S123.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, pages 507–512.

Lee, L.-F. (2007). *Self-Selection*, pages 383–409. Blackwell Publishing Ltd.

Madalozzo, R. (2010). Occupational segregation and the gender wage gap in brazil: an empirical analysis. *Economia Aplicada*, 14(2):147–168.

Magnac, T. (1991). Segmented or competitive labor markets. *Econometrica*, pages 165–187.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? evidence from sectoral transitions in mexico. *The World Bank Economic Review*, 13(2):275–302.

Maloney, W. F. (2004). Informality revisited. *World Development*, 32(7):1159–1178.

Mandel, H. and Semyonov, M. (2005). Family policies, wage structures, and gender gaps: Sources of earnings inequality in 20 countries. *American sociological review*, 70(6):949–967.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

Neuman, S. and Oaxaca, R. L. (2004). Wage decompositions with selectivity-corrected wage equations: A methodological note. *The Journal of Economic Inequality*, 2(1):3–10.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Oaxaca, R. L. and Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1):5 – 21.

Oaxaca, R. L. and Ransom, M. R. (1999). Identification in detailed wage decompositions. *The Review of Economics and Statistics*, 81(1):154–157.

Ogloblin, C. G. (1999). The gender earnings differential in the russian transition economy. *Industrial and Labor Relations Review*, 52(4):602–627.

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

Pagán, J. A. and Ullibarri, M. (2000). Group heterogeneity and the gender earnings gap in mexico. *Economía Mexicana, Nueva Época*, 9(1):23–40.

Pradhan, M. and Van Soest, A. (1995). Formal and informal sector employment in urban areas of bolivia. *Labour Economics*, 2(3):275–297.

Rauch, J. E. (1991). Modelling the informal sector. *Journal of Development Economics*, 35:33 – 47.

Reimers, C. W. (1983). Labor market discrimination against hispanic and black men. *The review of economics and statistics*, 65(4):570–579.

Ruhm, C. J. (1998). The economic consequences of parental leave mandates: Lessons from europe. *The Quarterly Journal of Economics*, 113(1):285–317.

Santos, R. V. and Ribeiro, E. P. (2006). Diferenciais de rendimentos entre homens e mulheres no brasil revisitado: Explorando o teto de vidro. *Centro Universitario Newton Paiva/MG e PPGE/UFRGS*.

Schmiedern, J. (2007). Returns to tenure: Is specific human capital acquired on the job? mimeo.

Sullivan, P. (2009). Estimation of an occupational choice model when occupations are misclassified. *Journal of Human Resources*, 44(2):495–535.

Tansel, A. (2001). Wage earners, self employed and gender in the informal sector in turkey. Economic Research Forum.

Wright, R. E. and Ermisch, J. F. (1991). Gender discrimination in the british labour market: a reassessment. *The Economic Journal*, 101(406):508–522.

Yun, M.-S. (2007). An extension of the oaxaca decomposition using generalized residuals. *Journal of Economic and Social Measurement*, 32(1):15–22.

A Multinomial logit estimates

Table 11: Labour market status

Relative risk ratios from the multinomial logit estimation. Base: formal employee

Women	Not working	Informal employee
Age	-0.205***	-0.061***
	(0.007)	(0.011)
Age^2	0.003***	0.001***
	(0.000)	(0.000)
Years of education	-0.210***	-0.213***
	(0.011)	(0.010)
Black	-0.080**	0.116***
	(0.026)	(0.024)
Having children	0.100***	-0.010
	(0.024)	(0.034)
...under 14	0.067	0.042
	(0.043)	(0.040)
Head of household	-0.201**	0.020
	(0.064)	(0.049)
Living in couple	0.355***	-0.040
	(0.061)	(0.034)
Lone mother	-0.277***	0.095
	(0.053)	(0.056)
Share of formal workers	-0.384***	-0.490***
in the household	(0.035)	(0.034)
Share of individuals	1.139***	0.093
who need care	(0.087)	(0.108)
Unemployment rate	7.973***	6.607***
(regional, by education group)	(0.620)	
Constant	4.415***	1.989***
	(0.305)	(0.312)

Notes: The share of working-age household members holding formal jobs excludes the respondent.

Table 12: Labour market status

Relative risk ratios from the multinomial logit estimation. Base: formal employee

Men	Not working	Informal employee
Age	-0.211***	-0.126***
	(0.008)	(0.008)
Age^2	0.003***	0.001***
	(0.000)	(0.000)
Years of education	-0.144***	-0.139***
	(0.006)	(0.006)
Bleck	-0.011	0.039
	(0.023)	(0.023)
Having children	0.141***	-0.079*
	(0.038)	(0.040)
...under 14	-0.349***	-0.068
	(0.041)	(0.036)
Head of household	-1.062***	-0.518***
	(0.050)	(0.034)
Living in couple	-0.390***	-0.146***
	(0.031)	(0.042)
Share of formal workers	-0.657***	-0.990***
in the household	(0.043)	(0.043)
Share of individuals	-0.118	0.178*
who need care	(0.112)	(0.082)
Unemployment rate	0.117	-2.044*
(regional, by education group)	(0.804)	(0.877)
Constant	4.594***	3.639***
	(0.201)	(0.189)

Notes: The share of working-age household members holding formal jobs excludes the respondent.