

Heterogeneity in the Returns to Education and Informal Activities

Marcelo Arbex*, Antonio F. Galvao, Jr.[†] and Fábio Gomes[‡]

August 2, 2010

Abstract

This paper investigates the effects of schooling on agent's informal earnings. We present a two-period model of informal activities and schooling, and test its implications using a survey of informal workers in Brazil. The model implies that, at least for high skilled workers, we should observe an education premium in the informal sector. When we introduce a non-trivial interaction between education and unobserved ability, the model also predicts that schooling variable is endogenous and its effect on earnings become heterogeneous, reflecting the distribution of this unobservable ability. Instrumental variable quantile regression approach is able to accommodate all these possibilities and, for this reason, we apply this technique. By means of Kolmogorov-Smirnov test we evaluate if there is an education premium in the informal sector, formally test if the education premium varies along the quantiles, and also test whether the education is endogenous. The two last features might be viewed as reflecting the distribution of unobservable ability. Indeed, we find that returns to schooling are far from zero, being positive and monotonically increasing along the quantiles of the earnings distribution. Finally, there is strong evidence suggesting that education is endogenous.

Keywords: Returns to Education; Informal Sector; Quantile Regression.

JEL Classification: C14; J24; J42; O17.

*Department of Economics, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4, Canada. *Email: arbex@uwindsor.ca.* † Department of Economics, University of Wisconsin-Milwaukee. 3210 N. Maryland Avenue, Milwaukee, WI 53201, United States. *Email: agalvao@uwm.edu.* ‡ Insper Institute. R. Quatá, 300, São Paulo, SP 04546-042, Brazil. *Email: fabioarg@insper.org.br.* We would like to thank Pedro de Sant'Anna, Marina D. Carvalho, Scott Drewianka, Todd Elder, Gabriel Montes-Rojas for their valuable comments and suggestions. The usual disclaimers apply.

1 Introduction

The informal sector is often defined as a sector that emphasizes small-scale, self-financed and unskilled labor intensive economic activities. In developing economies, workers employed in this sector tend to be younger, have less education, and earn less than their counterparts in the formal sector (Thomas, 1992; Maloney, 1999). Maloney (2004) presents evidence that Argentine and Mexican workers are less likely to leave formal employment for informal employment, as education levels increase. Nevertheless, there is a clear evidence of an educational heterogeneity with a significant proportion of workers with tertiary education engaged in informal activities. For instance, Amaral and Quintin (2006), Pratap and Quintin (2006) and Paula and Scheinkman (2010) find that at least ten percent of the informal labor force in Argentina and Brazil has post-secondary education.

While it should not be expected that more educated workers would face any difficulties or barriers to enter the formal labor market and compete for higher paying formal jobs, their engagement in informal activities is puzzling. In fact, the presence of highly educated workers in the informal sector could be indeed voluntary. Informal jobs may be a desirable alternative with more flexibility, a way to avoid government regulations, high tax rates and harassment by the tax administration. Going underground however implies giving up several fringe benefits. If the value of these benefits is non-negative, the return to education should be positive or at least high enough to offset the lack of benefits. In this paper we investigate whether there exists an education premium in the informal sector. Or in other words, what is the return of schooling, if any, on agent's informal earnings? Although, there is an extensive literature on returns to education, in particular for developed countries, the analyses are mainly restricted to agents working in the formal sector.¹ Despite the growing relevance of informal activities, there is scarce discussion in both theoretical and empirical literatures on returns to education for workers engaged in the informal sector.

In order to contribute to a better understanding of the presence of educated workers in the informal sector in developing countries, we begin by developing an analytical structure. In doing so, we seek to capture this salient and seemingly puzzling feature of developing economies. We motivate our analysis in the context of a two-period model drawn from Lemieux, Fortin and Frechette (1994) with important differences. First, we incorporate schooling decision and agents live two periods. They allocate time between leisure and education in the first period, and in the second period, agents decide how much time they work in each sector - formal and informal. Second, the production technology differs across sectors and physical capital is an additional factor of production in the formal sector. Informal production uses only human capital. Third, we focus on a corner solution when agents work only in the informal sector and investigate conditions under which we would observe more educated agents working only in this sector.

¹See, for instance, Card (1999) and Belzil (2007) for extensive discussion and surveys of the returns to education literature.

The theoretical model implies that there is an education premium in the informal sector, at least for high skilled workers. Moreover, for a particular class of preferences, it is possible to detail the relation between years of schooling and the education premium. The effect of the latter on the former is ambiguous because there is an interaction between the substitution and income effects and, the final result depends on a set of unknown parameters. We can show that, as expected, the lower the schooling level, the more willing agents will be to study more. This happens because a small increase in premium does not alter substantially the perceived income and the substitution effect dominates the income effect. Following Arias, Hallock and Sosa-Escudero (2001), we assume that human capital is composite by education and unobserved ability, allowing any kind of interaction between them. In this vein, ability can be viewed as any skill that improves the agent's level of human capital, generating higher schooling returns. As a consequence, two predictions arise. First, inasmuch as ability is heterogeneous among individual, so returns to education become heterogeneous too. Second, if beyond the effect through education, ability has a direct impact on earnings, the omission of ability implies that education is correlated with the error term, being an endogenous variable.

We empirically study the relationship between schooling and earnings and test the implications of our theoretical model for the Brazilian economy using the ECINF survey (Informal Urban Economy Survey) of informal workers for the year of 2000. In particular, we estimate and test the existence of an education premium in the informal sector, and investigate whether this premium varies along the conditional distribution of earnings. The examination of this unique dataset allows a better understanding of schooling and earnings of workers engaged in informal activities in Brazil. ECINF survey focus on self-employed and firms with up to five employees. These workers are defined, in our study, as informal workers. This dataset follows the International Labour Organization (ILO) definition of informal sector as enterprises with all or most characteristics in a list that includes family ownership, small scale of operations and labor-intensive methods. The definition of informality in this paper is then related with firm size, rather than tax avoidance or evasion. Our data on education is a categorical variable with values ranging from one to eight. It comprises education levels from no education to college degree and, thus, we estimate the return to one additional level of schooling.

To accommodate the possible heterogeneity in the distribution of earnings, conditional on education, we use quantile regression techniques. Indeed, this heterogeneity precludes the use of conventional least squares and two stage least squares results, while quantile regression approach constitutes a suitable way to investigate whether the returns to education differ along the wage distribution.² Quantile regression methods provide a framework for robust estimation and inference, being an essential tool to analyze the returns of education. However, the available data set does not allow us to control for ability. Thus, in order to deal with the endogeneity problem

²See Card (1995, 1999, 2001), Carneiro, Heckman and Vytlačil (2000) and Arias, Hallock and Escudero (2001) for a detailed discussion.

caused by the unobserved ability, and also allow for heterogeneity, our estimation strategy follows the instrumental variables quantile regression (IV-QR) method proposed by Chernozhukov and Hansen (2006, 2008). Finally, using the Kolmogorov-Smirnov test we address three questions: *(i)* is there education premium in the informal sector? *(ii)* does the education premium vary along the quantiles? and *(iii)* is there an endogenous relationship between education and earnings?

The IV-QR results show that returns to schooling are positive and monotonically increasing along the quantiles. For low quantiles of the conditional earnings distribution in the informal sector, there is approximately a fifty percent increase in the worker's wage per additional level of schooling. The returns increase as the quantile index increases toward the middle of the distribution and then remain approximately constant at eighty percent per additional level of schooling. This implies that larger gains to additional level of schooling exist for those at the middle of the earnings distribution. In addition, for high quantiles the returns of education are high, although the confidence interval becomes especially large for these quintiles. This heterogeneous education premium might be viewed as reflecting the distribution of unobservable ability. Last, the evidence suggests that education is an endogenous variable, requiring the use of instrumental variables to estimate the schooling premium. Thus, the findings from the empirical model corroborate the predictions of the theoretical model when unobserved ability is taken into account. In addition, regarding the results from our proposed empirical model, age has a positive impact on earnings, but this impact is larger for individuals at high quantiles of the conditional earnings distribution. Women tend to earn less than men in the informal sector in Brazil and experience has a negative impact on informal earnings, although the experience coefficient is very close to zero.

The rest of the paper is organized as follows. Section 2 presents the data and facts about the Brazilian informal sector. Section 3 presents the economy and the structure of the model. In Section 4, we describe the data used and present our estimation strategy and the empirical results obtained for informal workers in Brazil. Section 5 offers concluding remarks.

2 The facts and data

The presence of a substantial informal sector in developing countries has been well studied and documented.³ Before turning to the analytical model, it is useful to reflect briefly on the incentives people have to work in the informal sector. These relate broadly to issues such as government interventions - taxation and labor market regulations (Johnson et al., 1998, Friedman et al., 2000; Fugazza and Jacques, 2003); bureaucracy and corruption among other institutional and enforcement conditions (Busato and Chiarini, 2004; Choi and Thum, 2005; Dabla-Norris, Gradstein and Inchauste, 2008) - and access to capital markets and heterogeneity of firms and entrepreneurs (Antunes and Cavalcanti, 2007; Dessy and Pallage, 2003).

³Schneider and Enste (2000) provide estimates of the size of the informal economy in developed and developing nations.

The features of the informal labor market have also been of central interest in this literature. There are two main views. One of them argues that labor markets are segmented in developing countries. According to Dickens and Lang (1985), labor market segmentation implies a dualism in earnings for individuals with similar characteristics, as a consequence of the sector in which they work. Theoretical models explain labor market segmentation as a result of labor market policies, labor market institutions or the payment of efficiency wages in the formal sector (Rauch, 1991; Fortin, Marceau and Savard 1997; Straub, 2005). Barriers to entry, it is conjectured, prevent certain groups of workers from competing for higher paying formal jobs.

Recently, some studies have questioned the labor market segmentation hypothesis. Pratap and Quintin (2006) find no support for the hypothesis that observably similar workers earn more in the formal sector than in the informal sector in Argentina, a similar conclusion reached by Magnac (1991) and Maloney (1999) with data from Colombia and Mexico, respectively. Earle and Sakova (2000) find positive earning differentials for self-employed workers in the Transition countries. Bargain and Kwenda (2009) applied a fixed effect quantile approach to estimate the conditional distribution of the wage gap in South Africa, Brazil and Mexico, taking into accounting workers' unobserved heterogeneity. For all countries the authors concluded the informal sector wage penalty is large in the lower part of the distribution but almost disappears at the top.

Recent studies have investigated the issue of labor market segmentation in Brazil.⁴ They used data from the Brazilian household survey, *Pesquisa Nacional por Amostra de Domicílios (PNAD)* and the *Monthly Employment Survey (PME)*.⁵ Carneiro and Henley (2001) find that the differences in earnings are strongly explained by nonobservable characteristics of the workers who decide to join either the formal or the informal sector. Tannuri-Pianto and Pianto (2002) present selectivity corrected quantile regression models for earnings of both informal and formal workers and find that in the informal sector the unobservable characteristics increase the expected income for lower earnings quantiles while lowering those for higher quantiles. Filho et al. (2004), using a repeated cross-sections of Brazilian household surveys (PNAD), also concludes against the segmentation hypothesis. The empirical findings of Mattos and Ogura (2008) that formal firms are more productive than informal firms and high skilled workers are more likely to be employed in the formal sector. Botelho and Ponczek (2010) measure the degree of segmentation in the Brazilian labor market using data from the PME, a rotating panel of households in six metropolitan regions, and find that workers earn more in the formal sector, which supports the segmentation hypothesis. Ulyssea (2010) documented a sharp increase in the informal sector in Brazil in the late 80's. This coincides with a revision of the Brazilian Constitution, which greatly increased labor costs. Using the PME to analyze the mobility from one sector to another in Brazil, the author concluded that

⁴Ulyssea (2005) provides a comprehensive review of the literature addressing the informal market and the measurement of the degree of segmentation in Brazil.

⁵Paula and Scheinkman (2007) use data from the ECINF, but they address different aspects of informal activities in Brazil.

the transition rate from the formal to the informal sector is much lower than in the opposite direction, suggesting that work mobility is not as restricted as the traditional dualistic view would predict.

In these studies, a person is classified as an informal worker if she is an employee (domestic or otherwise) and works without a signed labor card, a definition of informal worker different than the one used in this paper. It is important to highlight that a key distinction between the *PNAD* and ECINF datasets refers to the dimension of the occupied personnel in the informal sector in Brazil. According to Jorge and Valadao (2003), the estimates gathered from the *PNAD* indicate an occupied contingent in the informal sector beyond the estimates found out by the ECINF. This result is particularly affected by the total of employees and unpaid family workers. It is not possible to identify, through the *PNAD*, the size of the activities in which they work. Therefore, the size of the informal sector is increased with the persons working in firms with more than five people occupied. Moreover, the ECINF excludes, from the scope of the informal sector, companies that are legally classified as stock companies or their revenue statement was filled in a given year. It indicates that although belonging to the definition of small business, those companies had a complete system of accounting. Through the *PNAD* it is not possible to proceed this filtering because such information is not gathered.

The number of own account workers, employers and employees is higher when the source is the survey of the labor force (*PNAD*). The estimates of the *PNAD*, considering only the urban areas in Brazil, indicate a total of 19 million people working in the informal sector, while the ECINF estimates a total of 13 million people engaged in informal activities. Hence, although the ECINF contains only information about informal workers, the advantage of this survey is that more precise information on informal labor characteristics can be obtained, information that is partially or not observed in other datasets. In addition, this dataset allows a more precise assessment of these activities in terms of job opportunities and earnings. The study of this unique dataset will allow a better understanding of the relationship between schooling and earnings for those workers engaged in informal activities in Brazil.

In this paper, we use the 2003 edition of the ECINF survey (*Informal Urban Economy Survey*), collected in October of 2003. Sampling strategy used the demographic census as a frame and the survey is limited to urban areas.⁶ Preliminary interviews were conducted to screen households for the presence of at least one entrepreneur with a business employing five or less people. Interviews were made aware that information collected for the survey was confidential and would only be utilized for statistical purposes.

The ECINF indicates that twelve percent of informal firms in Brazil have less than five employees and eighty eight percent are self-employed, that is a person who works by himself, without extra help and other workers. The revenue of firms with up to five employees is on average five

⁶For more details, see Almeida and Bianchini (1998), Jorge and Valadao (2003).

times higher than self-employed earnings. Only seven percent of the informal sector firms use the bank system for loans. Entrepreneurs own savings and informal credits play an important role in financing informal activities.

We obtain information on 60,233 informal workers in urban regions from all states in Brazil. Among these informal workers, sixty nine percent are self-employed and ten percent are employers. Employees with (10%) and without (6%) proper labor documents represent sixteen percent of this sample and five percent are non-paid workers. The informal job is the only and main job for ninety percent of the sample and ten percent has a second job. The majority of informal workers are men (64%). Of those informal workers, other than the owner or self-employed, one-third is between 18 – 24 years old and thirty six percent is in the 25 – 39 years old range. Informal jobs are found mainly through personal relationship and social ties. The majority of informal workers have been at the firm for a short period of time - thirty four percent of the workers for less than one year and thirty nine percent for 1 to 3 years.

Forty four percent of this sample works 40 – 60 hours per week. Earnings tend to be increasing in the education level (except for high-school drop-outs), as well as with age, for both men and women (except for those older than 60). One third of the informal workers drop out elementary school and high schoolers are around twenty percent of informal workers. Among informal firms' owners, twenty eight percent of employers has completed middle-school and forty percent of self-employed has not completed the elementary school. Table 1 presents for each education level the equivalent years of schooling and the percentage of individuals in each level.

Table 1 - Education profile of informal workers, Brazil, 2003
[ABOUT HERE]

Notice that one third of our sample has some level of schooling beyond "reads and writes". One quarter of the workers surveyed has completed the secondary school, which corresponds to twelve years of formal education. Thirteen percent of informal workers has completed at least eight years of formal education (primary education). Indeed, 59% of informal works have eight or more years of schooling. About 24% has completed at least eleven years of formal education (graduated secondary school) and, 36% of our sample has eleven or more years of schooling. To illustrate the heterogeneity in education in our sample, Figure 1 presents the unconditional density function estimate using a simple kernel estimator.⁷

Figure 1 - Education Distribution
[ABOUT HERE]

⁷The kernel is Gaussian and the bandwidth is the well known Silverman's rule of thumb.

3 A Model of Informal Activities and Schooling

We consider a two-sector economy. Time is discrete. Every period, a cohort of measure 1 of two-period-lived agents is born. As in most modern models of the informal sector, we assume that labor markets are competitive. In each period, agents are endowed with one unit of time.

In the first period of his life an agent allocates his time between leisure (l_t) and education (s_t). In the second period, the schooling decision is taken as given and the agent allocates his time among leisure (l_{t+1}), formal work (h_{0t+1}) and informal work (h_{1t+1}). The agent's time constraints in each period are as follows:

$$s_t + l_t = 1, \tag{1}$$

$$h_{0t+1} + h_{1t+1} + l_{t+1} = 1. \tag{2}$$

The agent has a nonlabor income I_t that includes any lump sum transfer or initial wealth in the first period. To finance future consumption, the agent can transfer wealth over time through bond holdings. One-period bonds (b_t) earn a risk-free interest rate ($1 + r_t$), measured in units of time $t + 1$ consumption good per time t consumption good, such as

$$c_t + b_t = I_t \tag{3}$$

where c_t represents a composite consumption good (the numeraire) in period t .

The formal sector (0) uses both human capital (H_{0t+1}), and physical capital (K_{0t+1}) by means of a concave Cobb-Douglas function as follows: $y_{0t+1} = A_0 H_{0t+1}^\theta K_{0t+1}^{1-\theta}$, where A_0 is the total factor productivity. Regarding the informal sector production technology (1), there exists robust empirical evidence that informal economic activities emphasize small-scale and labor intensive production (Fortin et al. 1997; Gauthier and Gersovitz 1997; de Soto 2000). Informal production tends to be much more labor-intensive than formal production. Thus, we assume that informal sector uses only human-capital (H_{1t+1}) as input and the production function is given by $y_{1t+1} = A_1 H_{1t+1}$, where A_1 is the total factor productivity.

In each sector, human capital accumulation is possible by combining inputs such as time spent in school (s_t) and time spent in the labor market: $H_{it+1} = e^{\gamma_i s_t} h_{it+1}$, where h_{it+1} is the time spent in sector $i = 0, 1$. That is, the agent can invest in education in the first period and earn more in the labor market in the second period. Our treatment of schooling and earnings reflects the Mincerian human capital literature in which education yields higher wage rates (Mincer, 1974). Assuming that labor markets are competitive, the earnings function become the time allocated to each sector times the respective marginal product:

$$\begin{aligned} w_{0t+1} &= \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta, \\ w_{1t+1} &= A_1 e^{\gamma_1 s_t} h_{1t+1}, \end{aligned}$$

where $\tilde{A}_0 \equiv \theta A_0 K_{0t+1}^{1-\theta}$, $\tilde{\gamma}_0 \equiv \theta \gamma_0$, and $\gamma_0, \gamma_1 > 0$ quantify the impact of schooling on formal and informal earnings, respectively. Therefore, in the second period the agent consume all resources, that is, the earnings of each sector plus the savings return, as follows:

$$c_{t+1} = \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta + A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t. \quad (4)$$

We consider a separable utility function:

$$U(c_t, c_{t+1}, l_t, l_{t+1}) = u(c_t) + v(l_t) + \beta [u(c_{t+1}) + v(l_{t+1})], \quad (5)$$

where the function $u(\cdot)$ and $v(\cdot)$ are strictly concave utility function ($u' > 0$, $u'' < 0$, $v' > 0$ and $v'' < 0$). The worker's problem is then to maximize (5) subject to (1), (2), (3), (4) and $h_{0t+1} \geq 0$, $h_{1t+1} \geq 0$, $s_t \geq 0$. This problem can be written in the Lagrangian form as follows:

$$\begin{aligned} \mathcal{L} = & u(I_t - b_t) + v(1 - s_t) + \beta u \left(\tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta + A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t \right) \\ & + \beta v(1 - h_{0t+1} - h_{1t+1}) + \lambda_1 h_{0t+1} + \lambda_2 h_{1t+1} + \lambda_3 s_t, \end{aligned}$$

where λ_1 , λ_2 , and λ_3 are the Lagrange multipliers. The solution of the problem is given by s_t^* , h_{0t+1}^* , h_{1t+1}^* , λ_1^* , λ_2^* , λ_3^* such as:

$$\frac{\partial \mathcal{L}}{\partial b_t} = -u'(c_t) + \beta u'(c_{t+1})(1 + r_t) = 0, \quad (6)$$

$$\frac{\partial \mathcal{L}}{\partial s_t} = -v'(1 - s_t) + \beta u'(c_{t+1}) \left(\tilde{\gamma}_0 \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta + \gamma_1 A_1 e^{\gamma_1 s_t} h_{1t+1} \right) + \lambda_3 = 0, \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial h_{0t+1}} = \beta u'(c_{t+1}) \theta \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^{\theta-1} - \beta v'(1 - h_{0t+1} - h_{1t+1}) + \lambda_1 = 0, \quad (8)$$

$$\frac{\partial \mathcal{L}}{\partial h_{1t+1}} = \beta u'(c_{t+1}) A_1 e^{\gamma_1 s_t} - \beta v'(1 - h_{0t+1} - h_{1t+1}) + \lambda_2 = 0, \quad (9)$$

$$\lambda_1 h_{0t+1} = 0, \lambda_2 h_{1t+1} = 0, \lambda_3 s_t = 0, \quad (10)$$

$$\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \quad (11)$$

$$h_{0t+1} \geq 0, h_{1t+1} \geq 0, s_t \geq 0. \quad (12)$$

Assuming an interior solution for s_t , conditions (7) and (10) imply that

$$v'(1 - s_t) = \beta u'(c_{t+1}) \left(\tilde{\gamma}_0 \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta + \gamma_1 A_1 e^{\gamma_1 s_t} h_{1t+1} \right), \quad (13)$$

i.e., the education decision depends on its (leisure) cost in the first period and its benefit in the second period (higher earnings discounted by β).

Regarding the optimal choices of h_{0t+1} and h_{1t+1} , we are interested in the corner solution for the informal sector. In this case, $h_{0t+1} = 0$ and $h_{1t+1} > 0$. When $h_{0t+1} = 0$, λ_1 can be positive or zero and condition (8) implies that

$$\beta v'(1 - h_{0t+1} - h_{1t+1}) \geq \beta u'(c_{t+1}) \theta \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^{\theta-1}, \quad (14)$$

i.e., the marginal cost of working in the formal sector is greater than or equal to its marginal benefit. On the other hand, when $h_{1t+1} > 0$, condition (10) implies $\lambda_2 = 0$ and, condition (9) becomes

$$\beta v'(1 - h_{0t+1} - h_{1t+1}) = \beta u'(c_{t+1}) A_1 e^{\gamma_1 s_t}, \quad (15)$$

i.e., the marginal cost of working in the informal sector is equal to its marginal benefit. Equations (14) and (15) together imply that the marginal benefit of working in informal sector is greater than or equal to the this marginal benefit in the formal sector.

Here, we focus on the earnings profile and schooling decision of workers engaged in informal activities only. In this case, if we assume that $h_{0t+1} = 0$, $h_{1t+1} > 0$ and $s_t > 0$, conditions (6), (7) and (9) result in the following system of equations:

$$u'(c_t) = \beta u'(c_{t+1}) (1 + r_t), \quad (16)$$

$$v'(1 - s_t) = \beta u'(A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t) \gamma_1 A_1 e^{\gamma_1 s_t} h_{1t+1}, \quad (17)$$

$$\beta v'(1 - h_{1t+1}) = \beta u'(A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t) A_1 e^{\gamma_1 s_t}, \quad (18)$$

where $c_t = I_t - b_t$ and $c_{t+1} = A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t$. This system has three unknowns, namely b_t , s_t and h_{1t+1} , but given its non-linearity it is virtually impossible to find, for instance, the optimal s_t^* as a function of quantities I_t , A_1 , γ_1 , r_t . The alternative would be to characterize the solution using the implicit function theorem for a system of equations. The difficult then becomes to find the signal of the derivative $ds_t/d\gamma_1$.⁸

Thus, we pursue the following strategy. Instead of calculating $ds_t/d\gamma_1$, we show that for large s_t , the model implies a positive impact of schooling on the agent's earnings ($\gamma_1 > 0$). Indeed, as schooling investment in the first period has a leisure opportunity cost, one would expect that even agents that work only in the informal sector demand a premium for education. For a particular functional form - the logarithmic utility - we can clearly identify a condition such that $ds_t/d\gamma_1$ is positive. Key for this result is the interaction between income and substitution effects. When the

⁸It would be necessary to calculate

$$\begin{bmatrix} \frac{\partial h_{1t+1}}{\partial \gamma_1} \\ \frac{\partial s_t}{\partial \gamma_1} \\ \frac{\partial b_t}{\partial \gamma_1} \end{bmatrix} = - \begin{bmatrix} \frac{\partial F_1}{\partial h_{1t+1}} & \frac{\partial F_1}{\partial s_t} & \frac{\partial F_1}{\partial b_t} \\ \frac{\partial F_2}{\partial h_{1t+1}} & \frac{\partial F_2}{\partial s_t} & \frac{\partial F_2}{\partial b_t} \\ \frac{\partial F_3}{\partial h_{1t+1}} & \frac{\partial F_3}{\partial s_t} & \frac{\partial F_3}{\partial b_t} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial F_1}{\partial \gamma_1} \\ \frac{\partial F_2}{\partial \gamma_1} \\ \frac{\partial F_3}{\partial \gamma_1} \end{bmatrix}$$

where F_1 , F_2 and F_3 are defined by equations (16), (17) and (18).

former is greater than the later, agents study less. In this case, a change in γ_1 alter substantially the perceived income by agents, leading to a decision to study less. Propositions 1 and 2 formalize these results, respectively.

Proposition 1 *For an agent that works only in the informal sector, there is a schooling level \bar{s} such that for $s_t \geq \bar{s} > 0$ impact of schooling on the agent's informal earnings (γ_1) must be greater than zero.*

Proof. See Appendix A.1. ■

The main implication of Proposition 1 is that at least for certain values of s_t we should observe an education premium for those engaged in the informal sector. In other words, the model implies that at least high skilled workers in the informal sector should earn some compensation for their investment in education.

Now we specialize the model to the case of logarithmic utility. With such preferences, it is possible to characterize the relation between years of schooling and the education premium, as presented in Proposition 2.

Proposition 2 *If preferences are logarithmic, agents work only in the informal sector and there is a interior solution for s_t , then $ds_t/d\gamma_1 > 0$ when $g(s_t) < \Gamma$, where*

$$g(s_t) = A_1 e^{\gamma_1 s_t} [1 - (1 + \beta) \gamma_1 (1 - s_t)]^2$$

$$\Gamma = I_t (1 + r_t) (1 + 2\beta)$$

Moreover, assuming that $\gamma_1 < \frac{1}{(1+\beta)t}$, we guarantee that $g'(s_t) > 0$.

Proof. See Appendix A.2. ■

When the opportunity cost of leisure, γ_1 , increases, the substitution effect leads to an increase in s_t . However, for large s_t a variation in γ_1 might represent a large variation in the perceived income and, the income effect can lead to a reduction in s_t . When $g(s_t)$ is lower than Γ , the first effect dominates. For $g'(s_t) > 0$, the larger s_t the greater $g(s_t)$. In this case, $g(s_t)$ can become greater than Γ , inverting the signal of $ds_t/d\gamma_1$. Indeed, if $s_t = 0$, there is no income effect because changes in γ_1 do not change the perceived income. Last, the term Γ increases with β , I_t and r_t . Large β means that agents are more willing to invest in education today to increase the earnings tomorrow.⁹ The term $I_t(1 + r_t)$ is the gross return on initial wealth. Given his/her initial wealth an agent can finance future consumption by allocating part of his/her wealth to bonds, earn interest and consume more in the future. If I_t is low, the agent is more willing to give up leisure and increase schooling to increase his/her income in future and consume more.

⁹Notice that, when $\gamma_1 < \frac{1}{(1+\beta)t}$, the larger the β , the lower $g(s_t)$. Thus, β increases Γ and, decreases $g(s_t)$.

NOVO

Obs 1. Fala do erwan: “Also, what can you say about the premium across the two sector, when both sectors are active? Is there anything special about the informal sector or would the same pattern arise in the formal sector? If the pattern is common to both sectors, why care at all about the formal/informal distinction?”

Obs 2: Isso abaixo foi o que eu consegui pensar. Não dá para dizer muito sobre o perfil do retorno da educação no setor formal e informal, pois γ_0 e γ_1 são exógenas. Só dá para dizer como as endógenas respondem a mudanças nestes parametros. O perfil/padrao de γ_0 e γ_1 são questões empíricas. No entanto, só podemos estimar γ_0 . Eu entendo o ponto dele como "estudar sempre tem custo, se não existe benefício pra que estudar? se isso é sempre verdade, para que fazer um modelo com setor formal e informal?". Tentei mostrar abaixo que o que acontece em um setor afeta o outro. Acho que não é uma resposta perfeita, mas pelo menos ele verá que tentamos (ou seja, não ignoramos o pedido dele). No mais, eu chutaria que ele leu rapidamente nosso paper e talvez por isso ele tenha esses pedidos para falarmos dos dois setores. Digo isso pois alguns dos papers adicionais que ele menciona são sobre dualismo/segmentacao. Bom, no lugar dele eu teria lido rapido nosso paper. Ele deve ser muito ocupado. Assim, estou especulando, mas talvez lendo com calma como editor ele perceba que o foco é mesmo setor informal.

As mentioned, the informal job is the only and main job for ninety percent of our sample and this is the reason why we analyzed the corner solution of the model. Of course, our propositions depend on this choice. For instance, if we check equation (7) while the education cost is given by the lost of leisure, its benefit comes from higher earnings in both sectors in the second period. Therefore, differently from proposition 1, when both sectors are active the education premium in the informal sector can be null and, even though, the optimal education level can be greater than zero. When we assume interior solution for h_{0t+1} and h_{1t+1} , besides $\gamma_1 = 0$, equation (7) becomes:

$$\frac{\partial \mathcal{L}}{\partial s_t} = -v'(1 - s_t) + \beta u'(c_{t+1}) \tilde{\gamma}_0 \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta + \lambda_3 = 0$$

And, $\lambda_3 \geq 0$ means that

$$v'(1 - s_t) \geq \beta u'(c_{t+1}) \tilde{\gamma}_0 \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^\theta$$

Equality means that $s_t^* > 0$, while inequality means that $s_t^* = 0$. If formal sector premium tends to zero, the RHS tends to zero and the optimal decision is null schooling. When this is not the case, for individuals in both sectors we can observe s_t greater than zero even when $\gamma_1 = 0$, because the formal sector guarantees a nonzero education premium. Obviously, if the difference between the return in the formal sector and informal sector becomes large, the individual will spend more time in the formal one. Indeed, for individuals in both sectors equations (8) and (9) imply that

marginal utility (return) of work in each sector identical:

$$\beta u'(c_{t+1}) \theta \tilde{A}_0 e^{\tilde{\gamma}_0 s_t} h_{0t+1}^{\theta-1} = \beta u'(c_{t+1}) A_1 e^{\gamma_1 s_t}$$

And, after some manipulation, it is easy to see that h_{0t+1} is a function of education premium in each sector:

$$h_{0t+1} = \left(\frac{\theta \tilde{A}_0}{A_1} \right)^{\frac{1}{1-\theta}} e^{\frac{1}{1-\theta}(\tilde{\gamma}_0 - \gamma_1) s_t}$$

4 Empirical Application

4.1 Description of Variables

The statistical description of the variables considered in this study is presented in Table 2. The variables *gender*, *age*, *days_month* and *hours_week* are self-explanatory. We use the variable *months_job*, that is the number of months in the job as a proxy for experience. *Education* is a categorical variable with values ranging from one to eight and comprise education levels from *no education* to *graduated college (college degree)*. The variable *wage_perhour* represents an informal worker earnings in October of 2003 (in R\$) per hour. *Work_link* refers to the type of work that is executed by the informal worker, i.e. if the worker works for limited or not limited time or works per task. Table A.3.1, Appendix A.3, presents detailed information for *payment*, *work_link*.

Table 2 - Summary Statistics
[ABOUT HERE]

4.2 Estimation Strategy and Empirical Results

One of the most widely studied topics in labor economics is the impact of education on earnings. The large volume of research in this area has been motivated by both interest in the causal effect of education on earnings as well as the inherent difficulty in measuring this effect. The difficulty arises due to the possible endogenous relationship between education and earnings. In particular, it seems likely that unobserved individual ability is correlated to both a person's education and wages, thus biasing standard regression estimates of the relation between schooling and earnings. In addition, economists have long believed that the returns to schooling may vary among individuals, further complicating the interpretation of conventional least squares and two stage least squares results.

Arias, Hallock, and Sosa-Escudero (2001) approach this issue by assuming a linear location-scale model where individual ability interacts with education. In our model this is equivalent to assume that $\gamma_1^i S_t^i = \tau_0 S_t^i + \tau_1 a^i S_t^i$, where i indicates the individual and a^i is his/her ability.

This approach means that human capital is composite by education and unobserved ability and schooling returns depend also on the ability level of the individual. In this vein, ability can be viewed as any factor that improves the agent's level of human capital, generating higher schooling returns. Because γ_1^i is the sum of τ_0 and $\tau_1 a^i$, and individuals abilities are different, they can have dissimilar return even when they have the same schooling level. Thus, our model suggests that the use of OLS or IV-OLS is not appropriated. Clearly, the location-scale model is very simple: ability has a monotonically positive or negative impact on education return. However, as emphasized by Arias, Hallock, and Sosa-Escudero (2001), quantile regression provides a more flexible approach to distinguishing the effect of education on different percentiles of the conditional earning distribution, being consistent with a non-trivial and, in fact unknown, interaction between education and ability.

It is important to discuss if our model is compatible with the common view about education endogeneity. Schooling decision, s_t is a function of γ_1 (see equation (26) in the Appendix). Of course, because a^i is not a decision variable, γ_1 can be replaced by $\gamma_1^i = \tau_0 + \tau_1 a^i$ (or any other function of a^i). If, beyond the effect through education, ability has a direct impact on earnings, the omission of ability implies that education is correlated with the error term. In this case, there is an endogenous relationship between education and earnings, being necessary the use of a instrumental variable approach in order to estimate the education return.

Therefore, in order to address the issue of heterogeneity in the returns to schooling and possible endogeneity predicted by our model, we estimate the instrumental variable quantile regression model of schooling on earnings for the informal sector in Brazil. The estimation strategy follows Koenker and Bassett (1978) for the usual quantile regression (QR) method and Chernozhukov and Hansen (2006, 2008) for the instrumental variables quantile regression (IV-QR) estimator. For completeness, we also provide results for the usual quantile regression estimates.

To investigate the returns to schooling of workers engaged in informal activities in Brazil, we estimate a linear conditional quantile model of the form

$$Q_{\ln(Y)}(\tau|X) = \gamma(\tau)S + X'\beta(\tau), \quad (19)$$

where Y is earnings per hour, S is education and X is a vector of covariates consisting of experience, experience squared, gender and age. In our theoretical model, the logarithm of earnings per hour means $\ln(w_1/h_1) = \ln A_1 + \gamma_1 s_t$, where s_t measure years of schooling. Note that, education is only a non-decreasing function of years of schooling and its coefficient, γ , is no longer γ_1 , although γ constitutes a measure of the return to education. Last, it is worth mentioning that in line with Light (2001), our estimates of the returns to schooling represent the labor market returns to schooling plus any work experience gained while in school. To test the implications of our model, we apply the Kolmogorov-Smirnov to evaluate if: *i*) there is a premium for education ($H_0 : \gamma(\tau) = 0$ versus $H_A : \gamma(\tau) \neq 0$); *ii*) the premium varies along the quantiles ($H_0 : \gamma(\tau) = \gamma$ versus

$H_A : \gamma(\tau) \neq \gamma$); *iii*) education is endogenous ($H_0 : \gamma(\tau) = \gamma_{QR}(\tau)$ versus $H_A : \gamma(\tau) \neq \gamma_{QR}(\tau)$), where γ_{QR} means the coefficient of the quantile regression without use of instrumental variables).

The database used in our study refers only to informal workers. There is abundant literature on the issue of sample selectivity bias in wage equations estimation using only data for working people (Heckman, 1979; Manski, 1989). Given the fact that the traditional selectivity correction procedure - Heckman's (1979) procedure - is not available when using the quantile regression model (Buchinsky, 2001) and the lack of good instruments to represent a labor market decision mechanism in our sample, we present results without sample selectivity correction. This is not an unusual decision when QR are estimated.¹⁰ We use the variable *work_link* as an instrument for education. We argue that this variable is correlated with the level of education of the worker and not correlated with the worker's ability. On the one hand, this variable refers to the type of work that is executed by the informal worker, thus this variable is clearly correlated with education. On the other hand, firms already have a pre-determined policy on hiring workers and contractors, and therefore ability is not observed when tasks are assigned. Thus, we expect a very small and insignificant correlation between *work_link* and the nonobserved individual ability. Moreover, a large portion of this variable in the sample refers to the owner or business partner and work for unlimited time, suggesting that the worker's ability plays a small role on the tasks undertaken by the informal worker.

We present results for the IV-QR estimator as well as for the usual quantile regression (QR) for comparison reasons. QR and IV-QR estimates of the schooling coefficient are provided in Table 3, for selected quantiles. Figures 2 and 3 present the results for the coefficients and confidence bands, for a range of quantiles, for IV-QR and QR, respectively. The shaded region in each panel represents the 95% confidence interval. In addition, the estimates for simple least squares (LS) and instrumental variables least squares (IV-OLS) appear in the respective figures, with dashed lines for confidence bounds. The least squares estimates show that after controlling for endogeneity the schooling return increases from 0.2816 to 0.8331.

Table 3 - Estimates Schooling coefficient, selected quantiles
[ABOUT HERE]

Figure 2 - Schooling Coefficient (QR)
[ABOUT HERE]

Figure 3 - Schooling Coefficient (IV-QR)
[ABOUT HERE]

¹⁰For discussion and references, see Montenegro (2001) and Pratap and Quintin (2006a).

Note that both the quantile regression and instrumental variables quantile regression estimates suggest that the returns to schooling vary over the earnings distribution. The variability of schooling returns is most apparent and dramatic in the IV-QR estimates. While the QR estimates do vary statistically, they are all closely clustered around the OLS estimate. The OLS value is 0.2816 while the QR estimate varies from 0.2440 to 0.3130, with average value equal to 0.2746. The practical lack of variability in the QR estimates is clearly demonstrated in Figure 2. The IV-OLS value is 0.8331 while the IV-QR ranges from 0.3200 to 1.9400, with mean value 0.8986 (Figure 3). Therefore, relative to the IV-QR estimates, the QR estimates appear to be approximately constant.¹¹

The shapes of the estimated coefficients for schooling are very interesting. The QR estimates exhibit a distinct U-shape, implying higher returns to schooling for those in the tails of the distribution than for those in the middle. However, if schooling is endogenous to the earnings equation, these estimates do not consistently estimate the parameters. The IV-QR estimates, on the other hand, are consistent under endogeneity and show quite different results than those obtained through standard QR. In particular, the IV-QR results show positive and monotonic increasing returns to schooling of approximately 50 percent per additional level of schooling at low quantiles of the earnings distribution. The returns increase as the quantile index increases toward the middle of the distribution and then remain approximately constant at levels near the IV-OLS estimates around 90 percent per additional level of schooling. This implies that the large gains to additional years of schooling accrue to those at the middle of the earnings distribution. In addition, for high quantiles the returns of education are extremely high, achieving 180 percent per additional level of schooling. This result can be, in part, associated to the fact that in Brazil a large portion of informal workers are self-employed. Besides, more educated and high skilled worker tend to be the owner or the person in charge of the informal business and therefore large returns are available from their high levels of education. In any case, it is important to mention that the confidence interval becomes substantially large for the highest quintiles.

Following the interpretation of Arias, Hallock, and Sosa-Escudero (2001) quantile regression methods allow us to estimate returns to schooling for individuals at different quantiles of the conditional distribution of earnings, which might be viewed as reflecting the distribution of unobservable ability. Thus, our empirical findings corroborate the predictions of the theoretical model, where the premium for more able workers is higher. However, it is necessary to test formally the conjecture that premium is not constant along the quantiles. This is especially important because the confidence interval is wider in the highest quantiles.

Table 4 reports the results for diagnosis tests using Kolmogorov-Smirnov type of tests for

¹¹The return of one year of schooling (γ_1) can not be directly associated with the estimated return of a degree. For instance, if the degree were formed by three years and using the OLS result (28, 16%), γ_1 could be approximated by $100(1.2816^{1/3} - 1) = 8.62\%$. However, the degrees do not have the same numbers of years. Thus, although we find clear evidence that there is an education premium, it is virtually impossible to recover γ_1 .

quantile regression. First, we test the hypothesis of a zero constant coefficient for the IV-QR estimates across quantiles, which is rejected at the 1% level of significance. Thus, we reject the hypothesis of no premium for education. Secondly, we test the hypothesis of constant effects, that is, $\gamma(\tau) = \bar{\gamma}$, where we set $\bar{\gamma}$ as the median estimate. We are also able to reject this hypothesis at 1% level of significance. Therefore, our results are consistent with the version of the model that includes unobserved ability. Thus, although the confidence interval of the IV-QR contains the point estimate of IV-OLS, for various intermediate quantiles, the evidence suggests that the effect of education on earnings is heterogeneous. Finally, we test exogeneity, i.e., whether $\gamma(\tau) = \gamma_{QR}(\tau)$ where $\gamma_{QR}(\tau)$ are the estimates from QR model without instruments. Again, we strongly reject the null hypothesis which is in line with a direct effect of ability on earnings, besides the effect through human capital.

Table 4 - Kolmogorov-Smirnov Tests
[ABOUT HERE]

Studying the informal sector wage gap in South Africa, Brazil and Mexico, Bargain and Kwenda (2009) find that the informal sector wage penalty is large in the lower part of the distribution but almost disappears at the top. We find strong evidence of a large education premium at the top of the conditional distribution of earnings. This is line with our result; however, the definition of informality used by Bargain and Kwenda (2009) corresponds to the lack/avoidance of formal registration, taxation, regulation of maximum working hours or workers protection standards, being different from ILO definition. Thus, these authors did not exclude large firms, differently from our study.

Although, the quantile regression results are based on the conditional wage distribution, Bargain and Kwenda (2009) analyzed the unconditional wage distribution of informal sector in attempt to explain their results. Focusing in Brazil, they found evidence that the top informal workers are statistically older, more often located in São Paulo (a large urban area), generally have high education level and hold managerial positions in large firms, concluding that the top informal employees are very comparable to the formal employees. In our case, the firm size factor cannot account for the shortening of the gap, which further strengthens the relevant of the unobserved skill factor. Even though, the comparison between unconditional and conditional wage distribution should be viewed with caution, if the top informal workers are similar to the top formal ones (increasing their mobility) plus the relevant unobserved ability, the informal sector should pay a high return for keeping these employees.

Regarding the covariates gender, age and experience in the IV-QR estimation, we observe that the 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. Women tend to earn less

than men in the informal sector in Brazil and a woman's expected earnings at the 90th and 93th percentiles are approximately 40 percent lower than a man's. Finally, experience has a negative impact on informal earnings, although the experience coefficient is very close to zero (Figure 4).

Figure 4 - Age, Gender and Experience (IV-QR)
[ABOUT HERE]

5 Conclusion

This paper investigates the effect of schooling on agent's decision to work in the informal sector. We present a two-period model of schooling and informal activities and test its implications using a survey of Brazilian informal workers. The schooling decision in the first period affects the agent's earnings in both sectors in the second period. For logarithmic preferences, the model implies an ambiguous relation between education premium and schooling decision, in the sense that increases in the former can lead to reductions or increases in the former. This derivative is a priori unknown; however, for a general class of utility functions we showed that at least for high educated persons, the informal sector becomes an interest option only if a premium exists. The reason is simple: schooling implies lost of leisure and, as consequence, should be reward. For high values of s_t , we proof that even when there is a leisure bliss point. We investigate the predictions of the model using a quantile regression framework. The results show that returns to education in the informal sector are positive and monotonically increasing along the quantiles. Large gains to additional years of schooling are observed for high quantiles of the earnings distribution, which seems to be consistent with the argument that unobservable ability impact the education return.

Appendix

A.1 - Proof of Proposition 1

Proof. The corner solution for informal sector means that $h_{0t+1} = 0$. Besides, $s_t \geq \bar{s} > 0$ and $\lambda_3 s_t = 0$ means that $\lambda_3 = 0$. Then, the derivative with respect to s_t becomes $v'(1 - s_t) = \beta u'(c_{t+1}) \gamma_1 \tilde{A}_1 e^{\gamma_1 s_t} h_{1t+1}^\alpha$, where $c_{t+1} = \tilde{A}_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t$. By contradiction, suppose that $\gamma_1 = 0$, which implies that $v'(1 - s_t) = 0$. If $v'(\cdot)$ is always positive, the contradiction is clear. Suppose that exist l^b such as $v'(l^b) = 0$, with $l^b < 1$. Define $\epsilon = 1 - l^b$. Since $l_t = 1 - s_t$, define $\bar{l} = 1 - \bar{s}$ and assume that $\bar{s} > \epsilon$. Then, $\bar{l} < l^b$, and $v'(\bar{l}) > 0$. Therefore, if $s_t \geq \bar{s}$, $v'(1 - s_t)$ is no longer zero. Of course, γ_0 cannot be negative, because $v'(\cdot)$ is never negative. ■

A.2 - Proof of Proposition 2

Proof. Assume that $u(c) = \ln c$ and $v(l) = \ln l$. Focusing on individuals that choose to work only in the informal sector ($h_{0t+1} = 0$, $h_{1t+1} > 0$, $\lambda_2 = 0$) and their education level is greater than zero ($s_t > 0$, $\lambda_3 = 0$), the conditions (6), (7), (8) and (9) become

$$\frac{1}{I_t - b_t} = \frac{\beta(1 + r_t)}{A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t} \quad (20)$$

$$\frac{1}{1 - s_t} = \frac{\beta \gamma_1 A_1 e^{\gamma_1 s_t} h_{1t+1}}{A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t} \quad (21)$$

$$\frac{\beta}{1 - h_{1t+1}} = \lambda_1 \quad (22)$$

$$\frac{1}{1 - h_{1t+1}} = \frac{A_1 e^{\gamma_1 s_t}}{A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t} \quad (23)$$

Equation (22) implies that $\lambda_1 > 0$. Using equation (20)

$$\begin{aligned} A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t &= \beta(1 + r_t) (I_t - b_t) \\ A_1 e^{\gamma_1 s_t} h_{1t+1} &= \beta(1 + r_t) I_t - \beta(1 + r_t) b_t - (1 + r_t) b_t \\ A_1 e^{\gamma_1 s_t} h_{1t+1} &= \beta(1 + r_t) I_t - (1 + \beta)(1 + r_t) b_t \\ h_{1t+1} &= A_1^{-1} e^{-\gamma_1 s_t} [\beta(1 + r_t) I_t - (1 + \beta)(1 + r_t) b_t] \end{aligned} \quad (24)$$

Substituting equation (24) in equation (21):

$$\begin{aligned} A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t &= \beta \gamma_1 A_1 e^{\gamma_1 s_t} h_{1t+1} (1 - s_t) \\ A_1 e^{\gamma_1 s_t} h_{1t+1} [1 - \beta \gamma_1 (1 - s_t)] &= -(1 + r_t) b_t \\ A_1 e^{\gamma_1 s_t} \{A_1^{-1} e^{-\gamma_1 s_t} [\beta(1 + r_t) I_t - (1 + \beta)(1 + r_t) b_t]\} [1 - \beta \gamma_1 (1 - s_t)] &= -(1 + r_t) b_t \\ [\beta(1 + r_t) I_t - (1 + \beta)(1 + r_t) b_t] [1 - \beta \gamma_1 (1 - s_t)] &= -(1 + r_t) b_t \\ [\beta I_t - (1 + \beta) b_t] [1 - \beta \gamma_1 (1 - s_t)] &= -b_t \\ \beta I_t [1 - \beta \gamma_1 (1 - s_t)] - (1 + \beta) b_t [1 - \beta \gamma_1 (1 - s_t)] &= -b_t \\ b_t \{1 - (1 + \beta) [1 - \beta \gamma_1 (1 - s_t)]\} &= -\beta I_t [1 - \beta \gamma_1 (1 - s_t)] \end{aligned}$$

$$b_t = \frac{-\beta I_t [1 - \beta \gamma_1 (1 - s_t)]}{1 - (1 + \beta) [1 - \beta \gamma_1 (1 - s_t)]} \quad (25)$$

Substituting equations (24) and (25) in equation (23):

$$\begin{aligned} A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t &= (1 - h_{1t+1}) A_1 e^{\gamma_1 s_t} \\ A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t &= A_1 e^{\gamma_1 s_t} - h_{1t+1} A_1 e^{\gamma_1 s_t} \\ 2A_1 e^{\gamma_1 s_t} h_{1t+1} + (1 + r_t) b_t &= A_1 e^{\gamma_1 s_t} \\ 2A_1 e^{\gamma_1 s_t} \{A_1^{-1} e^{-\gamma_1 s_t} [\beta (1 + r_t) I_t - (1 + \beta) (1 + r_t) b_t]\} + (1 + r_t) b_t &= A_1 e^{\gamma_1 s_t} \\ 2[\beta (1 + r_t) I_t - (1 + \beta) (1 + r_t) b_t] + (1 + r_t) b_t &= A_1 e^{\gamma_1 s_t} \\ 2\beta (1 + r_t) I_t - 2(1 + \beta) (1 + r_t) b_t + (1 + r_t) b_t &= A_1 e^{\gamma_1 s_t} \\ 2\beta (1 + r_t) I_t + (1 + r_t) b_t [1 - 2(1 + \beta)] &= A_1 e^{\gamma_1 s_t} \\ 2\beta (1 + r_t) I_t + (1 + r_t) \left\{ \frac{-\beta I_t [1 - \beta \gamma_1 (1 - s_t)]}{1 - (1 + \beta) [1 - \beta \gamma_1 (1 - s_t)]} \right\} [1 - 2(1 + \beta)] &= A_1 e^{\gamma_1 s_t} \\ 2\beta (1 + r_t) I_t + (1 + r_t) \left\{ \frac{-\beta I_t}{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)} \right\} [1 - 2(1 + \beta)] &= A_1 e^{\gamma_1 s_t} \end{aligned} \quad (26)$$

Notice that s_t is the unique unknown in equation (26). Define F such as

$$F = A_1 e^{\gamma_1 s_t} - 2\beta (1 + r_t) I_t + \frac{\beta I_t (1 + r_t) [1 - 2(1 + \beta)]}{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)}$$

Then, we can use the implicit function theorem to calculate how s_t respond to changes in γ_1 .

$$\frac{ds_t}{d\gamma_1} = -\frac{\frac{\partial F}{\partial \gamma_1}}{\frac{\partial F}{\partial s_t}}$$

$$\begin{aligned} \frac{\partial F}{\partial s_t} &= \gamma_1 A_1 e^{\gamma_1 s_t} - \frac{\beta I_t (1 + r_t) [1 - 2(1 + \beta)]}{\{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)\}^2} \left\{ -\frac{1}{[1 - \beta \gamma_1 (1 - s_t)]^2} \right\} \beta \gamma_1 \\ \frac{\partial F}{\partial s_t} &= \gamma_1 A_1 e^{\gamma_1 s_t} + \frac{\gamma_1 \beta^2 I_t (1 + r_t) (-1 - 2\beta)}{\{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)\}^2 [1 - \beta \gamma_1 (1 - s_t)]^2} \\ \frac{\partial F}{\partial \gamma_1} &= s_t A_1 e^{\gamma_1 s_t} - \frac{\beta I_t (1 + r_t) [1 - 2(1 + \beta)]}{\{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)\}^2} \left\{ \frac{1}{[1 - \beta \gamma_1 (1 - s_t)]^2} \right\} [\beta \gamma_1 (1 - s_t)] \\ \frac{\partial F}{\partial \gamma_1} &= s_t A_1 e^{\gamma_1 s_t} - \frac{\gamma_1 \beta^2 I_t (1 + r_t) (-1 - 2\beta) (1 - s_t)}{\{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)\}^2 [1 - \beta \gamma_1 (1 - s_t)]^2} \end{aligned}$$

Because $[-1 - 2\beta] < 0$, it is easy to see that $\frac{\partial F}{\partial \gamma_1} > 0$. Thus, $\frac{ds_t}{d\gamma_1} > 0$ only when $\frac{\partial F}{\partial s_t} < 0$. This occurs when

$$\gamma_1 A_1 e^{\gamma_1 s_t} < \frac{\gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta)}{\{[1 - \beta \gamma_1 (1 - s_t)]^{-1} - (1 + \beta)\}^2 [1 - \beta \gamma_1 (1 - s_t)]^2}$$

Define $\Lambda(s_t) \equiv [1 - \beta\gamma_1(1 - s_t)]$, then

$$\begin{aligned}
\gamma_1 A_1 e^{\gamma_1 s_t} &< \frac{\gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta)}{[\Lambda(s_t)^{-1} - (1 + \beta)]^2} \frac{1}{\Lambda(s_t)^2} \\
\gamma_1 A_1 e^{\gamma_1 s_t} &< \frac{\gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta)}{[\Lambda(s_t)^{-2} - 2(1 + \beta)\Lambda(s_t)^{-1} + (1 + \beta)^2] \Lambda(s_t)^2} \\
\gamma_1 A_1 e^{\gamma_1 s_t} &< \frac{\gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta)}{1 - 2(1 + \beta)\Lambda(s_t) + (1 + \beta)^2 \Lambda(s_t)^2} \\
\gamma_1 A_1 e^{\gamma_1 s_t} &< \frac{\gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta)}{[1 - (1 + \beta)\Lambda(s_t)]^2} \\
\gamma_1 A_1 e^{\gamma_1 s_t} [1 - (1 + \beta)\Lambda(s_t)]^2 &< \gamma_1 \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} [1 - (1 + \beta)\Lambda(s_t)]^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} \{1 - (1 + \beta)[1 - \beta\gamma_1(1 - s_t)]\}^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} \{1 - 1 - \beta + (1 + \beta)\beta\gamma_1(1 - s_t)\}^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} \{-\beta + (1 + \beta)\beta\gamma_1(1 - s_t)\}^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} \{-\beta[1 - (1 + \beta)\gamma_1(1 - s_t)]\}^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} (-\beta)^2 [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} \beta^2 [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 &< \beta^2 I_t (1 + r_t) (1 + 2\beta) \\
A_1 e^{\gamma_1 s_t} [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 &< I_t (1 + r_t) (1 + 2\beta)
\end{aligned}$$

Define

$$\begin{aligned}
g(s_t) &\equiv A_1 e^{\gamma_1 s_t} [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 \\
\Gamma &\equiv I_t (1 + r_t) (1 + 2\beta)
\end{aligned}$$

Therefore, when

$$g(s_t) < \Gamma \Leftrightarrow \frac{ds_t}{d\gamma_1} > 0$$

Lets analyze $g(s_t)$:

$$\begin{aligned}
g'(s_t) &= \gamma_1 A_1 e^{\gamma_1 s_t} \beta^2 [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 + 2A_1 e^{\gamma_1 s_t} \beta^2 [1 - (1 + \beta)\gamma_1(1 - s_t)] (1 + \beta)\gamma_1 \\
&= \gamma_1 A_1 e^{\gamma_1 s_t} \beta^2 [1 - (1 + \beta)\gamma_1(1 - s_t)]^2 \{1 + 2[1 - (1 + \beta)\gamma_1(1 - s_t)]^{-1} (1 + \beta)\}
\end{aligned}$$

$$g'(s_t) > 0 \Leftrightarrow 1 + 2[1 - (1 + \beta)\gamma_1(1 - s_t)]^{-1} (1 + \beta) > 0$$

$$\frac{2(1 + \beta)}{[1 - (1 + \beta)\gamma_1(1 - s_t)]} > -1$$

Supposing that $1 - (1 + \beta)\gamma_1(1 - s_t) > 0$, i.e. $\gamma_1 < \frac{1}{(1+\beta)(1-s_t)}$. Then,

$$\begin{aligned}2(1 + \beta) &> -1 + (1 + \beta)\gamma_1(1 - s_t) \\2 + 2\beta &> -1 + (1 + \beta)\gamma_1(1 - s_t) \\1 + 2\beta &> \gamma_1(1 - s_t) + \beta\gamma_1(1 - s_t) \\1 + 2\beta &> \gamma_1 l_t + \beta\gamma_1 l_t \\(1 - \gamma_1 l_t) + \beta(2 - \gamma_1 l_t) &> 0\end{aligned}$$

■

References

- Almeida, Rosangela, and Zelia Bianchini 1998. Aspectos de amostragem da pesquisa de economia informal urbana. Discussion Paper no.89, IBGE-Brasil.
- Amaral, Pedro, and Erwan Quintin 2006. A competitive model of the informal sector. *Journal of Monetary Economics* 53: 1541-1553.
- Antunes, Antonio, and Tiago Cavalcanti 2007. Start up costs, limited enforcement, and the hidden economy. *European Economic Review* 51:203-224.
- Arias, Omar, Kevin F. Hallock, and Walter Sosa-Escudero 2001. Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics* 26: 7-40.
- Bargain, Olivier and Prudence Kwenda 2009. The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data, UCD Centre for Economic Research, Working Paper Series, WP09/05.
- Belzil, Christian 2007. The return to schooling in structural dynamic models: a survey. *European Economic Review* 51:1059-1105.
- Botelho, F. and V. Ponczek. segmentation in the Brazilian Labor Market. Economic Development and Cultural Change (forthcoming).
- Buchinsky, Moshe 2001. Quantile regression with sample selection: estimating women's return to education in the U.S.. *Empirical Economics* 26:87-113.
- Busato, Francesco, and Bruno Chiarini 2004. Market and underground activities in a two-sector dynamic equilibrium model. *Economic Theory* 23:831-861.
- Card, David. 1995. Earnings, schooling, and ability revisited. In: Polachek, S. (ed.), Research in labor economics. vol. 14. Greenwich: JAI Press.
- _____. 1999. The causal effect of education on earnings. In: Ashenfelter, Orley, and David Card (eds.) Handbook of labor economics. vol. 3A. Elsevier.
- _____. 2001. Estimating the Return to Schooling: Progress on some persistent econometric problems. *Econometrica* 69(5), pp.1127-1160.
- Carneiro, Pedro, James Heckman, and Edward Vytlacil 2000. Estimating the returns to education when it varies among individuals. Working Paper, Department of Economics, University of Chicago.
- Carneiro, Francisco, and Andrew Henley 2001. Modelling formal vs. informal employment and earnings: microeconomic evidence for Brazil. U of Wales at Aberystwyth Management & Business Working Paper no. 2001-15. <http://ssrn.com/abstract=289826> or DOI: 10.2139/ssrn.289826
- Chernozhukov, Victor and Christian Hansen 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132:491-525.

- _____. 2008. Instrumental variable quantile regression: a robust inference approach. *Journal of Econometrics* 142:379–398.
- Choi, Jay, and Marcel Thum 2005. Corruption and the shadow economy, *International Economic Review* 46:817-836.
- Dabla-Norris, Era, Mark Gradstein, and Gabriela Inchauste 2008. What causes firms to hide output? the determinants of informality. *Journal of Development Economics* 85:1–27.
- de Soto, Hernando 2000. The mystery of capital: why capitalism triumphs in the west and fails everywhere else. New York: Basic Books.
- Dessy, Sylvain, and Stéphane Pallage 2003. Taxes, inequality and the size of the informal sector. *Journal of Development Economics* 70:225–233.
- Dickens, William, and Kevin Lang 1985. A test of dual labour market theory, *American Economic Review* 75:1-22.
- Earle, John, and Zuzana Sakova 2000. Business start-ups or disguised unemployment? evidence on the character of self-employment from transition economies. *Labour Economics* 7:575-601.
- Filho, N. A. M., Mendes, M., and de Almeida, E. S. (2004) “O Diferencial de Salarios Formal- Informal no Brasil: Segmentacao ou Vies de Selecao?” *Revista Brasileira de Economia* 58(2): pp. 235–248.
- Fortin, Bernard, Nicolas Marceau, and Luc Savard 1997. Taxation, wage controls and the informal sector. *Journal of Public Economics* 66: 293-312.
- Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton 2000. Dodging the grabbing hand: the determinants of unofficial activity in 69 countries, *Journal of Public Economics* 76:459-493.
- Fugazza, Marco, and Jean-Francois Jacques 2003. Labor market institutions, taxation and the underground economy. *Journal of Public Economics* 88:395-418.
- Gauthier, Bernard, and Mark Gersowitz 1997. Revenue erosion through tax exemption and evasion in poor countries, *Journal of Public Economics* 64:404-24.
- Jorge, Angela, and Lucilia Valadao 2003. Brazilian survey of the urban informal sector: main results and methodological evaluation. IBGE Brazil. http://mospi.nic.in/informal_paper_2003_03.htm
- Johnson, Simon, Daniel Kaufmann, and Pablo Zoido-Lobaton 1998. Regulatory discretion and the unofficial economy. *American Economic Review* 88:387-392.
- Heckman, James 1979. Sample selection bias as specification error. *Econometrica* 47:153-161.
- Koenker, Roger, and Gilbert Bassett (1978). Regression quantiles. *Econometrica* 46:33-50.
- Lemieux, Thomas, Bernard Fortin, and Pierre Frchette 1994. The Effect of taxes on labor supply in the underground economy. *American Economic Review* 84:231-254.

- Light, Audrey 2001. In-school work experience and the returns to schooling. *Journal of Labor Economics*. 19:65-93
- Magnac, Thierry 1991. Segmented or competitive labour markets. *Econometrica* 59:165-87.
- Manski, Charles 1989. Anatomy of the Selection Problem. *Journal of Human Resources* 24:343-360.
- Mattos, Enlison and Laudo Ogura 2008. Skill differentiation between formal and informal employment. *Journal of Economic Studies*. Forthcoming.
- Maloney, Willian 1999. Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *World Bank Economic Review* 13:275-302.
- Maloney, W. F. (2004) "Informality Revisited." *World Development* 32(7): pp. 1159–1178.
- Mincer, Jacob 1974. *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
- Montenegro, Claudio 2001. Wage distribution in Chile: does gender matter? A quantile regression approach. World Bank Development Research Group Working Paper no. 20.
- Paula, Aureo, and Jose Scheinkman 2007. The informal sector. Pennsylvania Institute for Economic Research Working Paper 07-035, University of Pennsylvania.
- Pratap, Sangeeta, and Erwan Quintin 2006. Are labor markets segmented in developing countries? A semiparametric approach. *European Economic Review* 50:1817-1841.
- _____. 2006a. The Informal Sector in Developing Countries: Output, Assets and Employment. UNU-WIDER. Research Paper no. 2006/130
- Rauch, James 1991. Modelling the informal sector formally. *Journal of Development Economics* 35:33-48.
- Schneider, Friedrich, and Dominik Enste 2000. Shadow economies: size, cause and consequences. *Journal of Economics Literature* 38:77-114.
- Straub, Stephane 2005. Informal sector: the credit market channel. *Journal of Development Economics*. 78:299-321.
- Tannuri-Pianto, M., and Pianto, Donald 2002. Informal employment in Brazil – a choice at the top and segmentation at the bottom: a quantile regression approach. Proceedings of the XXIV Meetings of the Brazilian Econometrics Society.
- Thomas, Jim (1992). *Informal Economic Activity*, London School of Economics Handbooks in Economics, London School of Economics: London.
- Ulyseia, G. (2005) "Informalidade no Mercado de Trabalho Brasileiro: Uma Resenha da Literatura." TR. 1070, Instituto de Pesquisas Economicas Aplicadas - IPEA.
- Ulyseia, G. (2010) "Regulation of entry, labor market institutions and the informal sector". *Journal of Development Economics* 91, 87-99.