Skill segmentation of informal employment

May 2007

Abstract
This paper examines the skill segmentation of informal and formal labor markets. A theoretical model is developed assuming that concealment of production is increasingly costly in the informal sector. The results match the existent empirical evidence, which indicates that informal firms employ a larger proportion of low skilled workers than formal firms. Using data on the Brazilian informal economy (PNAD-2005), an econometric analysis is presented to compare the plausibility of our model with the one proposed by Fortin et al. (1997). The estimations indicate that our model fits the data better.

Resumo
O artigo examina a segmentação entre o setor formal e informal dos mercados de trabalho. Um modelo teórico é desenvolvido assumindo que o o custo de permancer no setor informal é crescente na produo. Os resultados do modelo estão em acordo com a literatura empírica e indicam que firmas no setor informal empregam uma grande proporção de trabalhadores desqualificados quando comparados com empresas no setor formal. Usando dados da PNAD-2005 para a economia informal brasileira, os resultados apontam que nosso modelo teórico explica melhor a realidade brasileira quando comparados ao modelo apresentado em Fortin et al (1997).

JEL classification: H26, J24, O17.

Keywords: informal sector, segmentation, labor skill, less developed countries.

1 Introduction

The informal sector represents a large proportion of the production activity in less developed countries. Most societies attempt to control these activities through punitive or educational measures rather than through reforms that could improve the development of the formal economy. Shneider and Enste (2000) pose some causes for the growth of the informal sector such as the rise of the burden of taxes and social security contributions, increased regulation in the formal economy (e.g., reduction of weekly working hours), earlier retirement, unemployment, and the decline of civic virtue and loyalty toward public institutions combined with a declining tax morale.

Formal and informal jobs differ in several dimensions. While formal workers usually have access to some benefits, individuals employed in the informal sector are excluded from them. According to the model proposed by Rosen (1986), average earnings should be higher in the less desirable informal sector to compensate for the non-pecuniary benefits granted to registered workers. That is, assuming that the value of the benefits is non-negative, direct payments in the informal sector should be at least as high as those in the formal economy to offset the lack of these benefits. Botelho and Ponczek (2007) conjecture that the cumbersome Brazilian labor laws are the main reason for the rigidity in the labor market. Also, Maloney (2004) considers informal work as a transitory phase for entrants in the labor market. According to this view, informal jobs may be a desirable alternative, providing more flexibility, and allowing both sides of the market to avoid cumbersome and expensive regulations, and evade taxes. On the other hand, the recent literature on the emergence of informal activities relies on the heterogeneity of firms or entrepreneurs as the determinant factor (see Fortin, Marceau, and Savard (1997), thereafter cited as "FMS", Rauch (1991), Dessy and Pallage (2003), and Antunes and Cavalcanti (2003)). In this literature, the informal sector endogenously emerges due to the possibility of evasion, which reduces the marginal cost of being informal. It is shown that there is a segmentation in the scale of the firms and in wages, i.e., smaller firms with lower wages operate in the informal sector.

Moreover, empirical studies indicate that there are remarkable differences between the formal and the informal labor markets. For example, Funkhouser (1996) shows that the return to education in the informal sector is much lower than in the formal sector and the male-female differential is much larger in the informal sector in five Central America countries. Tannuri-Pianto and Pianto (2002) find that illiterate individuals are highly penalized in the formal sector in Brazil. College education is more advantageous in the formal sector, while the marginal benefit of elementary education is higher in the informal sector. The authors also argue that there is a wage differential between formal and informal workers even after controlling for their characteristics, i.e., with workers in the formal sector earning a higher return to education than informal workers. Indeed, several other papers discuss the existence of this wage differential.[1] The literature points out that the difference in

[1]See for instance Menezes-Filho et al. (2004), Carneiro and Henley (2001), and Funkhouser

earnings between sectors are explained by non-observable characteristics of workers. In addition, the wage differential may be compensated by the inefficacy of labor codes and/or low levels of human capital of workers in the informal sector.

In contrast with the previous literature, this paper attempts to analyze the emergence of the informal sector focusing on the heterogeneity of workers rather than firms. Thus, labor market considerations are incorporated into the analysis, which relies on a simplified general equilibrium model that takes into account the possibility of an informal sector, where firms do not pay taxes. Both formal and informal firms, however, produce the same good. Regarding workers, individuals may differ in their skill (productivity) level, with each offering one unit of labor. As the key assumption, an informal firm faces an increasingly cost of concealing production due to the need to hide production, in addition to the expected losses due to fines, confiscation of products, halted production, etc. Then, the emergence of informal activities happens due to evasion benefits for firms, but differences in private skill characteristics of workers and firms restricts the advantages of informal operation for certain firms. Accordingly, an industry becomes segmented in terms of evasion, but the segmentation depends on the type of workers required by the industry. This distinguishes our model from FMS, where workers are homogeneous in all characteristics. On the other hand, FMS also consider the effects of an imposed minimum wage rate in the formal sector, which is higher than the equilibrium wage paid in the informal sector, while we obtain our results without wage differentiation between the formal and the informal sectors (i.e., all workers obtain the same wage). It is also interesting to notice that the results depend on the concealment cost function. In particular, this function must be convex (the cost must get bigger at an increasing rate) with respect to the firm's production level. Facing an increasing cost of concealment, it is only optimal for firms to operate in the informal sector when the production level is low.[2] Finally, competition guarantees that only profitable firms stay in the market in the long run. Consequently, FMS's concealment cost function, which depends on the number of employed workers, generates results that are not well matched to the empirical evidence on labor skill segmentation mentioned before. This happens because labor demand may not increase fast enough with labor

(1996).

[2]Empirical evidence of this segmentation exists. See Shneider and Enste (2000).

productivity since wages also go up with productivity. Thus, high skilled labor could be driven to the informal sector.

This paper also presents econometric exercises using data on the Brazilian informal economy. This empirical analysis allow us to compare the plausibility of the model in this paper relative to FMS's model. Indeed, the results suggest that our model fits the data better. First, the probability that a higher skilled worker (when education level is considered) is employed in the formal sector is greater, corroborating the results of previous studies. Second, 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 (refuting the FMS's assumption of the existence of a wage differential between sectors). Finally, firms with the same number of workers are observed in both sectors, which is not consistent with FMS's model, where labor demand increases with firm productivity. In contrast, our model allows the number of workers to be the same in both sector because labor demand also depends on the skill level of workers employed by the firm.[3]

The paper is organized as follows. The next Section presents the model and analyzes the equilibrium distribution of workers between sectors for each level of labor skill. Section 3 presents an empirical test to compare the plausibility of FMS's model with ours. Section 4 offers concluding remarks.

2 The model: the case of costly production concealment

2.1 Setup

In this section, the model presented by FMS is modified to account for the fact that each worker has a skill qualification indexed by a scalar $\phi \geq 0$. As a simplification, however, each firm is skill specific, meaning that there is no substitution between workers of different labor skill levels. While the model assumes that all firms produce the same good with price normalized to \$1, firms that use workers with higher skill obtain a greater production.[4] A firm that requires high skilled labor will be called a "high-ϕ" firm, while a firm that uses low skilled labor will be

[3]This feature is explicit in the modified version of the model presented in a previous version of the paper.

[4]An alternative interpretation for this assumption is that firms using higher skilled workers produce a greater valued product. In this case, however, demand for each product would have to be perfectly elastic, so that relative prices are not affected by changes in the supply of products.

called "low-ϕ" firm. The set of firms that use the same type of labor skill will be called an "industry" (thus, each industry is characterized by a different labor skill requirement).

According to their skill level, workers get a wage rate $w(\phi)$. The determination of this wage function will not be studied here. However, one can interpret that competition in the labor market makes the wage rate to equal the opportunity cost incurred by a worker when supplying labor of quality ϕ. This cost includes learning effort (to obtain qualification) and the opportunity of getting a lower skilled job. In fact, we assume that wages increase with qualification, i.e., $w'(\phi) > 0$.

2.2 Sector choice by firms

In this paper, the term "formal sector" is used to refer to the portion of the economy characterized by firms binding to taxes and regulations vis-a-vis the "informal sector" where firms do not fully pay taxes and/or follow regulations.

In this subsection, a model is developed to show that highly efficient firms are more profitable in the formal sector, while less efficient firms find informal operation more advantageous. This result is analogous to the one presented by FMS. However, our model uses two distinct assumptions. First, constant returns to scale in firm-level production is adopted to simplify the analysis (in fact, additional results arises, making the model encompassing). Second, firms in the informal sector face an increasingly costly concealment of production (instead, FMS assumed that concealment costs increase with the number of workers). This is a key assumption to demonstrate (in the next subsection) that the proportion of employment by formal firms is greater in industries that use higher labor skill. In contrast, if concealment costs depend simply on the number of workers, then firms in high skill industries may prefer to operate informally (this result is shown in a previous version of the paper is available upon request).

Considering that firms are skill specific, the analysis in this subsection will be restricted to a particular skill level ϕ. The effects of changes in ϕ are analyzed in the next subsection.

Note that the aggregate production of each industry exhibits decreasing returns to scale due to the fact that firms in the industry are heterogeneous in their efficiency level. In other words, some firms are better technologically than others, producing

more with the same amount of labor. This heterogeneity may be due to differences in management skill, knowledge of the market, or even historical factors. The efficiency level of a firm is denoted by θ, which has an absolute frequency distribution $g(\theta, \phi)$ on $[0, \bar{\theta}]$. In order to simplify the analysis, it is assumed that each firm can hire only one worker.[5]

When using labor skill ϕ, firm production is $y^f = \theta\phi$ if in the formal sector. Accordingly, profit of a firm operating formally is

$$\pi^f = (1-\tau)[\theta\phi - (1+t)w(\phi)], \tag{1}$$

where τ is the profit tax rate and t is the labor tax rate. Note that (1) implies that a firm in the formal sector only produces if $\theta\phi \geq (1+t)w(\phi)$, i.e., if profitable. By contrast, a firm in the informal sector produces $y^i = a\theta\phi$, with $a < 1$ being parameter that represents the inefficiency inherent to informal operation (due to, for example, difficulties in obtaining capital, technology, other inputs, regular buyers, or financial credit). Note that a allows us to compare the effective revenue by a firm in each sector (the term "effective" will be used in this paper to refer to how much firms actually obtain or pay after all taxes are imposed): a firm in the formal sector ends up with $(1-\tau)\theta\phi$, while an informal firm gets $a\theta\phi$. Besides the inherent inefficiency, informal firms also incur a concealment cost $c(y^i)$, which is required to prevent punishment from authorities. This cost exists only when the firm produces a positive amount, increasing with production at a growing rate, i.e., $c(0) = 0$, $c'(.) > 0$, and $c''(.) > 0$. Hence, the profit of a firm in the informal sector is

$$\pi^i = a\theta\phi - w(\phi) - c(a\theta\phi). \tag{2}$$

Thus, a firm in this sector only produces if $a\theta\phi \geq w(\phi) + c(a\theta\phi)$.

Considering equations (1) and (2), a profit maximizing firm with efficiency level θ operates in the formal sector if

$$(1-\tau)\theta\phi - (1-\tau)(1+t)w(\phi) \geq a\theta\phi - w(\phi) - c(a\theta\phi). \tag{3}$$

Note that profit increases with θ at a constant rate $(1-\tau)\phi$ in the formal sector,

[5]Instead of one worker per firm, the same setup can be used if each industry has a representative producer with decreasing returns to scale, who would allocate labor to the sector that is marginally more profitable. On the empirical side, recent evidence on the Brazilian economy (from the PNAD 2005) indicates that about 88% of informal production is due to "firms" employing one worker only.

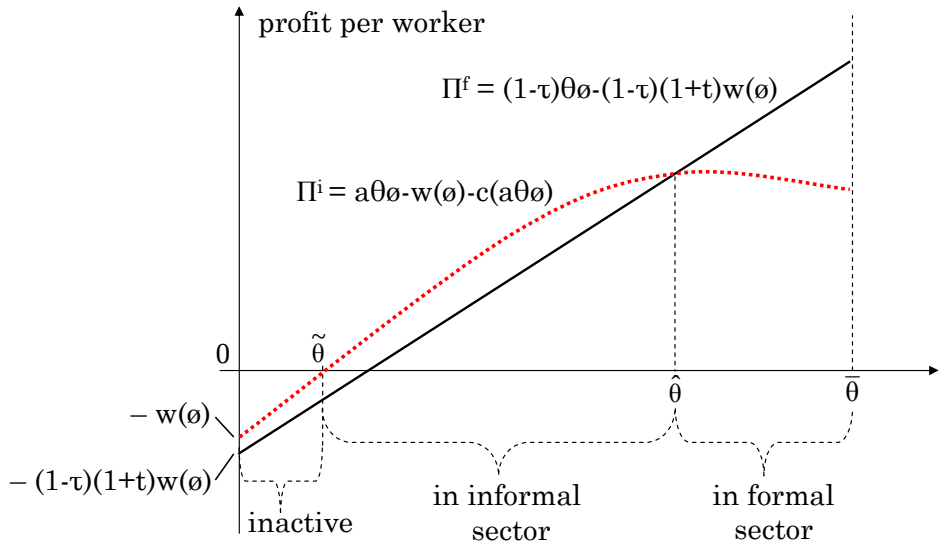

Figure 1: Potential firm profit in each sector

while in the informal sector this rate is $[1-c'(a\theta\phi)]a\phi$. Because $c'(.)$ is an increasing function, the growth rate in the informal sector becomes eventually smaller than in the formal sector. Thus, profit in the formal sector eventually exceeds the potential profit in the informal sector (unless $\bar{\theta}$ is not large enough, in which case all firms would be in the informal sector). Leaving aside the cases where all firms operate formally or informally for the moment, the result is the one depicted in Figure 1. In the diagram, $\hat{\theta}$ represents the threshold level of firm-efficiency such that only firms above $\hat{\theta}$ operate in the formal sector. In addition, note that the least efficient firms (i.e., firms with $\theta < \tilde{\theta}$ in the Figure) are not profitable, remaining inactive.

It is important to mention that the previous analysis supposes that low-θ firms find the informal sector more profitable. It is possible that firms would never operate in the informal sector: if the inefficiency inherent to informal operation is greater than the tax burden in the formal sector, i.e., $a < (1-\tau)$, and the effective labor cost is greater in the informal sector, i.e., $1 \geq (1-\tau)(1+t)$, then informal operation yields no advantage. These are sufficient, but not necessary conditions to have all firms operating formally. For example, if $c(.)$ grows very fast with production, then only very low-θ firms may find the informal sector advantageous, but these firms may be unprofitable. The diagram in Figure 1 also implies that other possible equilibrium cases exist. First, if the concealment cost does not grow relatively fast, $\hat{\theta}$ would be greater than $\bar{\theta}$, implying that all existent firms would be in the informal sector.

Second, on contrary, if the concealment cost grows too quickly, then all firms would be in the formal sector. Third, if the effective labor cost is higher in the informal sector, i.e., if $1 \geq (1-\tau)(1+t)$, then it is possible that the π^i curve crosses the π^f curve twice, first from below and then from above, implying that low and high-θ firms are in the formal sector, but medium θ-firms operate in the informal sector.[6] Since additional empirical evidence is needed to support the plausibility of these equilibrium cases, the paper will continue focusing on the case where low-θ firms operate informally and high-θ firms operate formally.

The following proposition summarizes the results derived in this subsection.

Proposition 1 *Consider the following conditions:*

(i) $a \geq (1-\tau)$,

(ii) $1 \leq (1-\tau)(1+t)$, *and*

(iii) $c(a\theta\phi)$ *grows relatively fast with production.*

While conditions (i) and (ii) imply that informal operation is advantageous for at least the lowest-θ firms, condition (iii) guarantees that at least the highest-θ firms will be formal. Therefore, if all conditions are satisfied, high-θ firms operate in the formal sector, while informal operation is more advantageous to low-θ firms (although these low-θ firms may be unprofitable).

In terms of government policies, note that the proportion of firms in the formal sector increases with the degree that condition (iii) in Proposition 2 is satisfied and the degree that conditions (i) and (ii) are not satisfied, holding everything else constant. Therefore, informality is reduced with greater and more increasing $c(a\theta\phi)$ on one hand, and smaller a, τ, or t on the other hand. In words, high tax rates and low monitoring and punishment (which implies low a and high $c(a\theta\phi)$) stimulates informality.[7]

2.3 Labor skill segmentation

This subsection analyzes how firms' choices between the formal and the informal sector determine the distribution of workers between these sectors. In addition,

[6] In this case, it is also possible that, if the concealment cost grows slowly, only low-θ firms would be in the formal sector.

[7] The effect of a change in τ may seem ambiguous when looking at conditions (i) and (ii) in Proposition 2, but recall that τ is the profit tax rate, which reduces formal profits when positive. Thus, a greater τ constrains formality.

changes in this distribution according to labor skill ϕ are studied.

In order to analyze labor demand by firms in the formal sector, recall that the frequency distribution of θ is given by the function $g(\theta, \phi)$ and each firm demands only one worker. Hence, in a given industry of labor-type ϕ, firms with efficiency level $\theta \geq \tilde{\theta}$ demand a total of $g(\theta, \phi)$ workers (firms with $\theta \leq \tilde{\theta}$ are unprofitable, not demanding any worker). Since all firms with $\theta \geq \hat{\theta}$ operate in the formal sector, the number of ϕ-type workers employed in the formal sector (denoted L_ϕ^f) is given by

$$L_\phi^f = \int_{\hat{\theta}}^{\bar{\theta}} g(\theta, \phi) d\theta. \tag{4}$$

Hence, the proportion of workers in the formal sector is given by $\alpha \equiv \frac{L_\phi^f}{L_\phi}$, where $L_\phi = \int_{\tilde{\theta}}^{\bar{\theta}} g(\theta, \phi) d\theta$. Because this proportion depends on the frequency distribution $g(\theta, \phi)$, it could increase or decrease with ϕ.[8]

In order to analyze how α is affected by an increase in ϕ, first notice that $\hat{\theta}$ decreases with ϕ if $\frac{d\pi^f}{d\phi} > \frac{d\pi^i}{d\phi}$ at $\hat{\theta}$. These derivatives can be obtained from (1) and (2):

$$\frac{d\pi^f}{d\phi} = (1-\tau)\hat{\theta} - (1-\tau)(1+t)w'(\phi) \tag{5}$$

and

$$\frac{d\pi^i}{d\phi} = [1 - c'(a\hat{\theta}\phi)]a\hat{\theta} - w'(\phi). \tag{6}$$

At $\hat{\theta}$, the slope of the π^i curve (as depicted in Figure 1) is smaller than the slope of the π^f curve, which implies that $[1 - c'(a\hat{\theta}\phi)]a < (1-\tau)$. Thus, because labor productivity significantly increases concealment cost at $\hat{\theta}$, revenues after concealment costs increase less in the informal sector than in the formal sector. On the other hand, the effective wage cost tend to increase proportionately more in the formal sector (supposing that $1 \leq (1-\tau)(1+t)$). Hence, potential profit increases relatively more in the formal sector if the concealment cost takes away a large part of the labor productivity gain, i.e., if $c'(a\hat{\theta}\phi)$ is sufficiently large relative to the tax rate τ, thus compensating the increased wage cost advantage obtainable in the informal sector, or if the wage cost advantage is very small (for example, if there was no difference in effective wage rates between sectors, then clearly profit in the formal sector increases

[8] Even though the paper does not analyze the equilibrium in the labor market, it is assumed that the wage rate $w(\phi)$ makes supply of each type of labor to exactly equal demand L_ϕ. Equilibrium should exist because even if supply was perfectly inelastic, L_ϕ decreases $w(\phi)$.

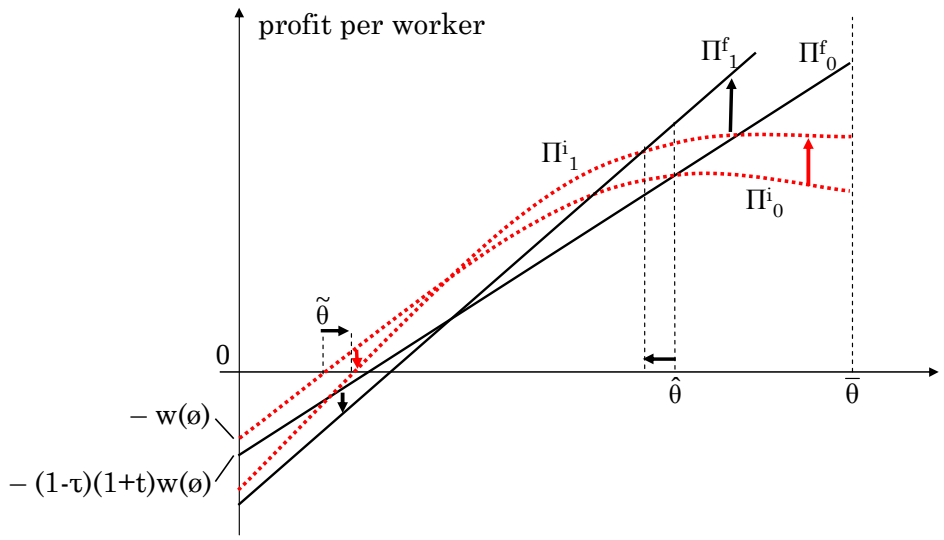

Figure 2: Changes in firm profit as ϕ increases

more).[9] In this case, the changes in the profit curves and the impact on $\hat{\theta}$ can be depicted as in Figure 2. The decrease of $\hat{\theta}$ results from the increase in firm production due to higher labor skill, which makes the concealment cost of informal operation larger, thus inducing more firms to become formal. Hence, a greater proportion of high-ϕ firms are induced to operate in the formal sector compared to low-ϕ firms.

Considering the case where $g(\theta, \phi)$ is constant with respect to ϕ, if $\hat{\theta}$ decreases with ϕ, then labor demand L_ϕ^f increases with ϕ (because $\hat{\theta}$ is the inferior limit of the integral in equation (4)). In addition, note that $\tilde{\theta}$ most probably increases with ϕ because for low-θ firms the marginal productivity increase due to higher ϕ is very small compared to the increased $w(\phi)$, reducing low-θ firms's profits. [10] Hence, α increases with ϕ. In other words, the proportion of formal employment is larger in high skill industries (i.e., among high skill workers). This result matches the empirical evidence mentioned in the Introduction.

[9]Notice that the magnitude of the decrease in $\hat{\theta}$ depends on the size of $c'(a\hat{\theta}\phi)$. Because $c''(a\hat{\theta}\phi) > 0$ and supposing that $\hat{\theta}$ does not decrease proportionately more than the increase in ϕ (which implies that $hat\theta\phi$ still increases), the marginal concealment cost $c'(a\hat{\theta}\phi)$ increases with ϕ, making it more likely that the change in profit is greater in the formal sector.

[10]A decrease in $\tilde{\theta}$ would happen if the informal profit for low-θ firms increases with ϕ. However, this only occurs if the wage rate $w(\phi)$ is not very sensitive to changes in ϕ. For example, if wages were the same regardless of skill, then higher labor skill increases profits of all firms, making some of the unprofitable firms to become profitable. Because obtaining higher skill requires costly learning by workers, it is not likely that the wage rate is invariant with skill.

Nonetheless, a further analysis is required because the frequency distribution $g(\theta, \phi)$ probably varies with ϕ in reality. First, note that firm productivity θ is determined in practice by capital use and technology (management or production technology). Since it is plausible that there is complementarity between labor skill and the amount of capital or technology used, then ϕ and θ are complements. Consequently, high θ firms are more frequent in high-ϕ industries than in low-ϕ industries, i.e., $g(\theta, \phi)$ becomes more skewed toward high-θs as ϕ increases (with $\bar{\theta}$ being probably greater in high-ϕ industries). According to equation (4), this change in distribution implies that the proportion of formal employment tends to be higher in high-ϕ industries even if $\hat{\theta}$ or $\tilde{\theta}$ increased slightly compared to lower-ϕ industries.

Proposition 2 summarizes the results in this subsection.

Proposition 2 *The proportion of formal employment is greater for high skilled workers than for low skilled workers if $g(\theta, \phi)$ is constant with respect to ϕ or becomes more skewed toward higher θ values as ϕ increases, and $(1-\tau)\hat{\theta} - [1 - c'(a\hat{\theta}\phi)]a\hat{\theta} > [(1-\tau)(1+t)-1]w'(\phi)$, which requires that either $c'(a\hat{\theta}\phi)$ is large relative to the profit tax τ or the marginal increase in the effective wage differential $[1-(1-\tau)(1+t)]w'(\phi)$ is relatively small.*

As a final remark, recall that Proposition 1 implied the theoretical possibility that all firms operate in the informal sector. Proposition 2 indicates that this possibility is greater for industries that require low labor skill because the threshold $\hat{\theta}$ is larger in this case, being possibly greater than the maximum existent $\bar{\theta}$. Intuitively, holding everything else constant, labor skill increases the amount produced, making it harder to conceal production. Hence, only industries that use low productive labor can be completely informal.

3 Empirical Analysis

Using data on the Brazilian informal sector, this section presents an econometric test of the plausibility of the theoretical model in this paper compared to FMS's model. One of the differences between the models is that FMS assumes $\overline{w} > w_i$, where $\overline{w}$ corresponds to the minimum wage established by law and w_i is the non-regulated wage in the informal sector. In contrast, in the present paper, the wage

rate is the same for workers with the same productivity regardless of the sector where the firm operates. In addition, FMSs results does not allow for firms with the same size in both sectors, while our model has no such restriction. The data is also used to reassess the hypothesis that formal employment increases with labor productivity.

3.1 Brazilian data on the informal sector

According to ENCIF 2003 (Informal Urban Economy Survey) elaborated by the IBGE (Brazilian Institute of Geography and Statistics), 88% of the informal firms in Brazil are classified as self-employed. This survey estimates that there are more than ten million informal firms in Brazil, mostly alocated in local commerce and small services. This sector employs 69% of all informal workers. Hence, it is essential to take the self-employed category into consideration when analyzing the Brazilian informal economy. Because there are also a large amount of formal self-employed workers, there is a coexistence of firms of the same size in both sectors. This empirical fact poses doubt on FMS's result that small firms operate in the informal sector, while formal sector has larger firms.

Data from the PNAD (Brazilian National Household Survey, which uses a representative population sample) is taken to test whether there is a difference between wages obtained by formal and informal self-employed workers in 2005. Self-employed workers represent firms that have only one employee.[11] In particular, we control for heterogeneous individual characteristics, focusing on the firms decision to operate either in the formal or in the informal sector. Table 1 summarizes the data used.

These variables attempt to capture individual as well as geographic characteristics. Concerning the first aspect, the variable experience represents how long the individual works as a self-employed, while age computes how old he/she is. Workers educational level are measured by dummy variables that assume value one if the individuals has concluded: elementary (4 years of schooling), secondary1(8 years of schooling), secondary2 (11 years of schooling), fundamental(8 years of schooling), highschool (11 years of schooling) graduate (15 years of schooling) and also by his/her years of schooling and the sum of family's years of schooling (family

[11]The terms "firm" and "self-employed worker" are used indistinctly. The advantage of restricting the sample to self-employed workers is that it is only necessary to control the estimation for individuals' characteristics, since they represent the firms.

Table 1: Descriptive Statistics

	Mean	Std. Deviation	Minimum	Maximum
experience	7.713	9.929	0	80
age	36.328	14.175	10	102
bornmunic	0.511	0.500	0	1
years of schooling	8.443	4.496	1	16
activity	5.672	3.579	1	13
formal	0.469	0.499	0	1
family school years	3.666	1.458	1	6
ln(w)	2.733	2.386	-3.871	27.631
male	0.559	0.497	0	1
white	0.468	0.499	0	1
elementary	0.107	0.310	0	1
secondary1	0.019	0.138	0	1
secondary2	0.014	0.118	0	1
fundamental	0.297	0.457	0	1
highschool	0.244	0.430	0	1
graduate	0.097	0.296	0	1
otherincome	0.019	0.136	0	1
children	1.713	1.279	0	13
urbanarea	0.807	0.395	0	1

school years)[12]. The variables white and male are dummies that assume one if the individual is white and male respectively. Last, the variables activity and formal, denote whether the individual works in the local commerce and in the formal sector respectively, while the variables otherincome and children represent how much the family's other member earn and how many children live with the worker. To capture geographic characteristics, the variables bornnmunic and urbanarea are dummy variables that assume value equal to one if the individual was born in that municipality (to caputre his/her network) and if that municipality belongs to an urban area.

3.2 Estimates

The empirical literature on wage differentials is extensive. [13] The main problem of such approach found to estimate the wage differential is to determine whether the decision of being formal or not is exogenous. In other words, one can say that a firm that has lower profits, may decide to not pay taxes and then become informal, or

[12]In Brazil, there were several changes in the educational levelss names. Depending on how old is the individual, he/she can fill out his/her educational achievements accordingly.

[13]See Leontaridi (1998).

vice-versa, or , the fact that a worker is in the informal sector might depend on her individual characteristics not observable by the researcher, i.e, the estimator can be biased due to endogeneity problem.[14]

To address the endogeneity issue describe above, we use the propensity score methodology. First, a probit estimation is run on the workers' individual characteristics to compute the probability of being formal versus informal. Then wages of similar self-employed workers are compared by choosing observations in each group (formal and informal) with similar propensity scores, i.e., only the observations with similar predicted probit value.[15]. The propensity score (the conditional probability of being formal) is a matching measure much simpler than conditioning on a large dimensional vector of covariates, i.e., one has to consider only the predicted probit value computed in the first stage (one dimension) rather than the full set of dependents (k-dimension) to compare the workers in the formal and informal sectors.[16]

To implement this methodology, denote the comparison group for the formal self-employed worker k with characteristics X_k as the set $H_j(X) = \{j/X_j \in c(X_k)\}$, where $c(X_k)$ is the characteristics neighborhood of x_k, i.e. those workers in the formal sector that present their characteristics close (in terms of Euclidian distance) to the individual k. Let N_infC denote the number of comparison group of informal self-employed workers (worker whose characteristics are close to the individual k) and let $h(k, j)$ represent the weight given to the j^{th} informal worker associated with the k^{th} formal worker, with $\sum_j h(k, j) = 1$. Formally, the general formula for the matching estimator is

$$\Delta w = \frac{1}{N_f} \sum_{k \in \{formal\}} [w_{f,k} - \sum_j h(k, j) w_{inf,j}],^{17} \tag{7}$$

where N_f is the number of formal workers, $\{formal\}$ is the set of all formal self-employed workers, $w_{f,k}$ is the wage of a formal worker k and $w_{inf,j}$ is the wage of the j^{th} informal worker that belongs to the comparison group of the formal worker k.

[14] See Botelho and Ponczek (2007).

[15] For a more detailed explanation, see Cameron and Trivedi (2005) and Rosenbaum and Rubin (1983).

[16] See Deheijia Wahba (2002).

[17] This is a general formula for the ATET - average treatment effect on the treated - estimator. (See Cameron and Trivedi (2005), p. 863.)

Note that $0 < h(k,j) \leq 1$. Different matching estimators are generated by varying the choice of $h(k,j)$. We consider the nearest-neighbor matching, choosing for every formal worker k the set $H_j(p(X)) = \{j/min\lambda_l p(X_k) - p(X_j)\lambda_l\}$, where $p(X)$ denotes the propensity score and $\lambda_l p(X_k) - p(X_j)\lambda_l$ is the Euclidian distance between the propensity scores of the formal k^{th} self-employed worker and j^{th} informal one. With this measure, the "closest" informal individual is considered when comparing to the formal self-employed worker. Table 2 presents the results of the first stage probit estimation,

Table 2: Formal sector entry probability (probit)

Explanatory Variable	Coefficient	(T-statistic)
bornmunic	-0.013	(0.049)
experience	0.021***	(0.007)
experiencesquared	-0.0004**	(0.000)
agesquared	-0.0006***	(0.000)
otherincome	-0.199***	(0.065)
elementaryschool	0.132	(0.082)
secondaryschool1	-0.016	(0.112)
secondaryschool2	-0.014	(0.134)
highschool	-0.073	(0.077)
graduate	0.487***	(0.114)
other	(0.223)	(0.202)
children	-0.065***	(0.023)
age	0.078***	(0.016)
white	0.376***	(0.048)
urbanarea	0.071	(0.102)
schoolyearsoffamily	0.230***	(0.031)
constant	-4.379***	(0.372)
LL	-1852.17	
LR $\chi 2(16)$	595.63	
P value	0	
Pseudo R^2	0.1385	
N	5502	

Note that the variables *experience* and *age* have the expected sign and significance. They indicate a concave function for the probability of entering the formal sector (since the signs of *agesquared* and *experiencesquared* are negative). Thus, higher age and experience increase the probability of having a formal firm, but at some point that probability starts to decrease. Also as expected, the education variables *graduate* (representing school graduation) and *schoolyearsoffamily* (sum

of the years of schooling of other family members) increase the probability of being formal. This reinforces our theoretical model which suggest that higher skilled workers tend to be employed in the formal sector. However, other sources of income (*otherincome*) reduces the formality probability. The same happens when the number of children (variable *children*) increases. This suggests that these children might contribute to the family's income, which reinforces the effect of *otherincome*.[18]

As mentioned before, the first stage aims to find what induces a self-employed to enter the formal sector. Each individual is now attributed an estimated propensity score. The next step is to compare the wages of formal and informal firms with similar propensity scores. Table 3 summarizes the results.

Table 3: Dependent variable - ln (wages)

Sample	Formal	Controls	Difference	St. Err.	t-stat.
Unmatched	3.268	2.352	0.916	0.093	9.830
ATT	3.268	3.057	0.211	0.214*	0.984
N	728	728			

* Calculated using bootstrap technique.

The first row of the table (Unmatched) presents the results of the comparison when no first stage is applied. In this case, observe that the difference in wages is positive (0.916) and statistically significant ($t - stat = 9.83$). However, after the application of the correction method (ATT), see that the difference becomes smaller (0.214) and statistically insignificant ($t - stat = 0.984$). This means that formal and informal firms with one employee present the same wage, controlling for heterogeneous but observables characteristics of the households. This result corroborates the theoretical model presented in this paper, since it was argued that there was no wage discrimination for firms with the same size and using the same labor skill.

Even though the result is robust (low $t - stat$), it is worth to point out that the potential future benefits of being in the formal sector (such as social security or pension funds) were ignored. Unfortunately, the data does not allow us to identify how much is contributed to that end by each formal worker.[19] Ideally, future benefits

[18]The income contribution by children might be due to child labor in poor families in Brazil (see Cardoso e Souza, 2005).

[19]The Brazilian law stipulates that the contribution rate to pension funds goes from 7.65% to 20%, but the exactly rate for each worker is not easily determinable with the data available.

would be brought to present value and added to the observed wage. The next step would be to test whether this "adjusted" wage obtained by formal workers is still statistically equivalent to the associated informal wage. Considering that the self-employed wage discriminated on the survey is the gross wage and noting that the *Lei Complementar 123* of 12/2006 (a federal regulatory bill) restricts how much each formal self-employed can receive (a maximum that corresponds to approximately to $US\$1,500$), we proceed as follows. The minimum contribution rate (7.5%) is discounted from the gross wage that the formal firm has to pay and the present value of the maximum that each worker can receive is added. Thus, formal wage is increased by 10% to 50% on average.[20] Using a conservative approach, wages of all formal self-employed workers were increased by 50%. The results of the estimation with the adjusted wages are presented in Table 4.

Table 4: Dependent variable - ln (wages) - adjusted for future benefits

Sample	Formal	Controls	Difference	St. Err.	t-stat.
Unmatched	3.675	2.3882	1.286	0.093	9.830
ATT	3.675	3.232	0.443	0.254*	0.984
N	602	602			

* Calculated using bootstrap technique.

The new estimates suggest that the results are robust to possible future payments to the self-employed worker. This works as long as this benefit does not go over 50% of the wage in the formal sector.[21]

Therefore, there is no evidence that formal workers obtain higher wages relative to informal workers.

4 Conclusion

Existent empirical evidence indicates that, compared to high skilled counterparts, low skilled workers are more predominant in smaller firms in the informal economy. This paper studies the economic reasons for this empirical fact. This analysis complements the model presented by FMS in the sense that heterogeneity is now assumed to exist on the worker side rather than on the firm side. In the

[20] The average life expectancy in Brazil, 85 years old, is used to perform the calculation.

[21] For instance, if the formal wage is increased by 60%, then there is no longer statistical significance for the equality of wages between sectors. The corresponding estimation result is available upon request.

model, the allocation of low skilled workers to informal firms results from profit maximization decisions by firms, which can choose to operate in either sector (formal or informal). The model shows that, in order to generate the result suggested by empirical evidence, production (or some other variable highly correlated with labor productivity) must be costly concealed in the informal sector. However, if the concealment cost depends only on the number of workers (as assumed by FMS), the results are ambiguous, meaning that low skilled workers could be more predominant in the formal sector.

In addition, the paper presents an empirical analysis (using data on the Brazilian informal economy) confirming that more educated workers tend to be found in the formal sector. The analysis also shows that workers with the same productivity receive the same net wage regardless of the sector where they work. This evidence corroborates the theoretical analysis, since this wage equality was assumed in theory. By contrast, FMS considered that wages were regulated in the formal sector, being higher than in the informal sector. The empirical analysis also shows that there is no complete segmentation in terms of labor skill, meaning that there are firms in both sectors using the same type of labor. Again, this is consistent with the results of our theoretical model. Finally, data analysis indicate that there are firms of the same size (regarding the number of workers) in both the formal and informal sectors, which is inconsistent with the results found by FMS (in their model, small firms operate in the informal sector, while larger firms operate in the formal sector).

References

Antunes, A. R., and Cavalcanti, T. V. (2003). Accounting for the hidden economy: barriers to legality and legal failures, working paper.

Alinghan, M. and Sandmo, A., (1972). Income tax evasion: a theoretical analysis. Journal of Public Economics, 1: 323-338.

Botelho, F. and Ponczek, V. (2007). Segmentation in the Brazilian labor market, mimeo.

Cameron, A. and Trivedi, P. (2005). Microeconometrics: methods and applications, Cambridge University Press.

Carneiro, F. G. and Henley, A. (2001). Modelling formal vs. informal employment and earnings: micro-econometric evidence for Brazil. Univ. of Wales at Aberystwyth Management & Business Working Paper No. 2001-15.

Cowell, F. A. (1990). Cheating the government: the economics of tax evasion. Cambridge: MIT press.

Cremer, H. and Gahvari, F. (1996). Tax evasion and the optimal general income tax. Journal of Public Economics, 60: 235-249.

Deheijia, R. and Wahba, S. (2002). Propensity score matching methods for nonexperimental causal studies. Review of Economics and Statistics, 84: 969-980.

Dessy, S., and Pallage, S. (2003). Taxes, inequality and the size of the informal sector. Journal of Development Economics 70, 225-233.

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

Funkhouser, E. (1996). The urban sector in Central America: household survey evidence. World Development 24, 1737-1751.

Leontaridi, M. (1998) Segmented labour markets: theory and evidence. Journal of Economic Surveys 12(1): pp. 103-109.

Loayza, N. (1996). The economics of the informal sector: a simple model and some empirical evidence form Latin America. Carnegie-Rochester Conference Series on Public Policy 45, 129-162.

Menezes-Filho, N. A., 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.

Rauch, J. E. (1991). Modelling the informal sector formally. Journal of Development Economics 35, 33-48.

Rosenbaum, P. and Rubin, D. (1983). The central role of propensity score in observational studies for causal effects, Biometrika, 70: 41-55.

Schneider, F. and Enste, H. (2000). Shadow economies: size, causes and consequences. Journal of Economic Literature, 38:77-114.

Tannuri-Pianto, M. Pianto, D. (2002). Informal employment in Brazil - a choice at the top and segmentation at the bottom: a quantile regression approach. XXIV Meeting of the Brazilian Econometrics Society.