# An Empirical Evaluation of an Evolutionary Game Theory Model of the Labor Market

Araujo, Ricardo Azevedo; Loureiro, Paulo Roberto and Souza, Nathalia Almeida
University of Brasilia

2011

# An Empirical Evaluation of an Evolutionary Game Theory Model of the Labor Market

Ricardo Azevedo Araujo
University of Brasilia, Brazil

Paulo Roberto Amorim Loureiro
University of Brasilia, Brazil

Nathalia Almeida de Souza
Anatel and University of Brasilia, Brazil

**Abstract**

In this paper we intend to perform an empirical evaluation of the evolutionary game theory model of the labor market developed by Araujo and Souza (2010). In order to accomplish this task we focus on the Brazilian labor market by using data from the National Household Sampling Survey – PNAD/IBGE, from 1995 to 2008. We used four different methodologies: the OLS, Pseudo-panel with fixed effects, Instrumental Variables and the Heckman Selection Model. Results indicate that the main difference between the 1995-2002 and 2003-2008 period is the impact of education over wages. According to these findings, investments in education were more profitable for the 2003-2008 period. However, all wage gaps between formal and informal markets reduced considerably.

**Keywords:** formal and informal and labor market, evolutionary game theory.

**JEL Classification:** E26, J62, C73

## 1. Introduction

There are many ways of defining informality and each definition may give rise to a different size to the informal market. Such activity may include self-employment, employment in micro-businesses or family run activity, as well as employment without appropriate access to social protection or formal registration. [see Henley et al. (2009)]. In the present paper we define informality as tax avoidance: firms and workers in the informal sector avoid tax payments but suffer other constraints such as limited access to legal protection, hedge and credit. Following this criteria De Paula and Scheinkman (2007) have estimated that approximately 10% of GDP in the United States was produced by individuals or firms that evaded taxes or engaged in illegal pursuits. It is also estimated that these activities produce 25 to 35% of aggregate output in Latin America, between 13 to 70% in Asian countries, around 15% in O.E.C.D. countries[1].

In this paper we do not treat informality as the result of exclusion but rather as the outcome of rational decisions taken by firms and workers. We acknowledge that the issue has many dimensions but we focus on the informality that arises as the optimal outcome of rational agents playing their best responses in a specific institutional set-up. This view is according to Hirschman (1970) who considers that workers and firms make implicit cost-benefit analysis when deciding to act in formal or informal sectors and it has found empirical support in a number of papers. In Maloney (2004), for instance, informal urban micro-firms can be seen as agents that, on behalf of the lack of enforcement, may opt for which degree of formal activity they wish to produce. He also observes that, even at lower wages, workers in these economies choose informality due to a level of dignity and autonomy that this type of work might offer.

Following this rationale, Fortin et al (1997) have developed a model with firm heterogeneity in which a formal and an informal market emerge in some sectors of the economy

[1] The World Bank estimates that this informal economy generates 40 percent of the GNP of low-income nations and 17 percent of the GNP of high-income ones. In some industries, such as retailing and construction, informality can account for as much as 80 percent of employment.

as the optimal response against the effects of taxation and wage controls. By studying the informal sector in the Republic of Cameroon, these authors have found that the average wage paid in the formal sector is higher than the average wage in the informal sector. This view was disputed by Mattos and Ogura (2008) studying the Brazilian economy: controlling for individual characteristics it is not possible to reject the hypothesis that wages in the formal sector are similar to wages in the informal sector. Besides they have concluded that in the same industry more efficient firms are formal while the less efficient ones are informal. They assume that the cost of informality increases with production, and this implies that formal firms have a greater probability of employing more workers than informal ones.

While most of these studies focus on informality from the viewpoint of firms Loayza and Rigolini[2] (2006) found that the size of informal employment is given by the proportion of workers whose skills fall below a threshold level where the worker is indifferent between the two sectors. They have found that informality not only responds to fundamental, long-run forces but also to inter-temporal economic conditions related to the business cycle and transient policies. Thus, for example, the informal sector could act as a buffer that expands in economic recessions or as an adjustment mechanism during temporarily high tax regimes. Fiess et al. (2010) have developed a labor market model and embedded it in a standard macroeconomic framework that allows capturing additional information on the sectoral origin of the shocks through the real exchange rate — a measure of relative prices of tradables and non-tradables. The findings suggest that the pro or counter-cyclicality behavior of the informal sector is more complex than that reported by Loayza and Rigolini (2006), since it depends on the sectoral origin of the shocks and the presence of binding wage rigidities. They found numerous examples

[2] Other papers that investigate causes and determinants of informality include Loayza (1996), Loayza et al (2005a, 2005b) and Maloney (2004). All of them point to some positive relation between the size of the informal sector and higher taxes, more labor market restrictions, and poorer institutions (bureaucracy, corruption and legal environment).

where either a positive productivity or demand shock to the non tradable/informal sector leads to its expansion.

Albretch et al. (2008) have built a search and matching model to analyze the effect of labour market policies in an economy with a significant informal sector. By running numerical simulations they found that a severance tax greatly increases average employment duration in the formal sector, reduces overall unemployment, reduces the number of formal sector workers, and reduces the number of workers who accept any type of offer (formal or informal). A payroll tax reduces average employment duration in the formal sector, greatly reduces the number of formal-sector workers, and significantly increases the size of the informal sector and the number of workers accepting any type of offer.

In this paper we intend to perform an empirical evaluation of the evolutionary game theory model of the labor market developed by Araujo and Souza (2010) by focusing on the Brazilian labor market by using data from the National Household Sampling Survey – PNAD/IBGE, from 1995 to 2008. According to this modeling workers and firms' decision to engage in the formal or informal sector arise as the outcome of bounded rational decisions based not only on the expected pay-offs in each of the sectors but also on the interaction with other agents. We must recall that preferences are wide-ranging and can include both pecuniary and non-pecuniary aspects of work. In fact, wages and tax schemes although important determinants of utility, are not unique. Other factors such as autonomy, flexibility, working hours, distance to work and opportunities offered in the informal sector also determine job satisfaction and may lead workers to choose informal employment[3].

Our empirical findings show that although wages play an import role in the decision of workers and firms all wage gaps between formal and informal markets reduced considerably as been pointed out by Mattos and Ogura (2008). However, an increase in one year of study would raise the probability of participating in the formal sector while diminishing the probability of

[3] For further information see Mulinge and Mueller (1998) and Saavedra and Chong (1999).

joining the informal sector. Besides, the skill premium in the formal sector is higher than in the informal sector.

This paper is structured as follows: in the next section we present the model with its main properties. Section 3 shows empirical evidence of determinants for labor participation in informality in Brazil. Section 4 brings conclusions to the study

## 2. Theoretical Model

The model departs from Araujo and Souza (2010) and corresponds to an asymmetric evolutionary game where there are two populations of interacting agents [See Gintis (2000)]: workers and firms. It is assumed that each identical worker has two possible strategies that is, supply labour in either formal or informal market at each period of time. Let $N$ be the number of workers, $N_f$ the number of workers that choose to supply labour in the formal sector – the formal strategy – and $N_i$ be the number of workers that choose the informal sector – the informal strategy. Let $n_f$ and $n_i$ be the proportions of workers that chooses the formal and informal strategies respectively, with $n_i + n_f = 1$. By choosing a strategy does not mean that the worker will be employed since to be hired it depends on matching a firm that has chosen the same strategy. Otherwise the worker will be unemployed. If she chooses the formal strategy then there exists a probability $\sigma$, $0 \leq \sigma \leq 1$, of finding a job in a formal firm. In this vein her instantaneous expected utility, $U_f^e$, is assumed to be given by:

$$U_f^e = \sigma(1-\tau)w_f \tag{1}$$

Where $w_f$ is the real wage discounted by the income tax $\tau$, $0 < \tau < 1$. Expression (1) shows that if the worker chooses the formal strategy there is no probability of punishment but she faces uncertainty related to finding or not a firm that also chooses the formal strategy to hire

her, what happens with probability $\sigma$. However, if the worker decides to act in the informal sector his expected utility, $U_i^e$, is given by[4]:

$$U_i^e = \phi[(1-\rho)w_i + \rho(w_i - m)] \quad (2)$$

Where □ , $0 \leq \phi \leq 1$, is the probability of finding a job in the informal sector and $w_i$ is the wage paid in the informal sector. The probability of being caught due to the operation in the informal sector is given by $\rho,\ 0 \leq \rho \leq 1$ . In this case the worker pays a fine, denoted by *m*, due to the choice of acting in the informal sector. These variables are assumed to be exogenous. Expression (2) shows that the worker who chooses the informal strategy faces two kinds of uncertainty: the first is related to the possibility of not finding a firm that chooses the informal strategy and the second is related to the possibility of being caught if hired by an informal firm.

In order to model the demand side of the labour market, let us assume following the literature of search and matching – see e.g. Pissarides (2000) – that the number of firms, denoted by *L*, is equal to the number of workers[5], that is $L = N$. Let $L_f$ be the number of firms that chooses the formal strategy and $L_i$ the number of firms that chooses the informal strategy. Analogous to the case of labour supply, each firm can demand labour in only one of the markets in each period of time. Let $\eta_i$ be the proportion of firms that chooses the informal strategy and $\eta_f$, the proportion of firms that chooses the formal strategy, with $\eta_i + \eta_f = 1$.

[4] An important difference between this approach and the one developed by Fortin et al (1997) is that in our model we model explicitly the possibility of being caught due to the operation in the informal sector while they consider that the firm in the informal sector faces a cost in order to avoid to be caught. The insight is that the higher the production of the firm the higher the cost in order to conceal its production.

[5] This is a usual assumption in the search and matching models and here it is adopted for tractability only. For a treatment of the labor market dynamics by using an evolutionary model in which the processes of vacancy setting is modeled through a process of searching and matching see Fagiolo et al (2004).

Following Pissarides (2000) assume that each firm hires only one worker who produces a fixed amount of product at a time. The price of the product is normalized to 1 and the amount of production in the formal sector is exogenously given by $y_f$. Being $\theta$, $0 \leq \theta \leq 1$ the probability of a firm that chooses the formal sector to find a worker that decides to supply labour in this sector, the profit of the firm if it decides to operate in the formal sector is given by:

$$\prod_f^e = \theta[(1-\gamma)y_f - w_f] \tag{3}$$

Where $\gamma$, $0 < \gamma < 1$, stands for the costs for being in the formal sector. Expression (3) shows that each firm has to pay $\gamma y_f$ as taxes. Both $y_f$ and $\gamma$ are assumed to be exogenous. If there is no matching between the formal worker and the formal firm then the profit of the firm is equal to zero, what occurs with a probability $1 - \theta$. In the informal sector the firm is also assumed to hire only one worker, but now it produces a smaller amount of product than in the formal operation due to limited access to public goods, capital goods etc. Let $y_i$ be the amount of product in informal operation, with $y_i < y_f$. In this vein the profit of the firm in the informal sector is given by:

$$\prod_i^e = \lambda\{(1-\psi)[y_i - w_i] + \psi[y_i - w_i - e]\} \tag{4}$$

Where $\lambda$, $0 \leq \lambda \leq 1$ is the probability of matching a worker in the informal sector, and $\psi$, $0 \leq \psi \leq 1$, is the probability[6] that the firm faces of paying a fine, expressed by $e$, due to the operation in the informal labour market. After some algebraic manipulations expression (4) yields:

$$\prod_i^e = \lambda(y_i - w_i - \psi e) \tag{5}$$

[6] We assume that this probability is the same of finding a worker in the informal sector. This assumption is made for the sake of convenience only but it expresses the fact that once a worker in the informal sector is detected then the corresponding firm is also found.

Since it is assumed that each firm hires only one worker the ratio of labour demanded in the formal sector, $\eta_f$, and the ratio of labour demanded in the informal sector, $\eta_i$, is proportional to the amount of firms in each sector. It is important to recall that if a worker who chooses the formal strategy does not match a firm with this strategy – an informal firm – then the pay-off of both worker and firm will be equal to zero. In this case the firm is assumed to produce zero output and the worker does not earn wage. This situation can be identified as unemployment from the viewpoint of the worker. We could assume alternatively that if a worker that chooses the formal sector matches a firm in the informal sector then both will obtain positive pay-offs but smaller than the pay-offs if both worker and firm choose the formal sector or informal sector simultaneously. It is easy to see that this game has two pure Nash equilibria namely {*f*,*f*} and {*i*,*i*} together with a mixed strategy equilibrium, in which both workers and firms randomly choose between being formal or informal.

In order to evaluate the dynamics of entrance and withdrawal of workers in the formal market we use a version of the dynamic replicator as proposed by Hofbauer and Sigmund (2003) adapted to the study of the labour market according to Araujo and Souza (2010). The dynamic movement of workers between the two strategies, namely formal and informal is given by the following expressions:

$$\dot{N}_f = N_f\left[U_f^e - \overline{U}_{f,i}\right] \tag{6}$$

$$\dot{N}_i = N_i\left[U_i^e - \overline{U}_{f,i}\right] \tag{7}$$

Where $\overline{U}_{f,i}$ is the average pay-off given by: $\overline{U}_{f,i} = n_f U_f^e + n_i U_i^e$. By inserting expressions (1) and (2) into (6) and (7) it is possible to show after some algebraic manipulation[7] that it yields the following equations for the dynamic behaviour of the ratios of workers in the formal and informal sectors.

[7] See Araujo and Souza (2010) for the derivation of expressions (8) and (9) from (3) and (4).

$$\dot{n}_f = n_f n_i \{\sigma[(1-\tau)w_f] - \phi u[w_i - \rho m]\} \quad (8)$$

$$\dot{n}_i = n_f n_i \{\phi[w_i - \rho m] - \sigma[(1-\tau)w_f]\} \quad (9)$$

These expressions show that the central planner can affect the supply of the labour in each sector by choosing the taxation, $\tau$, the probability of caught the worker in the informal sector, $\rho$ , and the fine to be paid in the informal sector, $m$. Until this point of the analysis the values of $\sigma$ and $\phi$ are exogenously considered but a further inquire on this probabilities by using a Bayesian inference may show that $\sigma = \eta_f$ and $\phi = 1 - \eta_f$ . Remember that firms have only two strategies, namely formal and informal. Even in the case where there is no matching between a firm choosing the formal strategy and a worker choosing the informal strategy their strategies are 'formal' and 'informal' despite the fact that the worker will be unemployed and the firm will produce nothing in that period of time. Hence all firms can be grouped into one of these categories: 'formal' or 'informal'. The probability that a worker faces of finding a 'formal' firm is given by $\sigma = \frac{L_f}{L} = \eta_f$ and the probability of finding a 'informal' firm is given by $\phi = \frac{L_i}{L} = \eta_i$ . Hence expression (8) may be rewritten as:

$$\dot{n}_f = n_f n_i \{\eta_f[(1-\tau)w_f] - (1-\eta_f)[w_i - \rho m]\} \quad (8)'$$

Following the same approach for the labour demand, the dynamic replicators for the firms are given by:

$$\dot{L}_f = L_f\left(\Pi_f^e - \overline{\Pi}_{f,i}\right) \quad (10)$$

$$\dot{L}_i = L_i\left(\Pi_i^e - \overline{\Pi}_{f,i}\right) \quad (11)$$

Where $\Pi_f^e$ stands for the expected profit of the formal strategy and $\Pi_i^e$ stands for the expected profit of the informal strategy and $\overline{\Pi}_{f,i}$ represents the average expected profit in the economy which is the average payoff for firms, given by: $\overline{\Pi}_{f,i} = \eta_f \Pi_f^e + \eta_i \Pi_i^e$ . By replacing expressions

(4) and (5) into expressions above and considering that $\eta_i + \eta_f = 1$ we obtain the following dynamic replicator in the simplex form:

$$\dot{\eta}_f = \eta_f \eta_i \{\theta[(1-\gamma)y_f - w_f] - \lambda[y_i - w_i - \rho m]\} \quad (12)$$

$$\dot{\eta}_i = \eta_i \eta_f \{\lambda[y_i - w_i - \rho m] - \theta[(1-\gamma)y_f - w_f]\} \quad (13)$$

By following the same rationale adopted for the labor supply it is possible to conclude that $\theta = n_f$ and $\lambda = 1 - n_f$. Expression (12) may then be rewritten as:

$$\dot{\eta}_f = \eta_f \eta_i \{n_f[(1-\gamma)y_f - w_f] - (1-n_f)[y_i - w_i - \rho m]\} \quad (12)'$$

The equilibrium or steady state solution of the model is obtained by considering that: $\dot{n}_f = \dot{\eta}_f = 0$. From expression (8)'' we have three possibilities, namely: $n_f = 0$, $n_f = 1$ or $f(\eta_f) = 0$. From expression (12)' we also have three possibilities, namely: $\eta_f = 0$, $\eta_f = 1$ or $g(n_f) = 0$. Hence, from the combination of these possibilities we are left with five possible cases that may be expressed as: (i) $n_f = 0$, $\eta_f = 0$; (ii) $n_f = 0$, $\eta_f = 1$; (iii) $n_f = 1$, $\eta_f = 0$; (iv) $n_f = 1$, $\eta_f = 1$; (v) $n_f = \dfrac{y_i - w_i - \psi e}{(1-\gamma)y_f - w_f + y_i - w_i - \psi e}$ and $\eta_f = \dfrac{[w_i - \rho m]}{[(1-\tau)w_f] + [w_i - \rho m]}$.

Araujo and Souza (2010) have shown that the most probable outcome is the fifth equilibrium in which the final solution is affected by the parameters chosen by the government such as taxation and fines. In the next section we perform an empirical evaluation of this model in order to investigate if economic variables such as wages, taxation and fines are the determinants of the workers' and firms' choices in the labor market. A number of other characteristics such as gender and race are also considered.

## 3. Empirical Analysis

In this section we aim to analyze the main factors that drive the Brazilian labor when deciding to participate in the formal or informal sectors. We accomplish this analysis by using four different methodologies, ascertaining for sample selection and endogeneity.

### *3.1. Data*

The data for this study were obtained at the National Household Sampling Survey – PNAD/IBGE, from 1995 to 2008, with an exception to the years of 1994 and 2000, when this survey did not occur. The cross-section data were pooled and divided in two periods: 1995 – 2002 and 2003 – 2008. The sample ranges individuals from 22 to 57 years of age, and the dependent variable is the logarithm of the real annual income (LOGSALM), deflated by consumer prices index – IPCA. The explaining variables in our analysis are: race (WHITE), gender, wages (SAL.M), years of study (YOSTUDY), square years of study (YOSTUDY2), urbanization index (URBAN), worked hours, union (UNION), formal (FORMAL) and informal (INFORMAL) markets, private sector, experience (EXPER) – defined by age less years of study, minus six – square experience (EXPER2), experience times education (EXPERYOSTUDY). Finally, other variables that can be included in the labor equation appear: family reference person (PESREF), age (AGE), square age (AGE2), kids (KIDS), firms with less than 11 employees (ESIZE), married man (MARRIEDM) and a dummy to indicate if the person lives in an urban area or not (URBAN).

We defined two sub-samples dedicated to the private sector. The first sample consists of 298.443 workers for the period of 1995-2002. The second sample consists of 345.032 workers for the 2003-2008 period. In each sub-sample, workers were split into formal and informal markets. The samples used consist of 219.709 formal workers and 78.838 informal workers for the period of 1995-2002; whereas, 259.460 formal and 90.192 informal workers composed the sample for the 2003-2008 period. Private sector's formal workers is composed by formal

working contracts while informal labor market is constituted of no contract based employees as well as self employment.

Table 1: Description of variables used in the Heckman selection equation for formal and informal labor market

| |
|---|
| Dependent Variables: |
| FlpF = participation of the worker on formal labor market. |
| FlpInf = participation of the worker on informal labor market. |
| Explaining Variables: |
| Yostudy = Years of study. |
| White = 1 if white, 0 otherwise. |
| Pessref = 1 if family reference, 0 otherwise. |
| MarriedM = 1 if married man, 0 otherwise. |
| Exper = Experience = age – education - 6. |
| Exper2 = Square Experience. |
| yostudy2 = Square years of study. |
| Experyostudy = years of experience times education. |
| Urban = urbanization index. |
| Tenure = years spent at the last job. |
| age = age in years. |
| Esize= Size of firms with up to 10 employees. |
| Age2 = Square Age. |
| Union = 1 if associated to union, 0 otherwise. |
| Lambda = inverse Mill`s ratio. |

Table 2: Description of variables used in labor market income equations

| Variables |
|---|
| ***Lnwage = Natural Logarithm of wage per worked hour (dependent variable).*** |
| White = 1 if white, 0 otherwise. |
| MarriedM = 1 if married man, 0 otherwise. |
| Exper = Experience = age – education – 6. |
| Exper2 = Square experience. |
| Yostudy = Years of study. |
| Experyostudy = Years of study times education. |
| Esize = Size of Firm with up to 10 employees |
| Kids = 1 if family have kids, 0 otherwise. |
| Lambda = inverse Mill's ratio. |

### *3.2. Descriptive analysis*

Assessing if there were any substantial changes in the private sector related to the formal and informal labor markets between the two sub-periods under analysis, we present the descriptive statistics on Table 1. First, we divided the samples into three different groups according to their age: ***group1*** is an array containing individuals from 22 to 33 years of age; ***group2*** goes from 34 to 45 years; ***group3*** ranges individuals with 46 to 57 years of age.

For the years of 1995-2002, group1 presents a wage differential between formal and informal workers of 22,19%, equivalent to a R$97,83 surplus. For group2, however, the wage differential is not so deep being only 3,40% higher than informal workers' wages. Group3 shows a mean of 18,22% higher wages for formal sector laborers, representing a R$92,84 gain for formality.

Analyzing the 2003-2008 period, group1's formal laborers obtain 20,32% higher wages than informal workers, representing a mean wage of R$ 812,16, in contrast to a R$ 675,01.

Thus, formal workers receive a R$ 137,15 bonus for legality. Group2, however, presents a lower differential for legality, such as in the first period's analysis. Here, formal workers have only a surplus of R$ 83,21, being formal wages 10,83% higher than informal ones. In the third group, the differential in wages rises, where formality shows a gain of 20,08% over informality, presenting a R$ 866,38 against a R$ 745,24. These findings contradicts the prediction of the model proposed by Rosen (1986), in which average earnings should be higher in the less desirable informal sector to compensate for the non-pecuniary benefits granted to registered workers and reinforces the findings of Fortin et al. (1997).

Here we might be observing the life cycle working over differential gains of formality/informality. Youngsters, or group 1, have problems in entering the labor market and therefore accept greater gain differentials due to job opportunities in either markets. They are also in the period of gaining experience and education in counterpart to monetary gaining objectives. On the other hand, group 2 is at the age of capturing money for later consumption, thus they are less willing to accept wage differentials. In this group we may be observing voluntary informality, and not the traditional view that informality is due to the lack of formal jobs. Members of group 3 are closer to retirement. Thus, many might face difficulties in finding a formal placement due to their age. Therefore, wage differentials among formal and informal jobs are to be expected.

The results show that in the last six years, the gap between formal and informal wages grew, what indicates that Public Welfare revenues increased. According to Gobetti and Orair (2010), fiscal revenues increased 1,93% during the 2003-2008, going from R$ 542.863 million, in 2003, to R$ 1.047.194 million in 2008. Despite the increase in gross base product for taxation, relative taxation burden augmented from 31,93% in 2003 to 34,85% of the GDP in 2008. Formal workers' real mean wages for all three groups of the 2003-2008 period are higher than those of the precedent period.

We must conclude, thus, that the gap increase among formal and informal wages is related to Brazilian economy's growth. Despite the increase of labor participation in the public sector, the wage gap among formality and informality remains significant. We can realize that when there is economic growth, the burden of being in the informal sector also grows, represented by the increase in the wage gap in the latter period analyzed. Thus we can perceive that access to public services becomes more relevant. It is also true that the size of the production with economic growth increases, which makes it easier for public monitoring to seize those who are under informal activity. Thus informal firms must remain small. As such, a 2,6% increase in the mean value for firm size (esize) with up to 10 employees is observed for the informal market, for the 2003-2008 cycle . Table 1 shows a 50,1% growth in the mean informality wage, what represents greater interest of small firms for the execution of smaller projects.

Descriptive statistics of table 1 also show that great part of younger workers – group1, are more present in informality for both periods. This represents that 56% of informality is in group1 in the first period and 55% in the second one. Group2, nonetheless, is more concentrated on formal market in both periods. As such, we tend to believe that, as age increases so does human capital, increasing formal market opportunities.

One of the most relevant variables for determining wage is education. When assessing table 1, we verify that education increased for both markets. Among formality, average years of education went from 7,99 to 9,24 in 2003-2008, whereas, in informal markets, the average went from 6,55 years to 7,78. When evaluating experience impact on data, we observe that, for the first period, formal workers possess 19,12 years, while informal workers had 20,05 years. In the second period of analysis this gap is reduced to 18,09 years for formal and 18,97 years for informal workers. However, it is interesting to note that experience's mean fell.

It is also observable from table 1 that the concentration of male workers is considerably higher in the formal sector for the 1995-2002 cycle, representing 67% of formal labor force. This massive participation continues for the second period, despite reducing to 64%. This male

predominance also presents itself in informality for both periods analyzed. The participation of non-white individuals in formal and informal labor markets showed an increase from the first to the second period analyzed. We can verify that white participation fell from 58% to 53% in formality and from 49% to 44% in informality.

Tenure in formality is higher than in informal markets. Under formality, tenure is around five years for both periods; while under informality, this value reduces to 3 years. In the 1995-2002 period, only 30% of formal workers were associated to union. This value grew considerably in the 2003-2008 cycle, where 45% of these laborers were under union filiations. Informal sector, however, presented only 5% filiations for both periods.

**Table 1 Descriptive Statistics – Formal and informal markets for periods of 1995-2002 and 2003-2008.**

| | **1995-2002** | | | | **2003-2008** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Formal** | | **Informal** | | **Formal** | | **Informal** | |
| | **Mean** | **sd** | **mean** | **Sd** | **mean** | **sd** | **mean** | **sd** |
| **salm** | 582.07 | 748.55 | 386.10 | 597.80 | 870.49 | 1016.89 | 747.25 | 1212.93 |
| **yostudy** | 7.99 | 4.01 | 6.55 | 4.13 | 9.26 | 3.80 | 8.31 | 4.47 |
| **exper** | 19.12 | 10.41 | 20.05 | 10.77 | 18.08 | 10.63 | 18.20 | 11.38 |
| **exper2** | 474.04 | 474.32 | 518.09 | 511.45 | 439.68 | 464.64 | 460.51 | 502.77 |
| **experyostudy** | 130.35 | 84.42 | 108.61 | 84.19 | 145.01 | 92.20 | 121.79 | 94.89 |
| **esize** | 0.77 | 0.42 | 0.35 | 0.48 | 0.77 | 0.42 | 0.36 | 0.48 |
| **kids** | 0.21 | 0.41 | 0.25 | 0.43 | 0.22 | 0.41 | 0.21 | 0.40 |
| **white** | 0.58 | 0.49 | 0.49 | 0.50 | 0.53 | 0.50 | 0.45 | 0.50 |
| **marriedM** | 0.57 | 0.50 | 0.56 | 0.50 | 0.53 | 0.50 | 0.40 | 0.49 |
| **yostudy2** | 98.51 | 74.45 | 75.19 | 69.98 | 120.78 | 73.63 | 107.41 | 82.42 |
| **age2** | 1245.77 | 650.39 | 1218.48 | 686.31 | 1280.35 | 677.65 | 1442.38 | 783.23 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **age** | 34.18 | 8.81 | 33.65 | 9.27 | 34.61 | 9.09 | 36.59 | 10.19 |
| **experyostudy** | 130.35 | 84.42 | 108.61 | 84.19 | 145.01 | 92.20 | 121.79 | 94.89 |
| **union** | 0.30 | 0.46 | 0.05 | 0.22 | 0.28 | 0.45 | 0.10 | 0.30 |
| **pessref** | 0.56 | 0.50 | 0.51 | 0.50 | 0.51 | 0.50 | 0.42 | 0.49 |
| **urban** | 0.96 | 0.20 | 0.93 | 0.26 | 0.96 | 0.21 | 0.86 | 0.34 |
| **group1** | 0.53 | 0.50 | 0.56 | 0.50 | 0.51 | 0.50 | 0.44 | 0.50 |
| **group2** | 0.34 | 0.47 | 0.30 | 0.46 | 0.34 | 0.47 | 0.33 | 0.47 |
| **group3** | 0.13 | 0.33 | 0.14 | 0.34 | 0.15 | 0.36 | 0.23 | 0.42 |
| **male** | 0.67 | 0.47 | 0.70 | 0.46 | 0.64 | 0.48 | 0.51 | 0.50 |
| **female** | 0.33 | 0.47 | 0.30 | 0.46 | 0.36 | 0.48 | 0.49 | 0.50 |
| **tenure** | 5.07 | 5.65 | 3.52 | 5.15 | 5.18 | 5.79 | 6.59 | 7.92 |

### *3.3. Estimation*

The econometric methods applied in this study are well known in the education return literature. The econometric equations are distinguished in four different estimation methods. The **first estimation** method is an **OLS** – ordinary least squares, which we use to compare to the other estimations.

However, there exists bias in the common OLS estimation of education impact over wages. This bias occurs through two main channels: individual ability or sample selection. The first bias, the individual ability issue, is due to the unobserved variable, that cannot be measured, and that influences individual's gains. This unobservable variable is consisted of abilities such as intelligence, motivation and dedication. To control for this bias, we use a **second estimation** model: **pseudo-panel**[8].

Heckman (1979) states that even if OLS methodology presents relevant results, there always lies a chance of these estimates being biased due to the omission of some regressors. This Pseudo-panel methodology also controls for this unaware omission. The Pseudo-panel

[8] The empirical treatment was based on pseudo-panel methodology used in Deaton (1985).

estimation can be done basically in two ways: (i) with a random effect, where the unobservable variable is random and the estimation is through a Feasible general least square, (ii) or through a Fixed effect model, where the unobservable variable has to be controlled for. To decide whether to use a random or fixed effect modeling, we apply a Hausman test. According to these results, the null hypothesis, that the fixed effect is not correlated to the explaining variables, is rejected and thus we use a fixed effects model.

However, even when controlling for unobserved effects, there may be an endogeneity problem within the estimation. When endogeneity is present in the explaining variable, caused by correlation between the error term and the explaining variables, the use of a **third method** to control for this bias is indicated: the **instrumental variable method** (IV). We then also apply this method, checking for endogeneity through a Sargan test.

According to Heckman's (1974, 1979) studies, the individual's decision for participating in the labor market depends on a reserve wage. The worker decides to accept a job offer, that is, decides to participate in the labor market if the wage offered is able to cover his opportunity cost of working, the reserve wage. However, in order to correct the sample selection bias caused by supply and demand of labor, we estimate a **fourth equation**, the **Heckman selection model**.

***Ordinary Least Squares***

The first estimation model we use is a Ordinary least Square. The OLS model can be obtained as follows:

$$\ln w_{ij} = \alpha + \Sigma \beta_i x_i + \varepsilon_i \tag{13}$$

Where ln*wij* is the logarithm of worker's *i* wage in the *j* labor market; $\alpha$ is the intercept; *x* is the vector of individual characteristics; $\beta i$ are parameters to be estimated and $\varepsilon i$ is the random error, independently and identically distributed.

***Pseudo-Panel with fixed effects***

The main objective of fixed effects is to control for specific characteristics of individuals, that are invariant in time. This method also allows reducing the bias over non specified independent variables. The intercepts $\alpha_i$ vary only from individual to individual and is able to capture all different characteristics along time. We admit that there are many unobserved characteristics that influence individual's economic performance, such as intelligence, motivation and dedication that are very difficult to measure and that are not easily instrumented for[9].

A way to evaluate whether to use a fixed effects model or a random effects model is using a Hausman test. In this test, we evaluate if the unobserved variable are correlated to the independent variables or not, that is, it tests for the presence of endogeneity. This test is constructed as follows:

Null Hypothesis: [ $\mathrm{E}(\alpha_i / x_{it}) = 0$ : non correlated].

Alternative Hypothesis: [ $\mathrm{E}(\alpha_i / x_{it}) \neq 0$ : correlated].

Once we rejected the null hypothesis in our data, we found that the use of random effects is inappropriate for estimating wages. We thus use a fixed effects modeling in the pseudo panel.

The fixed effects model is the following:

$$\ln(y_{it}) = \alpha_i + \beta' x_{it} + \lambda_t + e_{it} \qquad (i = 1,2,.....n; t = 1,...., T) \qquad (14)$$

where $y_{it}$ is a wage vector. The vector of observable individual charateristic is represented by $X_i$; $t$ measures time and n is the number of individuals. $e_{it}$ is an error component, which varies in time and individuals. A specific formulation of this equation, for controlling both for unobservable heterogeneity ($\alpha_i$) and for specific time effects ($\lambda_t$), can be estimated using the technique of fixed effects. These variables are expected to be stable for a short period of time.

***Instrumental Variables***

[9] See Hausman and Taylor (1981) for further information.

In the presence of endogeneity, having found correlation between the error term and the explaining part of the model, we used a model with instrumental variables. The variable that suffers from endogeneity must be instrumented for, so as to control for the bias. However, the instruments used must be non-related to the error term, but highly correlated to the variable being instrumented. This method allows the isolation of the endogenous variable [Arellano, M. and S. Bond (1991); Baum, Schaffer and S. Stillman (2003)], producing more robust estimations. As well discussed in literature, we also treat the variable education in our estimated model as being endogenous.

$$\ln w_{ij} = \alpha + \Sigma\beta_i x_i + \sum \theta_i y_i + \varepsilon_i \tag{15}$$

where $\ln w_{ij}$ ids the logarithm of worker`s i wage in the j labor market. $\alpha$ is the intercept, $x$ is the vector of explaining variables and $\beta_i$ are the parameters to be estimated. $\theta$ represents the instruments in the model and $\varepsilon_i$ is a random error term, independently and identically distributed. The model is estimated using a two stage least square.

***The Heckman Selection Model***

When solving for the sample selection bias, Heckman (1974, 1979 e 1980) proposed a procedure to obtaining consistent estimators. We therefore used this procedure in our labor income equation to correct for the bias generated in the supply and demand for labor market.

First, we estimated the equation for the workers' participation on the labor market over vector X, by using a probit model so as to obtain $\delta / \mu_i$ estimate. Using these estimates, we then obtained the inverse of Mill's ratio ($\lambda$). We then estimated the income equation over *Z*, using Mill's inverse ratio as explaining variable, through an OLS. The equation is given by

$$W_i = \delta Z_i + \lambda_i \Theta + \mu_i \tag{16}$$

where W is the logarithm of wage, *Z* is a vector or personal characteristics, $\delta$ is a vector of parameters and $\mu$ is a random error vector that assumes usual statistical assumptions.

### *3.4. Estimation Results*

To test the hypothesis of whether there are different returns to education among formal and informal labor markets in Brazil, for the 1995-2002 and 2003-2008 periods, we estimate the equation for participation in the labor market. This procedure makes possible identify the factors that affect workers' decision of participating or not of the labor market.

The explaining variables that we use in our model consist of, mainly, personal characteristics. The main personal characteristics are education (ESCOL), experience (EXP) and square experience (EXP2). Despite experience increases the probability for the individual to participate in labor force, this increase happens in a decelerated pace. The level of education also presents direct effect over the decision of participating on labor market. As the level of education increases, so does the opportunity of finding a job. Thus these are the main explaining variables.

However, other personal attributes commonly appear in the literature, such as race and gender. Hence, we also include these variables in our model. Although decision-making to entering labor market is firstly related to the worker's personal characteristics, may it be education, experience, race, gender or status, other variables such as kids, or whether the employer is formal or informal also have to be taken into account. In the appendix we present detailed description of the variables used in this research.

Table 2 presents the results for the Heckman selection equation for formal and informal workers in both 1995-2002 and 2003-2008 periods. All the variables in the model for formal labor present the expected sign. However, the analogous is not true for the informal market model. As an example, the higher the level of education, the lower the probability of obtaining a job in the informal sector. As such, an increase of one year of study would reduce the probability in 17,7% of obtaining a job in the first period and in 20,92%, in the second. The results show that the increase in one year of study would raise in 16,21% the probability of

participating of the formal sector in the first period and for the second, this probability rises to 22,77%.

From Table 2, we can also conclude that urban areas benefit from the availability of formal working posts. As such, formal labors in urban areas had a 16,1% probability of finding a formal job in the 1995-2002 period. However, this probability decreased in the later period to 12,12%. Union filiations strongly contribute to the decision of being formal or informal. Despite the high contribution to formality, filiations among formal workers have thoroughly decreased from 99,2% in the first period to 44,21%, in the 2003-2008 period, as can be seen on Table 2. The reduction of labor force being employed in production, combined with new managerial forms and great firm mobility among one and another sector, as well as the increase in temp jobs, have led to less union filiations.

**Table nº2 Heckman selection equation for formal and informal labor market for periods of 1995-2002 and 2003-2008.**

| | 1995-2002 | | 2003-2008 | |
|---|---|---|---|---|
| | Formal | InFormal | Formal | InFormal |
| | Coef | Coef | coef | coef |
| yostudy | 0.1621*** | -0.1765*** | 0.2277*** | -0.2092*** |
| | (-0.0047) | (0.00460 | (0.0029) | (0.0038) |
| yostudy2 | -0.0068*** | 0.0070*** | -0.0082*** | 0.0080*** |
| | (0.0002) | (0.0002) | (0.0001) | (0.0002) |
| age2 | -0.0007*** | 0.0007*** | -0.0010*** | 0.0009*** |
| | (0.0000 | (0.0000) | (0.0000) | (0.0000) |
| age | 0.0617*** | -0.0646*** | 0.0977*** | -0.0891*** |
| | (0.0031 | (0.0032) | (0.0022) | (0.0025) |

| | | | | |
|---|---|---|---|---|
| experyostudy | -0.0019*** | 0.0020*** | -0.0025*** | 0.0022*** |
| | (0.0001 | (0.0001) | (0.0001) | (0.0001) |
| esize | 1.0117*** | -0.9553*** | 0.9015*** | -0.9645*** |
| | (0.0066 | (0.0067) | (0.0049) | (0.0052) |
| kids | 0.0129 | -0.0106 | -0.0779*** | 0.0457*** |
| | (0.0099) | (0.0101) | (0.0066) | (0.0075) |
| white | 0.1412*** | -0.1270*** | 0.1168*** | -0.1415*** |
| | (0.0066) | (0.0068) | (0.0047) | (0.0052) |
| marriedM | -0.0471*** | 0.0388*** | 0.0783*** | -0.0490*** |
| | (0.0077) | (0.0080) | (0.0051) | (0.0057) |
| union | 0.9920*** | -0.8648*** | 0.4421*** | -0.9224*** |
| | (0.0103) | (0.0110) | (0.0074) | (0.0086) |
| pessref | 0.0910*** | -0.0917*** | 0.0127** | -0.0362*** |
| | (0.0093) | (0.0096) | (0.0051) | (0.0067) |
| urban | 0.1610*** | -0.1882*** | 0.1212*** | -0.1902*** |
| | (0.0141) | (0.0144) | (0.0086) | (0.0116) |
| _cons | -2.2750*** | 1.9080*** | -3.0348*** | 2.8570*** |
| | (0.0645) | (0.0648) | (0.0459 | (0.0531) |
| Wald | 24990.61 | 4994.05 | 76021.66 | 9279.85 |
| N | 192646 | 233272 | 343915 | 345985 |
| Rho | -.1259561 | .0631848 | .9120769 | -.0627006 |
| | (.0144699) | (.0287542) | (.0015331) | (.0182216) |

t statistics in parentheses * p<0.05, ** p<0.01, *** p<0.001

The LR chi2 tests in the adjusted models shows that all the models are significant to explain modifications occurred in the formal and informal labor market in the two periods under analysis. Observing the values of the z statistic for the four models of participation in the formal

and informal markets, all explaining variables, with exception to kids in the first period, are statistically different from zero at a 5% level of significance. Thus, we can say that each explaining variable statistically influences the decision of participating in the formal or informal labor market.

Tables 3 and 4 present estimation results, for formal and informal workers, of all four methods: (1) OLS; (2) pseudo-panel with fixed effects; (3) instrumental variables; and, (4) Heckman selection model. While table 3 brings results for the 1995-2002 period, table 4 shows results for the 2003-2008 period. A common way of testing for robustness of estimations is to do a sensitivity analysis, in which we include changes to some coefficients that could influence wage-education values. As can be assessed in tables 3 and 4, in spite of the many equations and methods used, most of the estimated coefficients were very close to one another, as well as presenting the same signs. A deeper analysis of tables 3 and 4 suggests that the estimators are consistent, in all four methods. They also present the expected sign, already consolidated in the literature. The coefficients of education and experience are always positive, while the square of experience is negative and its values reasonable, coherent with other values found in literature[10]. The estimated coefficients are all highly significant, at a 1% level of significance. In this manner, the explaining variables education, experience, square experience and education times experience are all consistent with human capital theory.

It is worth pointing out some aspects of the estimated equations. For example, the OLS estimated coefficient of years of study for formal labor shows its importance as a stimulating factor to wage. The increase in one year of study positively affects formal wages in a mean of 17,2%, for the first period and 19,8% in the second one. Returns on education for formal workers are greater than for informal laborers. In order to obtaining more precise information, we assess the second OLS equation for informality. In these results, the increase in one year of study corresponds to a 12,6% return over wages for the first period and 16,9% for the second,

[10] Loureiro and Galrão (2001); Sachsida, Loureiro and Mendonça (2004).

when estimated by the same technique. This is in accordance with the theory that if individuals invest on human capital, they obtain positive returns on their investment.

The education return in economic activity for formal and informal workers is characterized by great differentiation gap in the first 1995-2002 period analyzed. This means that allocation in the formal market is more attractive, for the existence of benefits such as welfare, health insurance and more stability, indicating that education is more valued in formality. However, this gap fell by almost half in the 2003-2008 period. Table 4 shows that returns to education maintained close to 21% for formality and 17,5% for informality in the second period of analysis. This means that, for each year of study, formal workers have at least a 19,7% up to 22,6%, when considering all the methods used, increase in wages. For informal workers this return is slightly lower, ranging from 16,9% to 18,2%, among the different methodologies.

The coefficient estimates for experience varied between 0,0694 a 0,0921 for the formal market and between 0,050 a 0,0885 for informality, among all four methodologies for either tables 3 and 4. Therefore, we can conclude that, due to little dispersion, the estimated data on experience returns are consistent. It is interesting to note that there are no significant gains for experience among working in formality or informality. Both value experience in very similar ways, in all four methodologies used.

The lambda variable, Mill's inverse ratio, was significant for both estimates. Thus its inclusion was necessary in order to eliminate tendency problems due to sample selection bias. In the first estimated equation for Heckman selection model, on formal workers for the 1995-2002, the lambda coefficient was negative and significant. This shows that non-measurable factors that would increase participation in the labor market, actually decreased wages for the first period. The other lambda coefficients are positive and statistically significant for informality in both cycles and for formality in the second period, indicating that these were correctly measured.

To be a married man in the labor market shows significant wage gains relative to non-married men. Married men in formal markets receive a mean of 27,4% more than single men, when observing the Heckman selection model in column 7, table 3. In the informal market, this gap reduces a little, representing 24,7% higher wages for married men. In the second period the wage gain seems to reduce a little, but is still high, representing a 17,5%, in formality, and 12,1%, in informality, wage surplus for marriages among men. It is interesting to note that values for all four methods generated similar coefficient values for each market sector: formal and informal.

Assessing the impact of firm size with up to 10 employees, available on table 3, we observe that wages in informal markets are greater than in formal markets. One of the main reasons for this differentiation is due to extreme rigidity and costs imposed by labor legislation. The concentration occurs in small firms where the presence of informal labor is more intense. On table 4, we observe this same behavior among firm size for the second period. This happens in the OLS estimation method (19,1% for formal and 26,1% for informal) for the Pseudo-panel with fixed effects (17,7% formal and 24,0% informal) and for the instrumental variables method (17,7% formal and 24,0% informal). However, for the Heckman selection model this tendency inverts. Formal firms tend to be more representative in formality than in informality for the 2003-2008 period, showing a 77% increase in formality, against a 30,8% in informality. In every model estimated we can verify a positive influence of being white over wages, that is, white people present higher wages than non-white individuals, both in formal and informal markets. In average, the wage gain for “white” is of almost a quarter more.

As can be observed on tables 3 and 4, the dummy variable for kids always has a negative influence over wages, as was expected. They are all statistically significant at a 1% significance level. This shows that having children has a negative impact on labor market and revenues. Children represent costs ranging from 12,9% to 15,3% for formal workers and 13,1% to 16,5% for informal laborers in the 1995-2002 period. For the 2003-2008 period this burden for having

kids decreased representing a 10,4% to 11% wage reduction for informal workers and 10,5% to 16,7% among formal workers.

Hence, the main difference between the 1995-2002 and 2003-2008 period is the impact of education over wages. Returns to education for the second cycle are greater than for the first period analyzed. The monetary gains for the first period vary from 16,5% to 18,2% for formality and from 12,2% to 14,3% for informal labor. Whereas, for the second period these gains range from 19,7% to 22,6% and from 16,9% to 18,2%, for formality and informality, respectively. According to these findings, investments made on education were more profitable for the 2003-2008 period. However, all wage gaps between formal and informal markets reduced considerably. This reinforces the fact that workers are increasingly deciding in which market to operate.

## 4. Conclusion

Empirical studies on informal employment in Brazil indicate that this market has gone through deep changes. Emerging evidence suggests that a share of informal employment is chosen and may offer specific benefits and opportunities to certain individuals. In particular, depending on their characteristics, some individuals may have a comparative advantage to work in the informal sector.

An empirical analysis was made to assess the main factors that drive the Brazilian labor when deciding to participate in the formal or informal sectors. In particular we've intended to observe if the game theory evolutionary model developed by Araujo and Souza (2010) is suited to describe the evolution of the informal market in the Brazilian economy. We used four different methodologies: the OLS, Pseudo-panel with fixed effects, Instrumental Variables and the Heckman Selection Model. The data for this study were obtained at the National Household Sampling Survey – PNAD/IBGE, from 1995 to 2008, with an exception to the years of 1994 and 2000, when this survey did not occur. The cross-section data were pooled and divided in two

periods: 1995 – 2002 and 2003 – 2008. The sample ranges individuals from 22 to 57 years of age.

Results indicate that the main difference between the 1995-2002 and 2003-2008 period is the impact of education over wages. Returns to education for the second period are greater than for the first period analyzed. The monetary gains for the first period vary from 16,5% to 18,2% for formality and from 12,2% to 14,3% for informal labor. Whereas, for the second period these gains range from 19,7% to 22,6% and from 16,9% to 18,2%, for formality and informality, respectively. According to these findings, investments made on education were more profitable for the 2003-2008 period. However, all wage gaps between formal and informal markets reduced considerably, which reinforces the fact that workers are increasingly deciding in which market to operate. In this vein we conclude that the model developed by Araujo and Souza (2010) provides a good description of the dynamics of the labor market when wages and tax schemes are taken into consideration as the key determinants of the choice of economic agents. However other variables such as education and skills have been pointed out as important factors that affect the decisions of workers and firms.

Table 3 Log-wage equations for the first cycle: 1995-2002

| VARIABLES | (OLS) 1-rwage FORMAL | (OLS) 2-rwage INFORMAL | (OLS_FE) 3-rwage FORMAL | (OLS_FE) 4-rwage INFORMAL | (XTIVREG2) 5-rwage FORMAL | (XTIVREG2) 6-rwage INFORMAL | (HECKMAN) 7-rwage FORMAL | (HECKMAN) 8-rwage INFORMAL |
|---|---|---|---|---|---|---|---|---|
| yostudy | 0.172*** | 0.126*** | 0.182*** | 0.134*** | 0.175*** | 0.143*** | 0.168*** | 0.122*** |
| | (73.92) | (32.58) | (70.30) | (30.43) | (6.01) | (13.11) | (73.58) | (30.01) |
| exper | 0.0755*** | 0.0514*** | 0.0921*** | 0.0636*** | 0.0694** | 0.0579*** | 0.0739*** | 0.0500*** |
| | (34.60) | (14.54) | (33.32) | (13.92) | (2.89) | (5.40) | (34.23) | (14.00) |
| exper2 | -0.000910*** | -0.000676*** | -0.00125*** | -0.000931*** | -0.000772* | -0.000693*** | -0.000885*** | -0.000656*** |
| | (-26.41) | (-11.89) | (-26.16) | (-11.59) | (-2.39) | (-3.92) | (-25.85) | (-11.51) |
| experyostudy | -0.00232*** | -0.00104*** | -0.00293*** | -0.00150*** | -0.00258* | -0.00211*** | -0.00226*** | -0.000983*** |
| | (-22.09) | (-6.14) | (-24.42) | (-7.63) | (-2.06) | (-4.13) | (-22.53) | (-5.96) |
| esize | 0.209*** | 0.256*** | 0.209*** | 0.255*** | 0.188*** | 0.209*** | 0.136*** | 0.208*** |
| | (31.59) | (24.12) | (31.58) | (24.06) | (11.92) | (7.35) | (12.58) | (8.67) |
| kids | -0.134*** | -0.134*** | -0.132*** | -0.133*** | -0.153*** | -0.165*** | -0.129*** | -0.131*** |
| | (-17.61) | (-10.65) | (-17.36) | (-10.59) | (-7.47) | (-4.71) | (-16.78) | (-10.26) |
| white | 0.267*** | 0.250*** | 0.267*** | 0.250*** | 0.181*** | 0.237*** | 0.259*** | 0.244*** |
| | (46.95) | (24.80) | (46.91) | (24.77) | (12.34) | (8.33) | (44.48) | (23.26) |
| marriedM | 0.274*** | 0.248*** | 0.274*** | 0.247*** | 0.306*** | 0.261*** | 0.274*** | 0.247*** |
| | (46.70) | (24.00) | (46.73) | (23.97) | (20.67) | (8.70) | (46.59) | (24.18) |
| group2 | | | 0.943*** | 1.361*** | | | | |
| | | | (24.89) | (21.89) | | | | |
| group3 | | | 0.921*** | 1.344*** | | | | |
| | | | (21.10) | (18.37) | | | | |
| group4 | | | 1.078*** | 1.471*** | | | | |
| | | | (24.06) | (19.46) | | | | |
| _cons | 1.109*** | 1.484*** | | | | | 1.264*** | 1.462*** |
| | (34.40) | (28.71) | | | | | (34.49) | (27.91) |
| lambda | | | | | | | -.0218648*** | .065274** |
| | | | | | | | (.0021591) | (.0297891) |
| N | 123393 | 44106 | 123393 | 44106 | 21589 | 5494 | 192646 | 233272 |
| F | 3876.83 | 1089.68 | 3114.80 | 874.21 | 481.79 | 112.31 | | |
| $R^2$-squared | 0.2009 | 0.1649 | 0.2015 | 0.1653 | 0.2076 | 0.1732 | | |
| Sargan-Hansen test | | | | | 0.8261(0.048) | 0.1612(1.963) | | |
| Wald chi2(8) | | | | | | | 24990.61 | 4994.05 |
| Log likelihood | | | | | | | -272898.2 | -156542.3 |
| LR test of indep. eqns. (rho = 0): chi2(1) = 62.53 Prob > chi2 = 0.0000 | | | | | | | | |
| LR test of indep. eqns. (rho = 0): chi2(1) = 5.06 Prob > chi2 = 0.0245 | | | | | | | | |

Table 4 Log-wage equations for the second cycle: 2003-2008

| VARIABLES | (OLS) 1-rwage FORMAL | (OLS) 2-rwage INFORMAL | (OLS_FE) 3-rwage FORMAL | (OLS_FE) 4-rwage INFORMAL | (XTIVREG2) 5-rwage FORMAL | (XTIVREG2) 6-rwage INFORMAL | (HECKMAN) 7-rwage FORMAL | (HECKMAN) 8-rwage INFORMAL |
|---|---|---|---|---|---|---|---|---|
| yostudy | 0.198*** | 0.169*** | 0.212*** | 0.182*** | 0.197*** | 0.171*** | 0.226*** | 0.173*** |
| | (116.16) | (59.40) | (112.22) | (55.83) | (85.87) | (43.78) | (119.69) | (55.83) |
| exper | 0.0763*** | 0.0698*** | 0.0991*** | 0.0885*** | 0.0780*** | 0.0739*** | 0.0910*** | 0.0718*** |
| | (49.36) | (26.64) | (49.99) | (25.90) | (37.64) | (20.47) | (53.02) | (26.74) |
| exper2 | -0.000693*** | -0.000780*** | -0.00109*** | -0.00110*** | -0.000718*** | -0.000848*** | -0.000932*** | -0.000807*** |
| | (-29.12) | (-18.90) | (-32.74) | (-18.80) | (-22.59) | (-14.94) | (-35.15) | (-19.19) |
| experyostudy | -0.00299*** | -0.00237*** | -0.00368*** | -0.00295*** | -0.00316*** | -0.00260*** | -0.00348*** | -0.00245*** |
| | (-41.65) | (-20.06) | (-44.56) | (-21.08) | (-32.90) | (-16.08) | (-43.92) | (-20.33) |
| esize | 0.191*** | 0.261*** | 0.190*** | 0.260*** | 0.177*** | 0.240*** | 0.770*** | 0.308*** |
| | (40.29) | (33.74) | (40.19) | (33.56) | (27.67) | (22.54) | (139.42) | (19.56) |
| kids | -0.112*** | -0.105*** | -0.109*** | -0.104*** | -0.114*** | -0.109*** | -0.167*** | -0.110*** |
| | (-20.91) | (-11.38) | (-20.32) | (-11.21) | (-16.01) | (-8.54) | (-28.05) | (-11.72) |
| white | 0.113*** | 0.115*** | 0.113*** | 0.115*** | 0.119*** | 0.116*** | 0.175*** | 0.121*** |
| | (28.22) | (15.19) | (28.15) | (15.15) | (22.08) | (11.13) | (38.62) | (15.57) |
| marriedM | 0.223*** | 0.178*** | 0.223*** | 0.177*** | 0.229*** | 0.189*** | 0.250*** | 0.181*** |
| | (54.31) | (23.64) | (54.30) | (23.60) | (41.80) | (18.28) | (54.27) | (23.87) |
| group1 | | | 1.454*** | 1.646*** | | | | |
| | | | (51.69) | (34.57) | | | | |
| group2 | | | 1.371*** | 1.567*** | | | | |
| | | | (42.27) | (27.82) | | | | |
| group3 | | | 1.515*** | 1.686*** | | | | |
| | | | (45.44) | (28.93) | | | | |
| _cons | 1.707*** | 1.864*** | | | | | 0.477*** | 1.865*** |
| | (70.94) | (47.22) | | | | | (17.71) | (47.22) |
| Lambda | | | | | | | -.123232*** | .065274** |
| | | | | | | | (.0142554) | (.0297891) |
| N | 254256 | 88355 | 254256 | 88355 | 134915 | 44297 | 343915 | 345985 |
| F | 3876.83 | 1089.68 | 3114.80 | 874.21 | 481.79 | 112.31 | | |
| $R^2$-squared | 0.1698 | 0.1561 | 0.9533 | 0.9335 | 0.2076 | 0.1732 | | |
| Sargan-Hansen test | | | | | 0.048(0.8261) | 1.963(0.1612) | | |
| Wald chi2(8) | | | | | | | 24990.61 | 4994.05 |

**References**


Arellano, M. and Bond, S. (1991). Some Tests of specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, Blackwell Publishing, 58(2), 277-97, April.

Araujo, R. and Souza, N. (2010). An Evolutionary Game Theory Approach to the Dynamics of the Labor Market: A Formal and Informal Pespective. *Structural Change and Economic Dynamics*, .

Baum, C., Schaffer, M. and Stillman S., (2003). Instrumental variables and GMM: Estimation and testing, *Stata Journal*, StataCorp LP, vol. 3(1), pages 1-31, March.

Deaton, A. (1985) Panel Data of Time Series of Cross-Section. *Journal of Econometrics*, 30, 109-126.

De Paula, and Scheinkman, J. (2007). The Informal Sector. *PIER Working Paper* 07-035.

Fiess, N. Fugazza, M and Maloney, W. (2010). Informal self-employment and macroeconomic fluctuations. *Journal of Development Economics* 91, 211–226

Fagiolo, G., Dosi, G. and Gabriele, R. (2004). Matching, Bargaining, and Wage Setting in an Evolutionary Model of Labor Market and Output Dynamics. *Advances in Complex Systems*, Vol. 7, No. 2 (2004) 157–186. World Scientific Publishing Company.

Fortin, B., Marceau, N. and Savard, L. (1997). Taxation, wage controls and the informal sector. *Journal of Public Economics* 66, 293 – 312.

Fortin, B. Marceau, N. and Savard, L. (1997). Taxation, wage controls and the informal sector. *Journal of Public Economics*, 66: 293-312.

Gintis, H. (2000). Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction. *Princenton University Press*. Princeton. New Jersey.

Gobetti, S. and Orair, R. (2010). Estimativa da Carga Tributária de 2002 A 2009. Diretoria de Estudos e Políticas Macroeconômicas. Notas Técnicas. Instituto de Pesquisa Econômica Aplicada (IPEA)N°16. Brasília, março de 2010

Hirschman, A. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms Organizations, and States*. Cambridge, MA: Harvard University Press.

Hausman, J. and Taylor, W. (1981). Panel data and unbservable individual effects. Econometrica, 49(6):1377-1398.

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply, *Econometrica*, 1, 679-694. 1974.

Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 47(1):153-161.

Heckman, J. (1980). Varieties of Selection Bias. *American Economic Review*, 80, 313-318. 1980.

Hofbauer, J. and Sigmund, K. (2003). *Evolutionary Games and Population Dynamics. Cambridge University Press*. United Kingdom.

Loayza, N. (1996). The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy* 45, 129-62.

Loayza, N., Oviedo, A. and Serven, L. (2005a). Regulation and Firm Dynamics. mimeo, *The World Bank*.

Loayza, N. and Oviedo, A. and Serven, L. (2005b). The Impact of Regulation on Growth and Informality: Cross Country Evidence. *World Bank Policy Working Paper* 3623.

Loyaza, N. and Rigolini, J. (2006). Informality trends and cycles. *World Bank Policy Research Working Paper* 4078.

Loureiro, P. and Galrão, F.(2001). Discriminação no mercado de trabalho: Uma análise dos setores rural e urbana no Brasil. *Economia aplicada*, v. 5, n. 3, p. 519-545.

Maloney, W. (2004). Informality Revisited. *World Development*, 32, 1159 – 1178.

Mattos, E. and Ogura, L. (2007). Skill differentiation between formal and informal employment. *Journal of Economic Studies* 36(5), 461 – 478.

Mulinge M.; Mueller C.W. (1998) Employee Job Satisfaction in Developing Countries: The Case of Kenya. *World Development* 26(12), 2189-2199.

Pissarides, C. (2000). Equilibrium Unemployment Theory. *The MIT Press*. Cambridge, Massachusetts.

Quandt, R. and Rosen, H., (1986). Unemployment, Disequilibrium and the Short Run Phillips Curve: An Econometric Approach. *Journal of Applied Econometric* 1(3), 235 – 53.

Saavedra, J. and Chong, A. (1999). Structural Reforms, Institutions and the Informal Sector in Peru. *Journal of Development Studies* 35 (4), 95-116.

Sachsida, A., Loureiro, P., and Mendonça, M. (2004) Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia,* v. 58, p. 249-265.