Formal-Informal Earnings Differentials in Brazil. A semi-parametric approach.

Maria Tannuri-Pianto [1] Donald M. Pianto [2]

April 15, 2002

[1]Department of Economics, University of Brasília. E-mail: tannuri@unb.br. Address: SQN 205 Bloco C apt. 606, CEP: 70843-030, Brasília-DF, Brazil. Phone: (55) (61) 30393548. Fax: (55) (61) 3402311.

[2]Department of Physics, University of Illinois at Urbana-Champaign.

Abstract

This paper studies the possible existence of sample selection bias for informal and formal employment in Brazil. We use semi-parametric methods that do not rely on the normality assumption and are capable of analyzing individuals at various points of the earnings distribution. We present selectivity corrected quantile regression models for earnings of both informal and formal workers. We find that in the informal sector the unobservable characteristics which cause selection increase the expected income for lower earnings quantiles while lowering those for higher quantiles. We also find that the earnings gaps between formal and informal workers are wider at low conditional quantiles than at high ones. Differences in returns to attributes explain around 30% of their earnings gap at low quantiles, while at high quantiles of the distribution, the gap is completely explained by differences in their individual characteristics.
Keywords: Informal Employment, Selection Bias, Earnings Differentials, Quantile Regression.
JEL Codes: C14, J21.

1 Introduction

Recent studies on wage differentials between formal and informal workers in Brazil (Henley and Carneiro, 2001; and Menezes-Filho, Mendes and Almeida, 2001) find that the differences in earnings are strongly explained by non-observable characteristics of the workers who decide to join each sector.

These results go against the hypothesis of labor market segmentation (Dickens and Lang, 1985), which postulates a dualism in earnings for individuals with similar characteristics, as a consequence of the sector in which they work. Their findings also reinforce a more recent body of work (Funkhouser, 1996; Maloney, 1998, 1999; and Marcoullier *et al.* 1997) which postulates that informal employment may be a desirable choice for workers, due to inefficient labor codes and/or low levels of human capital.

In Brazil, earnings of informal workers are on average lower than those of formal ones, but there is a significant overlap of their earnings distribution (see Figure 1). It is reasonable to assume that the motivations to join the informal sector are also substantially distinct for individuals at different points of the earnings distribution. This is especially true for the ones located at the two tails of the distribution, given that they typically have very different human capital endowments and opportunities. The articles by Carneiro and Henley (2001) and Menezes *et al.* (2001) study the presence of self-selectivity in earnings equations for the mean-earnings workers. Their analyses rely on parametric methods for the conditional mean and on the assumption of the normality of residuals in the participation equation. They are, therefore, incapable of analyzing if being part of the informal sector is really a choice of low-income individuals as well as of high-income ones, or if they just end up in the informal sector, because of the relative abundance of such jobs.

This paper estimates earning functions for informal and formal workers using quantile regression (Koenker and Bassett, 1978), but considering the possibility of self-selectivity into those sectors. We use a methodology developed by Buchinsky (1998) based on semi-parametric methods. The resultant models describe conditional earnings of individuals and returns to their observable characteristics at different points of the informal/formal earnings distribution, correcting for sample selectivity bias.

We find that unobserved characteristics which cause selection into the informal sector have a positive effect on earnings for lower quantiles and a predominantly negative effect for higher quantiles. The estimations also show that earnings of formal workers are higher than those of informal workers throughout the entire distribution, even after controlling for their socio-economic characteristics and correcting for sample selection bias. These earnings gaps are mostly explained by differences in the worker's attributes, with the exception of low income individuals, for whom differences in returns to attributes also play an important role in explaining their gaps.

The paper is organized as follows. Section 2 discusses the econometric methods. Section 3 presents the data. Section 4 describes the empirical findings and Section 5 concludes.

2 Econometric Methods

We control for sample selectivity in a quantile regression model by using a variant of the standard Heckman two-step. We employ a method introduced by Buchinsky (1998) and based on a non-parametric method for mean regression developed by Heckman (1980) and Newey (1991).

For the first step we could use a simple probit model to estimate the probabilities of workers being in the formal or informal sector. However, since we don't wish to impose the restriction of normality, we use a semi-parametric method developed by Ichimura (1993) which makes no assumptions about the distribution of the residuals. To perform this calculation we define a binary variable d such that

$$d \equiv \begin{cases} 1 & : \quad \text{informal} \\ 0 & : \quad \text{formal}, \end{cases} \tag{1}$$

and then assume the existence of a latent or index variable, g, such that

$$d_i = \begin{cases} 1 & : \quad g_i > 0 \\ 0 & : \quad g_i \leq 0. \end{cases} \tag{2}$$

We then set up the following model:

$$g_i = x_{1i}^{'}\gamma + \nu_i, \tag{3}$$

where x_{1i} is a set of characteristics which is thought to determine the likelihood of person i working in the informal sector. We describe the details of how γ is estimated in Section 2.1.

In the second step we assume the existence of an informal wage *offer* equation which represents the wage that a person would receive *if* he worked in the informal sector. In the mean this equation is,

$$y_i^O = x_{2i}^{'}\beta_0 + u_i. \tag{4}$$

Where x_2 is a subset of the characteristics x_1, used in Eq. (3). A quantile regression form of Eq. (4) is:

$$y_i^O = x_{2i}^{'}\beta_\theta + u_{\theta i}, \tag{5}$$

where

$$u_{\theta i} \equiv x_{2i}^{'}(\beta_0 - \beta_\theta) + u_i, \tag{6}$$

and the subscript θ refers to the conditional quantile at which the model is being estimated. In the mean we know that $\text{Mean}(u|x_2) = 0$ for the wage offer. The equivalent statement for the quantile model is that $\text{Quant}_\theta(u_\theta|x_2) = 0$.

However, we only observe the wages of those people who actually work in the informal sector (for whom $g > 0$). If these people are not drawn randomly from the distribution of residuals we say that there is selection bias when we estimate the wage equation. In the mean this generally implies that $\text{Mean}(u|x_2, g > 0) \neq 0$. In quantile regression the implication is that $\text{Quant}_\theta(u_\theta|x_2, g > 0) \neq 0$.

To get unbiased estimates of β_θ in the wage equation (not the wage *offer* equation) it is necessary to introduce an extra term in our model:

$$y \equiv x_2^{'}\beta_\theta + h_\theta(g) + \epsilon_\theta, \tag{7}$$

where

$$h_\theta(g) \equiv \text{Quant}_\theta(u_\theta|x_1, g > 0), \tag{8}$$

such that $\text{Quant}_\theta(\epsilon_\theta|x_1, g > 0) = 0$. This term includes information about the unobservable characteristics of informal workers (note the conditionality on x_1 and not x_2) which affect their choice of sector to work in. Unfortunately, we have no closed form for $h_\theta(g)$ and we must resort to approximations.

In our case we follow the lead of Buchinsky (1998) and approximate $h_\theta(g)$ by a power series whose coefficients are to be estimated by the regression. Ideally the power series expansion should be in a function which is larger when the impact of unobservable characteristics is larger. The smaller the probability of a worker being informal, while still being informal, the larger the role of unobservable characteristics in his choice of sector. A function which mimics this behavior is the inverse Mill's ratio, being small for those with a high probability of being informal and increasing monotonically as the probability of being informal decreases. By expanding $h_\theta(g)$ as a power series in the inverse Mill's ratio in our wage equation we control for the effect of sample selection bias.

2.1 First-step estimation — the SLS estimator

The coefficients γ in Eq. (3) are estimated by the semi-parametric least-squares (SLS) method (Ichimura, 1993):

$$\hat{\gamma} \equiv \arg\min_{\gamma \in \mathcal{R}^p} \frac{1}{n} \sum_{i=1}^{n} (d_i - \hat{\mathrm{E}}(d_i | x_{1i}, \gamma))^2, \tag{9}$$

where

$$\hat{\mathrm{E}}(d_i | x_{1i}, \gamma) = \frac{\sum_{j \neq i} d_j \kappa((x_{1i}^{'}\gamma - x_{1j}^{'}\gamma)/h)}{\sum_{j \neq i} \kappa((x_{1i}^{'}\gamma - x_{1j}^{'}\gamma)/h)} \tag{10}$$

is a kernel density estimate of d_i with kernel function $\kappa(\cdot)$ and its bandwidth h. The estimator is $\sqrt{n}$-consistent and asymptotic normal.

To solve this minimization problem, we obtain a starting value for γ from a probit solution of Eq. (3). Then we use the non-linear-minimization function provided in R (GNU Free Software Foundation) which implements a Newton-like minimization algorithm to solve the equation for $\hat{\gamma}$. Rather than performing the minimization in two steps, each with a different value of the bandwidth, as in Buchinsky (1998), we allow the bandwidth to vary during the minimization and use a Gaussian kernel with bandwidth suggested by Silverman (1986).[1]

We estimate the variance-covariance matrix, $\Lambda_\gamma = \Omega^{-1}\Sigma\Omega^{-1}$, for $\hat{\gamma}$ by using the following equations (Buchinsky, 1998):

$$\begin{aligned}
\hat{\Omega} &= \frac{1}{n} \sum_{i=1}^{n} \hat{f}^2(\hat{g}_i)(x_{1i} - \hat{\mathrm{E}}(x_{1i} | x_{1i}'\hat{\gamma})(x_{1i} - \hat{\mathrm{E}}(x_{1i} | x_{1i}'\hat{\gamma})' \\
\hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^{n} \hat{f}^2(\hat{g}_i)(d_i - \hat{\mathrm{E}}(d_i | x_{1i}, \gamma))^2 (x_{1i} - \hat{\mathrm{E}}(x_{1i} | x_{1i}'\hat{\gamma})(x_{1i} - \hat{\mathrm{E}}(x_{1i} | x_{1i}'\hat{\gamma})'
\end{aligned}$$

where $\hat{f}(g) = \mathrm{d}\hat{F}(g)/\mathrm{d}g$ and $\hat{F}(g) = \hat{\mathrm{E}}(d | x_1, \gamma) = \hat{\mathrm{E}}(d | g)$. The derivative is calculated by first fitting a monotone spline to $\hat{F}(g)$ (Ramsay, 1988) and then differentiating the fitted function.

[1] The bandwidth suggested by Silverman (1986) is $h = 4.24 \min(\sigma, R/1.34) n^{-1/5}$ where R is the interquartile range, n is the sample size, and σ is the standard deviation.

As noted in Buchinsky (1998), since the SLS estimate is consistent, independent of the distribution of ν, and the probit estimate is the efficient estimate under normally distributed ν, a Hausman type test can be performed where

$$\Delta_\gamma'(\hat{\Lambda}_\gamma - V_{P*})^{-1}\Delta_\gamma \to \chi^2(p-2) \quad \text{as } n \to \infty \tag{11}$$

under the null hypothesis. The constant term is not used in the SLS regression and all the parameters cannot be identified in a probit estimation, hence $\hat{\gamma}^{P*} = (\hat{\gamma}_3^P/|\hat{\gamma}_2^P|, \ldots, \hat{\gamma}_p^P/|\hat{\gamma}_2^P|)'$, V_{P*} is its variance, and $\Delta_\gamma = \hat{\gamma} - \hat{\gamma}^{P*}$.

2.2 Second-step estimation — quantile regression

In the second step, the values of $\hat{g} = x_1'\hat{\gamma}$ for the informal workers are used to expand $h(g)$, from Eq. (8), in a power series. Eq. (7) then becomes

$$y_i = x_{2i}'\hat{\beta}_\theta + \sum_{j=1}^{S}(\lambda(\hat{\mu} + \hat{\sigma}\hat{g}_i))^{j-1}\hat{\delta}_{\theta j} + \hat{\epsilon}_{\theta i} \tag{12}$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mill's ratio. We choose the scaling parameters, $\hat{\mu}$ and $\hat{\sigma}$, as the constant and slope coefficients from a probit regression of d_i on the index $\hat{g}(x_{1i}, \hat{\gamma})$. We also choose $S = 3$ for the series expansion.

Unfortunately, the $j = 1$ term in Eq. (12), whose coefficient is $\delta_{\theta 1}$, is equal to one and therefore it cannot be separately identified from the constant term in β_θ unless some additional information is available. However, the slope coefficients and the coefficients for the higher orders in the expansion are identifiable. Therefore, we remove the $j = 1$ term from the series expansion and estimate the resulting quantile model.

We then identify the constant term in the wage equation by using a subset of the data where the probability of informal sector participation is close to one. Following Buchinsky (1998) we partition β_θ into $\beta'_\theta = (\beta_{\theta 1}, \beta^{*\prime}_\theta)$ where $\beta_{\theta 1}$ is the constant coefficient and β^*_θ includes all the slope coefficients. We similarly partition the characteristics into $x'_{2i} = (1, x^{*\prime}_{2i})$ and define $e_i = y_i - x^{*\prime}_{2i}\hat{\beta}^*_\theta$, where $\hat{\beta}^*_\theta$ is the estimate of β^*_θ from Eq. (12) above with the $j = 1$ term removed. Finally, we determine $\hat{\beta}_{\theta 1} = \text{Quant}_\theta(e|\hat{g} > \text{Quant}_{0.95}(\hat{g}))$.

3 Data

The data used in this paper are from the 1999 Brazilian household survey, *Pesquisa Nacional por Amostra de Domicílios, (PNAD)*. We classify a person as an informal worker if she is an employee (domestic or otherwise) and works without a signed labor card. A person is classified as a formal worker if she has a signed labor card and she is not a public employee or a member of the military.

The 1999 PNAD surveyed a total of 206,061 individuals between the ages of 18 and 65. Among those, 153,373 were economically active, but only 138,734 worked during the week of the 19^{th} to 25^{th} of September.

The sample is further restricted to formal and informal employees only, as classified above, with available data on all the covariates, summing to 62,312 observations.[2] Our final sample is composed of 39,900 formal workers and 22,412 informal workers.

Table 1 reports summary statistics for both formal and informal workers, weighted by the PNAD's weights. Log of hourly earnings are used in the estimations. They are calculated by dividing monthly earnings from the main job by 4.33 times the reported hours worked per week. Average hourly earnings for formal workers, R$ 2.15, is almost twice as high as for informal

[2]Note that all employers, self-employed, public employees and members of the military are excluded here.

workers, R$ 1.17. Both groups have similar average years of age and "tenure at job". Formal workers are consistently more educated than informal. Formal workers have on average 7.9 years of education while informal workers have only 5.8 years. This can also be seen by the percentage of workers in each of the education categories. The informal sector has a much higher percentage of "less than elementary" and "illiterate" individuals, while the formal sector has a higher percentage of individuals with "secondary" and "college" education. 62.6% of the formal workers are "white", as opposed to 48.2% of the informal. The percentage of union affiliated workers in the formal sector is 27.9%, while it is only 5.8% at the informal sector.

A high percentage of formal workers, 63.72%, work at establishments with more than 10 employees, while only one third of this percentage of informal workers work in similar establishments. Their means of payment also differ substantially. The great majority of formal workers, 92.7%, get paid by salary only (*jornada*), while a substantial percentage of informal workers get paid by either commission, piece rate, or other contract. The percentage of individuals receiving some sort of housing, food, transport or health aid is considerably different in the two sectors. For example 45.5% and 50.7% of formal workers receive food and transport aid, respectively, while these figures are only 17.5% and 15.3% for the informal workers, respectively.

4 Results

As pointed out in Section 2, the general approach to estimate the quantile regression earnings functions corrected for selectivity bias is based on a two-step estimation. In the first step we estimate the semi-parametric least squares (SLS) model for the probability of being informal. In the second step we estimate the quantile models for the wage equation including the power series expansion in the inverse Mill's ratio.

4.1 Probability Models

Results for the probit and SLS probability models are presented in Table 2. Both the probability and the wage equation models include: 5-segment splines in age with knots at 25, 35, 45 and 55; 5-segment splines in "tenure at job" with knots at 2, 6, 13 and 19, representing the 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of the distribution of years of tenure, respectively; five dummies for education, indicating the achievement of the referred degree, with "less than elementary" being the excluded category; and dummies for gender, ethnicity, union participation, urban area, country regions, industry, and occupation. The probability models are further identified by the inclusion of variables for school attendance, hours worked per week, establishment size, more than one job held, head of household, spouse of head of household, other household income, payment type, and job benefits.

The probit model relies on the normality assumption for the residuals, but this hypothesis is tested and rejected by a *Hausman type test*[3]. Therefore, we use $\hat{\gamma}$ from the SLS model in the second-stage regressions. Both probability models are well fitted and their coefficients have similar signs and magnitudes.

Table 2 demonstrates the following results. The probability of working in the informal sector decreases as age increases from 18 to 25, stabilizes from 25 to 45, and increases above 45. Similarly, as tenure increases from 0 to 2 years the probability of being informal decreases and then stabilizes for tenure greater than two years. There is an increased probability of being informal for those people who are illiterate or hold more than one job. All the education

[3]See note 2 in Table 2 for the statistic and its p-value, and a discussion about the test in Section 2.1.

dummies are significant and show that as the level of education decreases the probability of being informal increases. Females, whites, unionized workers, and workers in establishments with more than 10 employees are less likely to be in the informal sector. Being the head of household or the spouse of the head of household does not have an impact on this probability.[4]

4.2 Selectivity Patterns

In the second-step estimations, quantile regression models for conditional log hourly earnings are estimated at the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of the distribution. Inverse Mill's ratio polynomials, based on the results from the SLS model, are included in the quantile models. The polynomials are of the third order and second degree: $1 + \lambda + \lambda^2$. As pointed out in Section 2.2, the quantile regression model intercepts represent the actual regression intercept plus the polynomial term of degree zero. Therefore, we perform some more estimations, based on a subsample of individuals that are highly likely to be in the informal sector, to disentangle these coefficients and plot the selectivity polynomials.

All the selectivity polynomials are highly significant for the informal workers, with the exception of the linear λ coefficient at the 25^{th} quantile model. This indicates the presence of sample selection bias for individuals throughout the entire earnings distribution. Figure 2 shows the impact of selection bias on informal sector earnings as a function of the index, $\hat{g} = x_1'\hat{\gamma}$, from the SLS model. The impact is negative for individuals at the 95^{th}, 90^{th} and 75^{th} quantiles of the conditional earnings distribution, but positive for individuals at the 50^{th}, 25^{th}, 10^{th} and 5^{th} quantiles. The positive effect at low

[4]The sets of dummies for region, industry and occupation are significant as groups in both models.

and median earnings quantiles indicates that these individuals receive higher earnings than one would expect based on their observable characteristics. This may be a result of specific abilities (such as manual, interpersonal, or artistic skills) not associated with any of the variables in our models. On the other hand, the negative impact at high earnings quantiles shows that high income informal workers earn less than would be expected by their observable characteristics. Perhaps these workers choose the informal sector for the benefit of untaxed earnings.

Figure 3 shows plots of the inverse Mill's ratio expansions for formal workers. Selectivity causes formal workers to earn less than expected based on their observable characteristics. This may be an indication that they choose being formal because they enjoy the safety of other forms of legal compensation, such as advanced notification of termination or employer's social security contributions. Such an interpretation is consistent with the SLS results for age greater than 45. After a worker has secured his retirement earnings in the formal sector, he is increasingly more likely to work in the informal sector. Henley and Carneiro (2001) estimate similar model specifications for the mean-earnings individuals using conventional Heckman two-step procedures and 1997 PNAD data and find positive, significant and very small in magnitude λ coefficients for both informal and formal workers. This indicates the presence of positive selection corrections on earnings in both sectors. Menezes-Filho *et al.* using repeated cross-sections, but also analyzing the mean-earnings individuals, find a positive correlation between non-observable individual characteristics (such as ability, intelligence and quality of education) and employment in the formal sector. Our results for formal workers are substantially different from theirs.

4.3 Formal $\times$ Informal Earnings Functions

We first comment on the informal workers earnings functions, corrected for selectivity bias, which are estimated at seven different quantiles of the earnings distribution (see Table 3). The linear coefficient on age indicates that it has a positive impact on earnings, but this impact is larger for individuals at high quantiles of the conditional earnings distribution. The other coefficients on age indicate the marginal effect of the age splines. The only marginal effect that is highly significant and negative for all quantiles is the first one (age ≥ 25), indicating a decrease in the slope of returns to age after 25 years of age.

The percentage increase in earnings from an extra year of "tenure at job" is around 3% for individuals at low and median conditional quantiles, while it is around 5.5% for individuals at high conditional quantiles. The splines on tenure tend to be more significant and negative at higher quantiles, indicating a downward break in the slope of tenure.

Illiteracy severely hurt informal workers located at high conditional quantiles of the earnings distribution (17% decrease at the 90^{th} and 95^{th} quantiles, and 5% and 8% decreases at the 50^{th} and 75^{th}, respectively). However, a surprising result is that illiteracy does not hurt informal workers at low quantiles of the conditional earnings distribution. Another interesting result is that women earn a lot less than men at high quantiles (-25% and -27% at the 90^{th} and 95^{th} quantiles), than at low quantiles (around -17% at the 5^{th}, 10^{th} and 25^{th} quantiles). A similar pattern, only different in sign, is observed for white informal workers. Their earnings differences relatively to non-whites are higher at high quantiles than at low quantiles. Being a union member only slightly increases the earnings of high quantile informal workers, while it does not affect the earnings of low and median quantile workers. The sets of dummies for region, industry and occupation are significant as groups in all the models.

We now discuss the selectivity corrected earnings function estimates for the formal workers, at different quantiles (see Table 4). The most striking differences between the formal and informal workers estimates—other than returns to education, which we will discuss next—are their rewards/penalties for illiteracy, gender, union affiliation and working in the urban sector. There is a very strong penalty for illiterate individuals in the formal sector throughout all quantiles, while in the informal sector this is only true for individuals at high quantiles. In addition, women are hurt more severely at high quantiles of the formal sector earnings distribution than they are at similar quantiles of the informal sector distribution. A woman's expected earnings at the 90^{th} and 95^{th} percentiles are approximately 30% lower than a man's. Being a union member has a positive and significant effect on earnings of formal workers throughout all quantiles and low income individuals receive a higher premium for union affiliation than high income ones. Finally, working in the urban sector benefits formal workers in a very similar way throughout the quantiles, while for the informal workers the premium increases with quantile and it is almost negligible at low quantiles.

4.4 Formal $\times$ Informal Returns to Education

Table 5 summarizes the results for returns to education in both informal and formal sectors, at the different quantiles. The first interesting result to note is that the returns to an elementary school degree are lower for formal workers than for informal workers at all analyzed quantiles. As one would expect, informal workers at the 5^{th} percentile of the earnings distribution are the ones who benefit the most from getting an elementary school degree.

The returns to a primary school degree are very similar in magnitude for both informal and formal workers, but they increase with the quantiles. In the informal sector these returns vary from 9.9% at the 5^{th} quantile to 23.9% at the 95^{th} quantile. The same pattern is observed for the secondary school degree, but the returns are higher in magnitude.

The returns for getting a college degree are very high. Low and median income individuals in the formal sector receive higher returns for a college degree than similar individuals in the informal sector. At high quantiles, informal workers benefit more than formal ones from this degree. The individuals that benefit the most from going to college are located at the 75^{th} and 90^{th} quantiles of the earnings distributions, in both informal and formal sectors.

4.5 Formal × Informal Earnings Gaps

A little bit more of notation is introduced here to facilitate the understanding of Table 6. We are interested in estimating the difference in the log hourly wages between the formal and informal sectors:

$$\ln w_f^\theta - \ln w_i^\theta = \bar{X}'_f \hat{\beta}_f^\theta - \bar{X}'_i \hat{\beta}_i^\theta \tag{13}$$

where $\ln w_k^\theta$ is the log hourly wages in sector k evaluated at quantile θ, $\hat{\beta}_k^\theta$ is a vector of estimated coefficients for sector k evaluated at quantile θ, $\bar{X}_k$ is a vector of median[5] characteristics of workers in sector k, and $k = f, i$ denotes the formal and informal sectors, respectively.

The earnings differentials can be decomposed into a first term representing differences in endowments between formal and informal sector employees and referred to as the covariates differential, and a second term representing

[5]Except for the dummy variables, which are evaluated at their mean values.

sectoral differences in returns to these endowments, called the coefficients differential:

$$\bar{X}_f'\hat{\beta}_f^\theta - \bar{X}_i'\hat{\beta}_i^\theta = (\bar{X}_f - \bar{X}_i)'\hat{\beta}_f^\theta + \bar{X}_i'(\hat{\beta}_f^\theta - \hat{\beta}_i^\theta). \tag{14}$$

Or, using the reversed order decomposition:

$$\bar{X}_f'\hat{\beta}_f^\theta - \bar{X}_i'\hat{\beta}_i^\theta = (\bar{X}_f - \bar{X}_i)'\hat{\beta}_i^\theta + \bar{X}_f'(\hat{\beta}_f^\theta - \hat{\beta}_i^\theta). \tag{15}$$

The top panel of Table 6 shows conditional earnings estimates for the models that are not corrected for sample selection bias, while the bottom panel shows similar estimates for the models with sample selection correction. It is worth noting that these earnings are calculated for some representative median characteristic individual, that does not necessarily exist in the data set. The model with sample selection correction includes a polynomial in the inverse Mill's ratio and unbiased β coefficients. The first two columns which decompose the earnings gaps between the impact of covariates and coefficients are based on Eq. (14) and the two subsequent columns are based on Eq. (15). When calculating earnings of a informal worker as if she was paid according to the formal sector payment scheme[6] for the model with selectivity correction, we use the inverse Mill's ratio that a formal worker with the same characteristics would have.

Table 6 shows that even after controlling for socio-economic characteristics, the earnings gaps are substantially large. One sees in the bottom panel that, at the 5^{th} quantile, a formal worker earns 96% more than an informal worker. At higher quantiles the gaps are smaller, dropping to 57% at the 95^{th} percentile. Menezes *et al.* (p. 856) claim that after controlling for education the earnings in the informal sector are higher than those in the formal sector. However, by looking at the columns titled "Coefficients" in Table 6 one

[6] Using the β's from the formal sector.

sees that, evaluating the formal and informal earnings functions at the same characteristics, formal earnings are greater than informal earnings, except at the highest quantiles.

In our models differences in covariates practically explain the entire earnings gap for individuals at high quantiles. However, differences in coefficients (returns to covariates) explain a significant proportion of the gaps at median and lower quantiles (*e.g.* 36% at the 5^{th} quantile using the first decomposition). This demonstrates that high wage individuals in the informal sector earn less because they are less skilled, while the low income ones, besides being less skilled, get lower returns to their skills as well. These findings seem to corroborate the hypothesis of segmentation of the labor markets previously reported for Brazil (see Fernandes, 1996; Pero, 1992; Cacciamali and Fernandes, 1993).

5 Conclusion

This paper investigates if there is sample selection bias in informal and formal employment in Brazil. We use semi-parametric methods that do not rely on the normality assumption of residuals and are capable of analyzing individuals at various points of the earnings distribution. We present selectivity corrected quantile regression models for earnings of both informal and formal workers.

The most striking differences between the formal and informal workers earnings functions estimates are their rewards/penalties for illiteracy, gender and union affiliation. There is a very strong penalty for illiterate individuals in the formal sector throughout all quantiles, while in the informal sector this is only true for individuals at high quantiles. In addition, women are hurt much more severely at high quantiles of the formal sector earnings distribution than they are at similar quantiles of the informal sector distribution. Being a union member has a positive and significant effect on earnings of formal workers throughout all quantiles, with low income individuals receiving a higher premium for union affiliation than high income ones. Only informal workers at high quantiles receive some benefit from union affiliation.

The returns to getting an extra degree of education vary substantially for informal and formal workers. As an illustration of this pattern, we observe that low and median income individuals in the formal sector receive higher returns for a college degree than similar individuals in the informal sector. At high quantiles, informal workers benefit more than formal ones from this degree. On the other hand, informal workers are the ones who benefit the most from getting an elementary school degree.

An important result is that the unobserved characteristics of people in the informal sector (those characteristics which cause them to be in the informal sector even though the probit and SLS models give significant probabilities for them to be in the formal sector) also impact their earnings. These unobservable characteristics increase the expected income for lower earnings quantiles while lowering those for higher quantiles. Selection also affects earnings in the formal sector, but here the unobservable characteristics decrease the expected income for all earnings quantiles. These findings contradict the previous literature.

Another important result is that informal workers earn much less than the formal ones, even after controlling for their characteristics. High earning individuals in the informal sector earn less because they are on average less skilled than formal workers, while the low earning ones, besides being less skilled, receive lower returns to their skills as well. It is very difficult to deny the hypothesis of segmentation in the labor markets based on these findings. Individuals at the bottom of the informal sector earnings distribution earn substantially less than formal workers, they are less skilled than formal workers at similar positions of the earnings distribution, they receive lower returns to their skills, but they still receive positive rewards to their unobservable characteristics when in the informal sector. It seems to be the case that there is no space for these individuals in the formal sector.

References

Buchinsky, M. (1998): "The dynamics of changes in the femal wage distribution in the USA: A quantile regression approach," *Journal of Applied Econometrics*, 13, 1–30.

Cacciamali, M. C., and R. Fernandes (1993): "Distribuição dos trabalhadores e diferencias de salários entro os mercados de trabalho regulamentado e não-regulamentado," *Pesquisa Planejamento Econômico*, 23(1), 135–156.

Dickens, W., and K. Lang (1985): "A test of dual labour market theory," *American Economic Review*, 75(4), 1–22.

Fernandes, R. (1996): "Mercado de trabalho não-regulamentado: participação relativa e diferenciais de salarios," *Pesquisa e Planejamento Econômico*, 26(3), 417–442.

Funkhouser, E. (1996): "The urban informal sector in Central America: Household survey evidence," *World Development*, 24(11), 1737–51.

Heckman, J. (1980): "Sample selection bias as a specification error," in *Evaluation Studies*, ed. by E. Stromsdorfer, and G. Farkas, vol. 5 of *Review Annual*, pp. 61–74. Sage, Beverly Hills.

Henley, A., and F. G. a. Carneiro (2001): "Modelling Formal vs. Informal employment and earnings: microeconometric evidence for Brazil.," *Annals of the XXIX National Meeting of Economics of ANPEC.*

Ichimura, H. (1993): "Semiparametric least-squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58, 71–120.

Koenker, R., and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

Maloney, W. (1998): "Are LDC Labor Markets Dualistic?," Working paper, World Bank.

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

Marcoullier, D., V. Ruiz de Casilla, and C. Woodruff (1997): "Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru," *Economic Development and Cultural Change*, 45(2), 367–92.

Menezes-Filho, N. A., M. Mendes, and E. S. o. Almeida (2001): "O diferencial de salários formal - informal: segmentação ou viés de seleção?," *XXIII Meeting of the Brazilian Econometrics Society.*

Newey, W. (1991): "Two step series estimation of sample selection model," MIT.

Pero, V. L. (1992): "A carteira de trabalho no mercado de trabalho metropolitano brasileiro," *Pesquisa Planejamento Econômico*, 22(2), 305–342.

Ramsay, J. O. (1988): "Monotone regression splines in action," *Statistical Science*, 3, 425–461.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

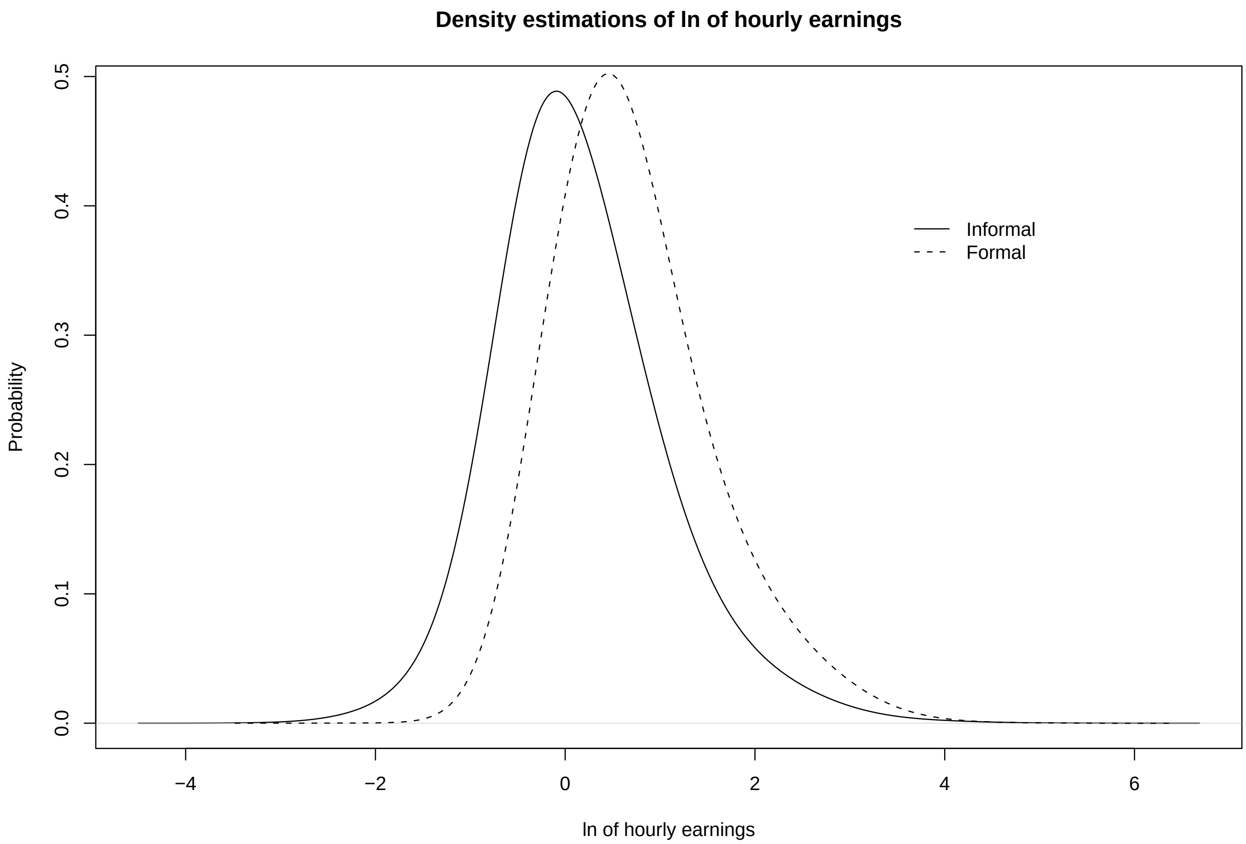

Figure 1: Density of the log of hourly wages for the formal and informal sector.

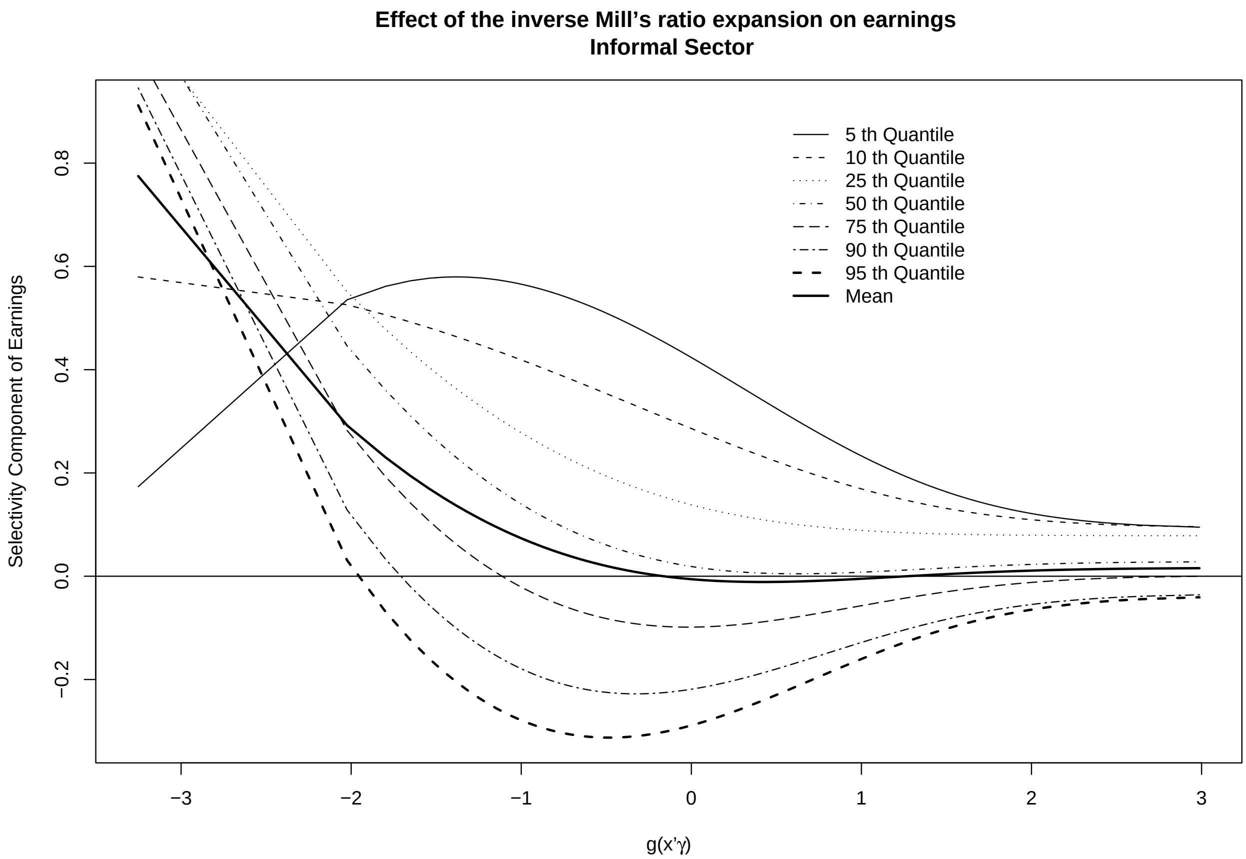

Figure 2: The effect of unobservable variables on the income of informal workers at different conditional quantiles of the wage distribution as a function of $\hat{g}$.

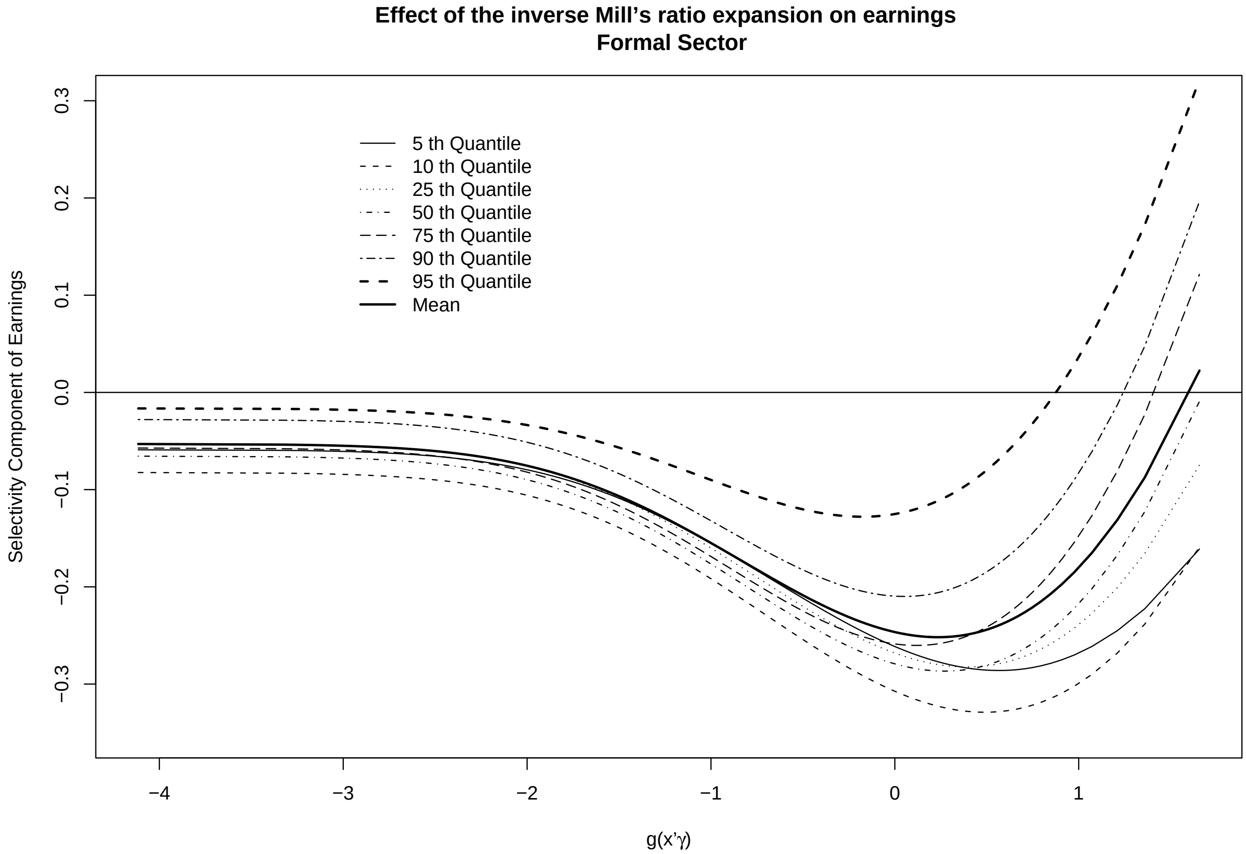

Figure 3: The effect of unobservable variables on the income of formal workers at different conditional quantiles of the wage distribution as a function of $\hat{g}$.

Table 1: **Summary Statistics.** Weighted sample averages and standard errors.

	Formal Workers		Informal Workers	
	Mean	Std. error	Mean	Std. error
Log (hourly earnings)	0.7622	0.0044	0.1586	0.0061
Age (years)	33.7476	0.0595	32.7099	0.0861
Tenure at job (years)	5.2665	0.0343	4.0631	0.0473
Education (years)	7.9343	0.0230	5.7895	0.0313
Less than elementary	0.1424	0.0019	0.3309	0.0034
Elementary	0.2837	0.0025	0.3069	0.0034
Primary	0.1948	0.0022	0.1511	0.0026
Secondary	0.2957	0.0026	0.1709	0.0027
College	0.0798	0.0015	0.0387	0.0014
Masters	0.0036	0.0003	0.0014	0.0003
Attending school	0.1053	0.0017	0.1243	0.0024
Illiterate	0.0409	0.0011	0.1602	0.0026
Female	0.3416	0.0027	0.2858	0.0033
White	0.6259	0.0027	0.4820	0.0037
Union	0.2790	0.0025	0.0580	0.0017
Urban	0.9019	0.0016	0.7643	0.0031
Hours worked per week	44.4558	0.0519	42.8356	0.0946
More than 10 co-workers	0.6372	0.0027	0.2090	0.0030
More than 1 job held	0.0309	0.0009	0.0483	0.0015
Head of household	0.5195	0.0028	0.4768	0.0036
Spouse head of household	0.1554	0.0020	0.1375	0.0025
Other hh income (R$/month)	697.87	6.4339	581.05	8.2565
Paid by piece rate	0.0022	0.0003	0.0522	0.0016
Paid by commission	0.0243	0.0009	0.1383	0.0025
Paid by salary only (*jornada*)	0.9270	0.0014	0.7644	0.0031
Paid by commission & *jornada*	0.0448	0.0011	0.0251	0.0011
Paid by other contract	0.0018	0.0002	0.0200	0.0011
Housing aid	0.0588	0.0013	0.0843	0.0020
Food aid	0.4553	0.0028	0.1747	0.0027
Transport aid	0.5074	0.0028	0.1534	0.0026
Health aid	0.2717	0.0025	0.0301	0.0013
Sample Size	39,900		22,412	

Table 2: **Selection Model – informal employment**

	SLS Model		**Probit Model**	
Variables	Coefficient	Std. Error	Coefficient	Std. Error
Intercept			2.0960	0.1232***
Age	−0.0364	0.0055***	−0.0423	0.0044***
Age $\geq$ 25 (mg. effect)	0.0373	0.0074***	0.0435	0.0061***
Age $\geq$ 35 (mg. effect)	−0.0042	0.0058	−0.0029	0.0049
Age $\geq$ 45 (mg. effect)	0.0137	0.0075*	0.0160	0.0062***
Age $\geq$ 55 (mg. effect)	0.0000	0.0125	0.0000	0.0101
Tenure	−0.2322	0.0168***	−0.2428	0.0096***
Tenure $\geq$ 2 (mg. effect)	0.2102	0.0204***	0.2213	0.0145***
Tenure $\geq$ 6 (mg. effect)	0.0103	0.0129	0.0118	0.0112
Tenure $\geq$ 13 (mg. effect)	−0.0035	0.0147	−0.0039	0.0129
Tenure $\geq$ 19 (mg. effect)	0.0438	0.0138***	0.0423	0.0117***
Elementary school	−0.0387	0.0245	−0.0450	0.0216**
Primary school	−0.1241	0.0297***	−0.1324	0.0252***
Secondary school	−0.1924	0.0318***	−0.1766	0.0259***
College degree	−0.1646	0.0473***	−0.0997	0.0362***
Masters/Doctorate	−0.2170	0.1586	−0.2257	0.1174*
Attending school	0.0452	0.0257*	0.0525	0.0219**
Illiterate	0.2351	0.0344***	0.2424	0.0285***
Female	−0.0794	0.0222***	−0.0946	0.0181***
White	−0.0595	0.0169***	−0.0436	0.0141***
Union	−0.5659	0.0378***	−0.7139	0.0212***
Urban	−0.0309	0.0253	−0.0508	0.0212**
Hours worked per week	−0.0088	0.0009***	−0.0083	0.0006***
More than 10 co-workers	−0.7132	0.0410***	−0.7650	0.0149***
More than 1 job held	0.1568	0.0404***	0.2177	0.0326***
Head of household	−0.0009	0.0215	−0.0009	0.0181
Spouse head of household	0.0406	0.0280	0.0633	0.0234***
Other hh income (R$/month)	0.0000	0.0000 ***	0.0000	0.0000***
Paid by piece rate	1.3041	0.1012***	1.5208	0.0678***
Paid by commission	1.1711	0.0783***	1.0555	0.0376***
Paid by salary only	0.0892	0.0406**	0.0578	0.0313*
Paid by other contract	1.3534	0.1355***	1.2914	0.0809***
Housing aid	−0.4188	0.0415***	−0.2936	0.0246***
Food aid	−0.1983	0.0211***	−0.2025	0.0157***
Transport aid	−0.6822	0.0397***	−0.6842	0.0147***
Health aid	−0.5278	0.0416***	−0.6294	0.0236***

Note 1: 63,312 obs, ***, **, * - denotes significance at 1%, 5% and 10%, respectively. Both models include 2-digit industry dummies (10), 2-digit occupation dummies (7) and regional dummies (4). *Note 2*: The *Hausman test* statistic for normality of the probit residuals is $204.52 \sim \chi^2_{55df}$, with p-value = 0.00.

Table 3: **Quantile Regression, informal employment with selectivity correction.** (Log hourly earnings)

Variable	5th quantile	10th quantile	25th quantile	50th quantile	75th quantile	90th quantile	95th quantile
Intercept	−1.4409 (0.1258)***	−1.2744 (0.0906)***	−0.8635 (0.0785)***	−0.5022 (0.0728)***	−0.2115 (0.0909)**	0.1327 (0.1176)	0.2644 (0.1311)**
λ^1	0.5298 (0.0707)***	0.2699 (0.0438)***	0.0141 (0.0348)	−0.1050 (0.0291)***	−0.2460 (0.0333)***	−0.3809 (0.0516)***	−0.4850 (0.0592)***
λ^2	−0.1441 (0.0340)***	−0.0375 (0.0208)*	0.0765 (0.0175)***	0.1167 (0.0133)***	0.1518 (0.0154)***	0.1874 (0.0277)***	0.2146 (0.0297)***
Age	0.0258 (0.0047)***	0.0300 (0.0035)***	0.0283 (0.0028)***	0.0317 (0.0026)***	0.0342 (0.0032)***	0.0361 (0.0043)***	0.0457 (0.0053)***
Age≥ 25	−0.0143 (0.0074)*	−0.0219 (0.0050)***	−0.0182 (0.0042)***	−0.0218 (0.0039)***	−0.0190 (0.0047)***	−0.0182 (0.0063)***	−0.0259 (0.0076)***
Age≥ 35	−0.0154 (0.0059)***	−0.0062 (0.0048)	−0.0060 (0.0038)	−0.0044 (0.0036)	−0.0110 (0.0040)***	−0.0100 (0.0053)*	−0.0118 (0.0059)**
Age≥ 45	0.0039 (0.0087)	−0.0023 (0.0066)	−0.0064 (0.0051)	−0.0079 (0.0050)	−0.0072 (0.0052)	−0.0001 (0.0075)	−0.0002 (0.0082)
Age≥ 55	−0.0287 (0.0192)	−0.0278 (0.0072)***	−0.0085 (0.0079)	−0.0080 (0.0072)	−0.0065 (0.0080)	−0.0337 (0.0105)***	−0.0236 (0.0168)
Tenure	0.0301 (0.0114)*	0.0350 (0.0089)***	0.0223 (0.0071)***	0.0345 (0.0066)***	0.0501 (0.0083)***	0.0597 (0.0110)***	0.0568 (0.0133)***
Tenure≥ 2	−0.0299 (0.0171)*	−0.0418 (0.0141)***	−0.0106 (0.0113)	−0.0247 (0.0101)**	−0.0339 (0.0126)***	−0.0372 (0.0169)**	−0.0360 (0.0208)*
Tenure≥ 6	−0.0207 (0.0134)	−0.0122 (0.0117)	−0.0282 (0.0100)***	−0.0186 (0.0084)**	−0.0269 (0.0093)***	−0.0331 (0.0129)**	−0.0342 (0.0165)**
Tenure≥ 13	0.0181 (0.0179)	0.0298 (0.0156)*	0.0348 (0.0121)***	0.0165 (0.0105)	0.0149 (0.0103)	0.0225 (0.0156)	0.0228 (0.0192)
Tenure≥ 19	−0.0038 (0.0203)	−0.0210 (0.0142)	−0.0302 (0.0087)***	−0.0156 (0.0079)**	−0.0120 (0.0085)	−0.0237 (0.0127)*	−0.0133 (0.0141)
Elementary	0.1491 (0.0318)***	0.1152 (0.0165)***	0.0990 (0.0155)***	0.0860 (0.0134)***	0.0920 (0.0151)***	0.0740 (0.0207)***	0.1107 (0.0235)***
Primary	0.2433 (0.0357)***	0.2377 (0.0210)***	0.2242 (0.0189)***	0.2373 (0.0173)***	0.2675 (0.0201)***	0.2930 (0.0304)***	0.3253 (0.0261)***
Secondary	0.4054 (0.0343)***	0.3580 (0.0244)***	0.3827 (0.0215)***	0.4495 (0.0200)***	0.5519 (0.0239)***	0.6301 (0.0363)***	0.6705 (0.0372)***
College	0.8958 (0.0997)***	0.9217 (0.0352)***	1.0215 (0.0417)***	1.1270 (0.0374)***	1.3768 (0.0557)***	1.4585 (0.0466)***	1.4056 (0.0733)***
Masters	1.6940 (0.0525)***	1.8466 (0.0526)***	1.6942 (0.0501)***	2.0279 (0.1474)***	2.1589 (0.0722)***	1.8237 (0.1109)***	1.6022 (0.0504)***
Illiterate	0.0056 (0.0419)	0.0095 (0.0276)	−0.0208 (0.0173)	−0.0525 (0.0138)***	−0.0846 (0.0159)***	−0.1802 (0.0203)***	−0.1864 (0.0217)***
Female	−0.1863 (0.0195)***	−0.1828 (0.0145)***	−0.1890 (0.0119)***	−0.2214 (0.0115)***	−0.2373 (0.0132)***	−0.2937 (0.0184)***	−0.3176 (0.0235)***
White	0.0829 (0.0179)***	0.0971 (0.0130)***	0.0943 (0.0103)***	0.1085 (0.0100)***	0.1015 (0.0113)***	0.1170 (0.0160)***	0.1555 (0.0209)***
Union	−0.0116 (0.0328)	0.0155 (0.0354)	0.0347 (0.0262)	0.0271 (0.0194)	0.0528 (0.0178)***	0.0629 (0.0336)*	0.0545 (0.0194)***
Urban	0.0231 (0.0261)	0.0453 (0.0157)***	0.0652 (0.0132)***	0.0644 (0.0116)***	0.0812 (0.0128)***	0.1217 (0.0173)***	0.1451 (0.0189)***

Note: 22,412 obs, ***, **, * - denotes significance at 1%, 5% and 10%, respectively. All models include 2-digit industry dummies (10), 2-digit occupation dummies (7) and regional dummies (4).

Table 4: **Quantile Regression, formal employment with selectivity correction.** (Log hourly earnings)

Variable	5th quantile	10th quantile	25th quantile	50th quantile	75th quantile	90th quantile	95th quantile
Intercept	−0.8626 (0.0737)***	−0.7433 (0.0715)***	−0.5405 (0.0672)***	−0.2873 (0.0620)***	−0.0701 (0.0781)	0.2310 (0.1085)**	0.3035 (0.1268)**
λ^1	−0.3815 (0.0276)***	−0.4353 (0.0340)***	−0.4325 (0.0267)***	−0.4523 (0.0236)***	−0.4645 (0.0255)***	−0.4378 (0.0333)***	−0.3234 (0.0541)***
λ^2	0.1603 (0.0152)***	0.1921 (0.0225)***	0.2035 (0.0163)***	0.2313 (0.0139)***	0.2656 (0.0138)***	0.2633 (0.0136)***	0.2347 (0.0336)***
Age	0.0240 (0.0022)***	0.0266 (0.0027)***	0.0302 (0.0024)***	0.0308 (0.0022)***	0.0338 (0.0027)***	0.0377 (0.0038)***	0.0418 (0.0045)***
Age≥ 25	−0.0156 (0.0035)***	−0.0176 (0.0038)***	−0.0177 (0.0033)***	−0.0153 (0.0030)***	−0.0158 (0.0036)***	−0.0185 (0.0053)***	−0.0199 (0.0065)***
Age≥ 35	−0.0048 (0.0033)	−0.0047 (0.0029)	−0.0076 (0.0026)***	−0.0104 (0.0024)***	−0.0102 (0.0027)***	−0.0053 (0.0040)	−0.0072 (0.0053)
Age≥ 45	−0.0071 (0.0042)*	−0.0084 (0.0042)**	−0.0108 (0.0036)**	−0.0118 (0.0034)***	−0.0146 (0.0040)***	−0.0237 (0.0051)***	−0.0163 (0.0067)***
Age≥ 55	−0.0163 (0.0053)***	−0.0141 (0.0060)*	−0.0073 (0.0071)	−0.0040 (0.0057)	−0.0047 (0.0087)	−0.0003 (0.0085)	−0.0124 (0.0140)
Tenure	0.0221 (0.0056)***	0.0259 (0.0057)***	0.0195 (0.0056)***	0.0233 (0.0051)***	0.0229 (0.0063)***	0.0278 (0.0087)***	0.0433 (0.0117)***
Tenure≥ 2	−0.0074 (0.0085)	−0.0087 (0.0082)	0.0045 (0.0080)	0.0049 (0.0073)	0.0062 (0.0088)***	0.0000 (0.0120)	−0.0241 (0.0168)
Tenure≥ 6	−0.0017 (0.0070)	−0.0038 (0.0071)	−0.0099 (0.0056)*	−0.0146 (0.0052)***	−0.0117 (0.0056)**	−0.0116 (0.0083)	0.0051 (0.0115)
Tenure≥ 13	−0.0165 (0.0101)	−0.0066 (0.0099)	−0.0021 (0.0066)	0.0048 (0.0060)	0.0095 (0.0073)	0.0119 (0.0099)	−0.0036 (0.0115)
Tenure≥ 19	0.0086 (0.0124)	−0.0047 (0.0094)	−0.0062 (0.0071)	−0.0104 (0.0058)*	−0.0236 (0.0107)**	−0.0109 (0.0075)	−0.0084 (0.0100)
Elementary	0.0697 (0.0138)***	0.0735 (0.0109)***	0.0821 (0.0115)***	0.0922 (0.0107)***	0.0813 (0.0128)***	0.0668 (0.0177)***	0.0725 (0.0198)***
Primary	0.1718 (0.0155)***	0.1734 (0.0131)***	0.1881 (0.0134)***	0.2264 (0.0121)***	0.2484 (0.0150)***	0.2594 (0.0215)***	0.2726 (0.0248)***
Secondary	0.3196 (0.0156)***	0.3379 (0.0156)***	0.3941 (0.0142)***	0.4618 (0.0130)***	0.5342 (0.0153)***	0.5973 (0.0215)***	0.6575 (0.0274)***
College	0.8827 (0.0306)***	0.9425 (0.0285)***	1.0895 (0.0244)***	1.2011 (0.0202)***	1.2919 (0.0224)***	1.3589 (0.0317)***	1.3725 (0.0381)***
Masters	1.1214 (0.3727)***	1.3651 (0.3268)***	1.5173 (0.0721)***	1.6491 (0.0679)***	1.6915 (0.1024)***	1.7947 (0.1168)***	1.7210 (0.0775)***
Illiterate	−0.0975 (0.0153)***	−0.0862 (0.0153)***	−0.1020 (0.0154)***	−0.1115 (0.0168)***	−0.1502 (0.0231)***	−0.1836 (0.0207)***	−0.2428 (0.0315)***
Female	−0.1919 (0.0095)***	−0.2126 (0.0089)***	−0.2489 (0.0082)***	−0.2959 (0.0076)***	−0.3306 (0.0088)***	−0.3608 (0.0125)***	−0.3669 (0.0163)***
White	0.1018 (0.0087)***	0.1119 (0.0083)***	0.1176 (0.0077)***	0.1206 (0.0070)***	0.1131 (0.0084)***	0.1304 (0.0113)***	0.1592 (0.0151)***
Union	0.1041 (0.0105)***	0.0836 (0.0102)***	0.0641 (0.0087)***	0.0671 (0.0083)***	0.0551 (0.0094)***	0.0509 (0.0125)***	0.0509 (0.0167)***
Urban	0.0809 (0.0123)***	0.0829 (0.0143)***	0.0767 (0.0117)***	0.0826 (0.0105)***	0.0847 (0.0131)***	0.0754 (0.0191)***	0.0870 (0.0185)***

Note: 39,900 obs, ***, **, * - denotes significance at 1%, 5% and 10%, respectively. All models include 2-digit industry dummies (10), 2-digit occupation dummies (7) and regional dummies (4).

Table 5: **Returns to education in the formal and informal sectors.**

Informal	5th quantile	10th quantile	25th quantile	50th quantile	75th quantile	90th quantile	95th quantile
Elementary	16.1 3.8	12.2 2.9	10.4 2.5	9.0 2.2	9.6 2.3	7.7 1.9	11.7 2.8
Primary	9.9 2.4	13.0 3.1	13.3 3.2	16.3 3.9	19.2 4.5	24.5 5.6	23.9 5.0
Secondary	17.6 6.9	12.8 4.1	17.2 5.4	23.6 7.3	32.9 9.9	40.1 11.9	41.2 12.2
College	63.3 13.0	75.7 15.1	89.4 17.3	96.9 18.5	128.2 22.9	129.0 23.0	108.6 20.2
Masters	122.2 49.0	152.2 58.5	96.0 40.0	146.2 56.9	118.6 47.9	44.1 20.0	21.7 10.3
Formal							
Elementary	7.2 1.8	7.6 1.9	8.6 2.1	9.7 2.3	8.5 2.1	6.9 1.7	7.5 1.8
Primary	10.7 2.6	10.5 2.5	11.2 2.7	14.4 3.4	18.2 4.3	21.1 4.9	22.1 5.1
Secondary	15.9 5.1	17.9 5.6	22.9 7.1	26.5 8.2	33.1 10.0	40.2 11.9	46.9 13.7
College	75.6 20.6	83.0 22.3	100.4 26.1	109.4 27.9	113.3 28.7	114.2 28.9	104.4 26.9
Masters	27.0 12.7	52.6 23.5	53.4 23.8	56.5 25.1	49.1 22.1	54.6 24.3	41.7 19.0

Note: The top number represents the percentage increase in income that one receives from the given level of schooling relative to the level below. For 'Elementary' the level below is no degree. The under-set number gives the returns received for each additional year of schooling assuming that each year of additional study from the level below yields the same return.

Table 6: **Earnings Differentials – Formal × Informal, by quantile.**

Without Sample Selection Correction							
	(A)	(B)	(A) - (B)	Covariates	Coefficients	Covariates	Coefficients
	$\bar{X}_f'\hat{\beta}_f$	$\bar{X}_i'\hat{\beta}_i$		$(\bar{X}_f - \bar{X}_i)'\hat{\beta}_f$	$\bar{X}_i'(\hat{\beta}_f - \hat{\beta}_i)$	$(\bar{X}_f - \bar{X}_i)'\hat{\beta}_i$	$\bar{X}_f'(\hat{\beta}_f - \hat{\beta}_i)$
5%	−0.0445	−0.6993	0.6548	0.3541	0.3007	0.3611	0.2937
10%	0.1149	−0.4911	0.6060	0.3732	0.2328	0.3281	0.2779
25%	0.3948	−0.1897	0.5845	0.3942	0.1903	0.3308	0.2537
50%	0.7097	0.1629	0.5468	0.4174	0.1294	0.3731	0.1738
75%	1.0350	0.5353	0.4997	0.4372	0.0625	0.4354	0.0643
90%	1.3675	0.8918	0.4757	0.4486	0.0271	0.4793	−0.0036
95%	1.5968	1.1527	0.4440	0.4526	− 0.0085	0.4791	−0.0351
Mean	0.7306	0.1791	0.5515	0.4216	0.1299	0.4115	0.1400

With Sample Selection Correction							
	(A)	(B)	(A) - (B)	Covariates	Coefficients	Covariates	Coefficients
	$\bar{X}_f'\hat{\beta}_f$	$\bar{X}_i'\hat{\beta}_i$		$(\bar{X}_f - \bar{X}_i)'\hat{\beta}_f$	$\bar{X}_i'(\hat{\beta}_f - \hat{\beta}_i)$	$(\bar{X}_f - \bar{X}_i)'\hat{\beta}_i$	$\bar{X}_f'(\hat{\beta}_f - \hat{\beta}_i)$
5%	−0.0359	−0.7097	0.6738	0.4312	0.2425	0.4826	0.1911
10%	0.1248	−0.5019	0.6267	0.4605	0.1662	0.4488	0.1779
25%	0.4099	−0.2170	0.6270	0.4701	0.1569	0.4682	0.1588
50%	0.7081	0.1322	0.5760	0.4852	0.0907	0.4865	0.0894
75%	1.0279	0.5077	0.5201	0.4847	0.0354	0.5109	0.0092
90%	1.3514	0.8614	0.4900	0.4809	0.0091	0.5323	−0.0423
95%	1.5786	1.1249	0.4538	0.4569	−0.0032	0.5124	−0.0587
Mean	0.7269	0.1594	0.5675	0.4795	0.0880	0.4893	0.0782