Skill differentiation between formal and informal employment

Enlinson Mattos*
Sãao Paulo School of Economics, Fundação Getulio Vargas, São Paulo, SP, Brazil
Laudo M. Ogura
Department of Economics, Grand Valley State University, Grand Rapids, MI, USA

April 2008

Purpose: This paper examines the existence of skill differentiation between formal and informal labor markets.
Design/methodology/approach: First, a theoretical model is developed under the assumption that concealment of production is increasingly costly for informal firms. Second, using data on the Brazilian self-employed economy, two methods are utilized to compare earnings in the formal versus the informal economy: propensity score and instrumental variable (IV) methods. For the IV estimations, state-level variables are used as instruments for individual's decisions. In addition, the effect of schooling on formality choice is analyzed.
Findings: The theoretical model implies that more productive firms tend to operate formally and the proportion of workers employed by formal firms is larger for higher skilled workers. In the empirical analysis, we find that formal firms are more productive than informal firms, controlling for workers' characteristics, and that higher workers' skill increase the probability of formal operation, as predicted by the theoretical model.
Originality/value: The paper provides an original theoretical model of skill differentiation in labor markets and empirically evaluates the implications of the model.
Keywords: informal sector, labor skill, less developed countries.

1 Introduction

The informal sector represents a large proportion of production activity in less developed countries. Shneider and Enste (2000) review the literature on informal economic activity and discuss the causes for the growth of the informal sector such as the rise of the burden of taxes and social security contributions, increased regulation (e.g., reduction of weekly working hours), earlier retirement, high unemployment rates, and the decline of civic virtue and loyalty toward public institutions combined with a declining tax morale. On the other hand, to control informal activities, most societies have adopted punitive or educational measures rather than reforms that reduce the cost of formal operation.[1]

The recent literature on the emergence of informal activities has relied on the heterogeneity of firms and entrepreneurs as the determinant factor, given the degree of regulation and the size of tax burden. For instance, see Gang and Gangopadhyay (1990), Rauch (1991),

*Corresponding author (e-mail address: enlinson.mattos@fgv.br). The authors would like to thank the helpful comments provided by an anonymous referee. Usual disclaimers apply.

Fortin, Marceau, and Savard (1997), Dessy and Pallage (2003), Straub (2005), Amaral and Quintin (2006), and Antunes and Cavalcanti (2007). In this literature, the informal sector endogenously emerges because tax evasion reduces the cost of operation. Fortin et al. (1997) also shows that smaller firms operate in the informal sector that pays lower wages. Straub (2005) and Amaral and Quintin (2006), on the other hand, focuses on how accessibility to credit markets affect the formality decision of heterogeneous firms.

Regarding labor markets, while formal workers usually have access to additional benefits (besides wages), individuals employed in the informal sector are excluded from them. A model based on this observation is presented by Rosen (1986), concluding that earnings should be higher in the less desirable informal sector to compensate the lack of fringe benefits. However, other authors conjecture that regulations create barriers to entry into the formal labor market, generating wage segmentation, i.e., earnings are higher in the formal labor market for similar workers. Tannuri-Pianto and Pianto (2002) and Botelho and Ponczek (2006), for instance, find evidence of the existence of wage differentials. Other authors have found the existence of wage differentials, but explain it based on non-observable characteristics of workers who joined each sector. See for instance Carneiro and Henley (2001) and Menezes-Filho, Mendes, and Almeida (2004). On the other hand, Maloney (2004) argues (based mostly on evidence from surveys) that most informal workers voluntarily choose informality because of non-pecuniary advantages (for instance, more independence in the case of self-employed workers), which could compensate for lower earnings. Besides differences in earnings, there are other differences between formal and informal labor markets. For instance, Funkhouser (1996) shows that returns on education are much lower in the informal sector and the male-female differential is much larger in the informal sector in five Central America countries. Tannuri-Pianto and Pianto (2002) find similar evidence: illiteracy is highly penalized and college education is highly rewarded in the formal sector in Brazil, but they also find that the marginal benefit of elementary education is higher in the informal sector.[2]

In this paper, we analyze the emergence of the informal sector focusing on the heterogeneity of not only firms, but also workers. Thus, labor market considerations are incorporated into the analysis. The key assumption in the model is that informal firms face an increasing cost of concealing production (due to the need to hide output from authorities, in addition to the expected losses from fines, confiscation, halted production, etc.). Hence, the emergence of informal activities occurs due to evasion benefits that firms can obtain, but differences in private skill characteristics of workers and in firms' efficiency levels induce high output firms to operate formally. Accordingly, in the same industry (characterized by a given type of labor skill), more efficient firms are formal while the less efficient ones are informal. In addition, since the cost of informality increases with production, industries that employ more productive workers tend to have a greater proportion of firms in the formal sector. This implies that formal firms have a greater probability of employing high skilled workers (as suggested by existing empirical evidence).

While the model presented in this paper follows the setup introduced by Fortin et al. (1997), thereafter referred as "FMS", the dependence of formality choice on heterogeneous labor skill distinguishes our model from FMS's one (workers are homogeneous in FMS's model). Moreover, it is important to stress that formality choice of firms using higher skilled workers depends on the concealment cost function. In particular, to credibly obtain the result that skilled workers are more commonly employed in the formal sector, the concealment cost function must be convex with respect to the firm's production level (i.e., this cost must go up at an increasing rate). In this case, informal operation is only optimal

when the production level is low (which is true when either workers are low productive or firms are inefficient). In contrast, the concealment cost in FMS's model is an increasing function of the number of employed workers. In that case, greater labor skill increases productivity, but also wage rates. Hence, labor demand by firms may not increase much or it could even decrease depending on the change in the productivity-to-wage ratio, stimulating informality in industries that employ high skilled workers.[3]

Amaral and Quintin (2006) also attempts to explain why workers in the formal sector are more educated on average. In their model, this result is due to restricted capital financing in the informal sector, making informal firms to substitute low-skilled labor for capital. Comparatively, in our model, capital is not explicitly considered and firm efficiency is taken as given, with productive firms facing higher cost of being informal because it is costly to conceal production. The two models, however, can be reconciled by arguing that the concealment cost in our model is largely determined by the cost of getting capital needed to become more productive.

Last, this paper also presents supporting empirical evidence. The econometric analysis presented here uses data on the Brazilian self-employed economy. For the sake of robustness, different methods (OLS, probit, instrumental variable, and propensity score) were used, with results indicating that formal firms are indeed more productive than informal ones (controlling for workers' characteristics) and that high skilled workers are more likely to be employed by formal firms.

The remaining of the paper is organized as follows. The next Section presents the theoretical model and analyzes the equilibrium distribution of workers between the formal and the informal sectors for each level of labor skill. Section 3 presents an empirical analysis that supports the theoretical result. Section 4 offers concluding remarks. In the Appendix, it is shown that theoretical results are sensitive to the specification of the concealment cost function (in particular, FMS's assumption that concealment costs are based on the need to hide workers is used to show that the results may not match the existent empirical evidence).

2 The model

2.1 Setup

In this section, a theoretical analysis of informal activity is presented, focusing on the heterogeneity of firms and workers. The setup of the model is based on FMS's paper, which analyzes how informality arises in a model with heterogeneous profit maximizing firms that can choose between formal and informal operation. In FMS's model, heterogeneity of firms is due to differences in technology or management ability, affecting their efficiency (productivity) level. In the present paper, workers are also heterogeneous due to differences in skill, with qualification indexed by a scalar $\phi \geq 0$. According to their skill level ϕ, workers get a wage rate $w(\phi)$, increasing in qualification, i.e., $w'(\phi) > 0$. This wage rate is assumed to be the same for workers in the formal and in the informal sectors, i.e., there is no wage differential (as Amaral and Quintin (2006) points out, there is no definitive evidence that wage differentials between formal and informal workers exist in practice for the same level of qualification). The determination of the wage function and workers' choices for skill qualification will not be formally analyzed here, although one can expect 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, a cost that includes learning effort

(to obtain qualification) and the opportunity of getting a lower skilled job. Based on the wage rates observable in the market, workers are supposed to choose a qualification level taking into consideration how much additional effort is needed to get additional income.

To simplify the analysis, each firm is skill specific, meaning that there is no substitution between workers of different labor skill levels.[4] The model assumes that all firms produce the same good with price normalized to \$1, although firms that use workers with higher skill obtain greater production.[5] In the remaining of the paper, firms that require high skilled labor will be called "high-ϕ" firms, while the ones that use low skilled labor will be called "low-ϕ" firms. The set of firms that use the same type of labor skill forms an "industry" (hence, each industry is characterized by a different labor skill requirement).

2.2 Formality choice by firms

In this subsection, a model is developed to show that highly efficient firms in each industry are more profitable in the "formal sector", i.e., in 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). This result is analogous to the one found by FMS, although our model adopts two distinct assumptions. First, to simplify the analysis, constant returns to scale in firm-level production are assumed rather than decreasing returns to scale. Second, firms in the informal sector face a concealment cost that increases with production rather than with the number of workers. As shown in the next subsection, this assumption makes it more likely that the proportion of formal employment is greater in industries that use higher labor skill. In contrast, if the concealment cost depends simply on the number of workers as in FMS's model, then firms in high skill industries may prefer to operate informally (this result is presented in the Appendix). Considering that firms are skill specific, the analysis in this subsection will be restricted to a particular skill level ϕ. Changes in the formality choice as ϕ increases are analyzed in the next subsection.

The aggregate production in 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 technologically better than others, producing more with the same amount of labor. This heterogeneity may be due to differences in management skill, technology, and other factors (locational advantages, relationship with buyers, etc.). 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.[6]

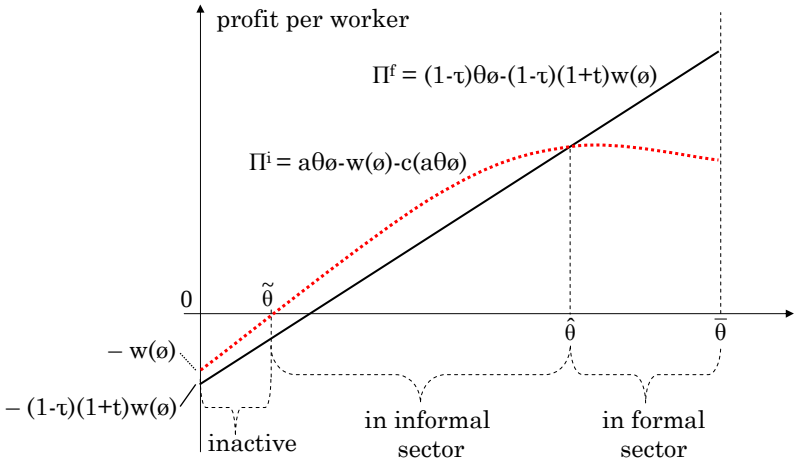
In the industry that uses labor skill ϕ, a formal firm can produce $y^f = \theta\phi$. Thus, its profit is

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

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 $y^i = a\theta\phi$, where $a < 1$ is a parameter that represents the inefficiency inherent to informal operation (due to, for example, difficulties in obtaining capital, technology, material, regular buyers, or financial credit). Besides the inherent inefficiency, informal firms also incur a concealment cost $c(y^i)$, which represents the cost of preventing legal punishments and the lost production due to punishment. This cost only exists when the firm produces a positive amount and increases with production at a growing rate, i.e., $c(0) = 0$, $c'(.) > 0$, and $c''(.) > 0$. Then, the profit of a firm in the

Figure 1: Potential firm profit in each sector



informal sector can be written as

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

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

Comparing profits given by equations (1) and (2), it can be concluded that a profit maximizing firm with efficiency level θ operates in the formal sector if

$$(1-\tau)\theta\phi - (1-\tau)(1+t)w(\phi) \geq a\theta\phi - w(\phi) - c(a\theta\phi) \geq 0. \tag{3}$$

Note that profit increases with θ at a constant rate $(1-\tau)\phi$ in the formal sector, while the rate is $[1 - c'(a\theta\phi)]a\phi$ in the informal sector. Because $c'(.)$ is an increasing function, the growth rate with respect to θ 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 for the moment the cases where all firms operate formally or informally, 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. Another interesting result is the discontinuity in the distribution of firm size in terms of production level: at $\hat{\theta}$, a firm would be indifferent between operating formally or informally, but production would be greater if the firm operated in the formal sector. Thus, there is a "missing middle" in terms of production level, i.e., as θ increases, production jumps from $a\hat{\theta}\phi$ to $\hat{\theta}\phi$. This "missing middle" result is also found in FMS's model, although firm size in their paper is measured by number of workers rather than by output level.

The following proposition summarizes the main result derived in this subsection.

Proposition 1 *Consider the following conditions:*

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

(ii) $(1-\tau)(1+t) \geq 1$, *and*
(iii) $c(a\theta\phi)$ *grows relatively fast with production.*
If the above conditions are satisfied, high-θ *firms operate in the formal sector, while low-*θ *firms operate informally.*

Conditions (i) and (ii) in Proposition 1, i.e., the inefficiency inherent to informal operation is lower than the tax burden in the formal sector and the effective labor cost is higher in the formal sector, imply that informal operation is advantageous for at least the lowest-θ firms. Condition (iii), on the other hand, guarantees that at least the highest-θ firms will be formal. In terms of government policies, note that the proportion of firms in the formal sector increases with the degree that condition (iii) in Proposition 1 is satisfied and the degree that conditions (i) and (ii) are not satisfied, holding everything else constant. Therefore, informality is reduced with greater and faster-increasing concealment cost on one hand, and smaller a, τ, or t on the other hand. In words, high tax rates, low enforcement, and light punishment stimulate informality.[7]

The conditions in Proposition 1 are based on the premise that low-θ firms find the informal sector more profitable. It is possible that firms would never operate informally. For instance, if conditions (i) and (ii) are not satisfied, then informal operation yields no advantage. Also, 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 (i.e., they would not exist in practice). On the other hand, if condition (iii) is not satisfied, then all firms would be informal because $\hat{\theta}$ would be greater than $\bar{\theta}$. Finally, if condition (ii) only is not satisfied (i.e., the effective labor cost is smaller in the formal sector), then it is possible that the π^i curve crosses the π^f curve twice, first from below and then from above, implying that both low and high-θ firms are in the formal sector, but medium θ-firms operate informally (in an extreme case, it is also possible that only low-θ firms would be in the formal sector if the concealment cost grows too slowly). Since empirical evidence is needed to support the plausibility of these alternative equilibrium cases, this paper will continue focusing on the case where low-θ firms operate informally and high-θ firms operate formally. It is important to notice, however, that the model is flexible enough to explain other possible cases that may be empirically found under specific industry conditions.

2.3 Skilled labor allocation

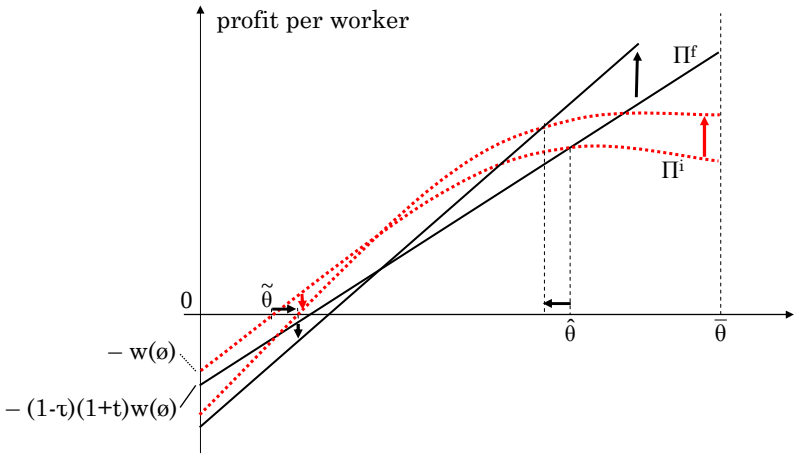
This subsection analyzes how formality choices by firms determine the allocation of workers between the formal and the informal sectors. The ultimate objective is to study how this distribution of workers varies with labor skill ϕ.

To figure out the labor demand by 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 the industry of labor-type ϕ, operating firms with efficiency level θ demand a total of $g(\theta, \phi)$ workers. 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 $L_\phi = \int_{\underline{\theta}}^{\bar{\theta}} g(\theta, \phi) d\theta$. Because α depends on the frequency distribution $g(\theta, \phi)$, it could increase or

Figure 2: Changes in firm profit as ϕ increases



decrease with ϕ.[8]

In order to analyze how the proportion of formal employment α changes with greater ϕ, first notice that $\hat{\theta}$ decreases with ϕ if $\left.\frac{d\pi^f}{d\phi}\right|_{\theta=\hat{\theta}} > \left.\frac{d\pi^i}{d\phi}\right|_{\theta=\hat{\theta}}$. The derivatives in this inequality can be obtained from (1) and (2):

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

and

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

Note that at $\hat{\theta}$ the slope of the π^f curve depicted in Figure 1 is greater than the slope of the π^i curve, which implies that $(1-\tau) > [1 - c'(a\hat{\theta}\phi)]$. Hence, formal operation revenue increases more than informal operation revenue net of concealment cost. This happens because the concealment cost is significantly increased by higher labor productivity at $\hat{\theta}$. On the other hand, because $(1-\tau)(1+t) \geq 1$ is assumed, the effective wage cost tend to increase more in the formal sector. In other words, greater labor productivity affects both revenues and costs more heavily in the formal sector. The potential profit, thus, could increase more or less in the formal sector. However, if the concealment cost takes away a large part of the labor productivity gain or if the effective wage cost advantage in the informal sector is very small, then the potential profit tends to increase more in the formal sector. Figure 2 depicts the changes in profit curves and the corresponding effect on $\hat{\theta}$. In summary, $\hat{\theta}$ decreases because higher labor productivity increases firm production, which thus makes the concealment cost of informal operation larger, inducing more firms to become formal. Therefore, a greater proportion of high-ϕ firms operate in the formal sector compared to low-ϕ firms.

However, the distribution of workers across sectors is also affected by changes in the firm distribution $g(\theta, \phi)$. In the simple case where $g(\theta, \phi)$ is constant with respect to ϕ, formal

labor demand L_ϕ^f increases with ϕ if $\hat{\theta}$ does so (because $\hat{\theta}$ is the inferior limit of the integral in equation (4)). Moreover, $\tilde{\theta}$ most likely increases with ϕ because the marginal productivity of low-θ firms should increase less than the increase in $w(\phi)$, making these firms less profitable. Therefore, α increases with ϕ, i.e., the proportion of formal employment is larger in high skill industries. This result is corroborated by the empirical evidence mentioned in the Introduction. Further supporting evidence is presented in the empirical part of this paper (next Section).

Nonetheless, a further analysis is required because the frequency distribution $g(\theta, \phi)$ should vary with ϕ in reality. However, note that firm productivity θ is determined in practice by capital use and technology. Since it is plausible that there is complementarity between labor skill and the amount of capital or technology used,[9] then ϕ and θ are complements. Consequently, high θ firms are more frequent in high-ϕ industries than in low-ϕ industries. In other words, $g(\theta, \phi)$ is more skewed toward high-θs as ϕ increases. For the same reason, the maximum firm efficiency $\bar{\theta}$ is probably greater in high-ϕ industries. According to (4), this change in distribution implies that the proportion of formal employment is likely higher in high-ϕ industries even if $\hat{\theta}$ and/or $\tilde{\theta}$ are slightly higher than in low-ϕ industries.

Proposition 2 summarizes the results in this subsection.

Proposition 2 *Consider the following conditions:*

(i) $g(\theta, \phi)$ *is either constant with respect to* ϕ *or becomes more skewed toward higher* θ *values as* ϕ *increases, and*

(ii) $(1-\tau)\hat{\theta} - [1 - c'(a\hat{\theta}\phi)]a\hat{\theta} > [(1-\tau)(1+t) - 1]w'(\phi)$.

If these conditions are satisfied, the proportion of formal workers is greater among high skilled workers than among low skilled ones.

As explained above, condition (i) in Proposition 2 guarantees that firms employing high skilled labor are not less intrinsically efficient, while condition (ii) implies that the incentive to be formal is increased when more productive workers are employed (basically because the greater productivity makes the concealment cost higher, i.e., because $c'(.)$ is large enough).

As the final remark, recall that Proposition 1 implies the 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 $\hat{\theta}$ tend to be larger in this case (thus there is a greater likelihood that $\hat{\theta} > \bar{\theta}$). Intuitively, holding everything else constant, completely informal operation can only happen in industries that use low productive labor, with low output per firm, because then the concealment cost would be low.

3 Empirical Analysis

Data on the Brazilian informal economy is used to perform econometric assessments of the implications of the theoretical model proposed in the paper. The main testable results of the model are stated in Propositions 1 and 2. In short, more productive firms and high skilled workers tend to be found in the formal sector. Simple tests are proposed to verify these assertions. First, because firms that operate more productively tend to be in the formal economy, in equilibrium formal firms should be more productive than informal ones, controlling for labor skill. Second, the probability of formal firm operation should be greater if they employ higher skilled workers.

Table 1: Descriptive Statistics

Variable	Mean	Std. Deviation	Minimum	Maximum
output_hour	20.74	40.50	0	500
tenure	7.713	9.929	0	30
age	36.328	14.175	10	60
schooling	8.443	4.496	1	16
family_schooling	3.666	1.458	1	6
white	0.468	0.499	0	1
formal	0.469	0.499	0	1
otherincome	0.019	0.136	0	1
children	1.713	1.279	0	13
bornplace	0.511	0.500	0	1
urbanarea	0.807	0.395	0	1
pensioners87	.052	.0153	0	0.083
eduexp87	5.43E-11	2.83E-11	0	1.75E-10

3.1 Brazilian self-employed economy

According to the 2003 ENCIF (Informal Urban Economy Survey) elaborated by the IBGE (Brazilian Institute of Geography and Statistics), there are more than ten million informal firms in Brazil, mostly allocated in local commerce and small services (69% of all informal workers are employed in these activities). The survey also indicates that 88% of informal firms in Brazil are classified as self-employed workers. Hence, it is essential to take the self-employed category into consideration when analyzing the Brazilian informal economy. Because there is also a large amount of formal self-employed workers (15% in our sample), this type of firms can be easily used to test hypothesis on formality choice. One advantage of this strategy is that more precise information on informal labor characteristics can be obtained when asking self-employed workers than when asking firms with larger number of workers (in which case the information may not be reliable because managers of informal firms are reluctant to talk about their operations).[10] Also notice that firms with the same number of workers commonly coexist in the formal and the informal sectors, supporting our view that formality choice is related to firms' production level rather than to the number of workers employed by firms (which was the case assumed in FMS's model, where only firms with large number of employees end up operating formally).

The data used in the empirical analysis is from the 2005 PNAD (Brazilian National Household Survey). Table 1 provides statistical description of the variables considered. The output per hour of self-employed workers is represented by the variable *output_hour*, which is computed by adding three components of income: revenue, value of the merchandise produced/sold that was consumed by the household (rather than actually sold), and other monetary benefits from the activity, divided by the number of work hours.[11] For the estimations, only self-employed workers employed in the commerce activity were considered because it is easier to compare production levels when all workers have the same activity. These workers represent more than 90% of the workers in the sample available. The other variables in Table 1 attempt to capture individual and geographical characteristics of workers.

Regarding individual characteristics, the variable *formal* is a dummy for working in the formal sector (value 1 in this case). The variable *tenure* represents how long the individual

has been working as a self-employed, while *age* indicates how old she is. Educational level is measured by years of schooling of the worker (*schooling*) and the average years of schooling of other family members (*family_schooling*). *white* is a dummy for race (value 1 if white) and *otherincome* is a dummy variable that takes the value 1 if the worker has an additional income source. Finally, the variable *children* indicates how many children live with the worker.

Geographic characteristics are represented by the variables *citynative* and *urbanarea*, which are dummy variables with value 1 if, respectively, the individual was born in that municipality (to capture the extension of her social network) and if that municipality belongs to a metropolitan area.

Last, the per capita state expenditure on education in 1987 (*eduexp87*, measured in the national currency unit of 2005) and the percentage of beneficiaries of public pension funds in the state population in 1987 (*pensioners87*) are also taken into consideration as proxies for, respectively, schooling availability and perception on fringe benefits from formality. Because these variables are exogenous to the individual's decision to study and to work, they will be used for robustness check. Their use and justification are explained later.

3.2 Estimation

First, we test if firms with higher production per hour tend to operate formally, controlling for the characteristics of workers. In particular, we estimate

$$output\_hour = \beta_0 + \beta_1 formal + \beta_2 controls + \epsilon, \tag{7}$$

where βs and γ correspond to the coefficients of the explanatory variables (*controls* is a vector of the additional control variables, explained below) and ϵ is the error term. Then, we test if more skilled workers tend to be in the formal sector by estimating the effects of schooling (the measure for skill) on formal entry probability:

$$formal = \delta_0 + \delta_1 schooling + \delta_2 controls + \epsilon. \tag{8}$$

Regarding these tests, there is a related literature that focuses on wage differentials between formal and informal sectors (see Leontaridi (1998), for instance, who surveys the literature on labor market segmentation). One problem in the estimation of such wage differentials is that workers' skill may be affected by the decision to work formally or informally because returns from skills accumulation are different in each sector. In other words, while workers decide on formal or informal employment based on the returns from their skill level, skill level choices depend on the returns from higher education in the sector where each worker expects to be employed. Hence, the estimator might be biased due to endogeneity (this problem is discussed in more detail in Botelho and Ponczek (2006)).

The same endogeneity issue described above applies to our tests for differentials in output and labor skill across sectors. While formality of firms depends on their workers' skill, skill choices may have been affected by workers' choices on formality, which is affected by the availability of formal and informal jobs. Because output depends on labor productivity, the endogeneity problem also applies to testing output differentials.[12] Considering the limited availability of data to overcome this problem, several alternative procedures are presented for the sake of robustness.

First, the propensity score (PS) method is used, consisting of a probit estimation run on the workers' individual characteristics to compute the probability of being formal. Then,

output levels of firms corresponding to similar workers are compared by choosing observations in each group (formal and informal) with similar PSs, i.e., observations with similar predicted probit value.[13] The PS (a conditional probability of being formal) is a matching measure much simpler than conditioning the probability 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 (multiple dimensions) to compare workers in the formal and informal sectors (see Deheijia Wahba (2002) for details on this advantage of the PS methodology). To implement the 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 set of X_k, i.e., those workers in the formal sector who present close characteristics (in terms of Euclidean distance) compared to the individual k. Let N_i denote the number of informal self-employed workers that will be used to compare with the formal worker k (i.e., of informal workers whose characteristics are similar to the individual k). Let $h(k, j)$ represent the weight given to the j^{th} informal worker associated with the k^{th} formal worker, with $\sum_j h(k, j) = 1$. Generally, the formula for the matching estimator is

$$\Delta q = \frac{1}{N_f} \sum_{k \in \{formal\}} [q_{f,k} - \sum_j h(k, j) q_{i,j}], \tag{9}$$

where Δq is the output differential between sectors, N_f is the number of formal workers, $\{formal\}$ is the set of all formal self-employed workers, $q_{f,k}$ is the output of the formal worker k and $q_{i,j}$ is the output of the j^{th} informal worker that belongs to the comparison group of worker k.[14] Note that $0 < h(k, j) \leq 1$. Different matching estimators are generated by varying the choice of $h(k, j)$. In this paper, the nearest-neighbor matching is used, choosing the set $H_j(p(X)) = \{j : min\,|p(X_k) - p(X_j)|\}$ for every formal worker k, where $p(X)$ denotes the PS and $|p(X_k) - p(X_j)|$ is the Euclidean distance between the PSs of the formal k^{th} self-employed worker and the j^{th} informal one. In other words, the "closest" informal self-employed individual is considered when comparing to the formal self-employed worker. To avoid the endogeneity problem, *eduexp87* and *pensioners87* are included under the assumption that they cause exogenous variation of individual's choices regarding education and formality respectively. The justification for using 1987 data is that the current education and formality choices are likely to be correlated to the state government spendings in education and the extension of the formal pension system when workers were still young, at an age where study and work decisions are crucial in determining their future education and formality choices (the average worker in the sample was 18 years old in 1987).[15] Other control variables included are *tenure*, *tenure*^{2}, *tenure*^{3}, *age*, *age*^{2}, *age*^{3}, *white*, *otherincome*, *children*, *citynative*, and *urbanarea*.[16] Conditional on these variables, the decision to enter the formal sector is assumed to be random, thus allowing us to find an unbiased estimator for the output differential between formal and informal sectors. The results of the PS estimation are presented in Table 2, column 5. In the table, the estimated production differential for the self-employed worker with mean characteristics is shown. The results presented in Table 2 will be discussed after the alternative approaches are explained.

The second approach considered is the instrumental variable (IV) method. The state-level variable *eduexp87* is used as instrument for education level, while *pensioners87* is used as instrument for formality choice. The state government expenditure on education in each state when the worker was (on average) 18 years old seems to be a good instrument for schooling because it is not caused by the current individual earnings.[17] Regarding the

Table 2: Estimations of Production-per-hour Differential

Controls	(1) Mean	(2) OLS	(3) 2SLS[a]	(4) Treated[a]	(5) PS[a]
(none)	0.916***				
	(0.097)				
Schooling[b]		0.332***	4.5***	1.7***	0.153
		(0.116)	(1.58)	(0.517)	(0.245)
I. Schooling[b]		0.60***	4.6***	3.5***	0.415**
		(0.11)	(1.867)	(0.112)	(0.207)

All values reported in the table above corresponds to the estimated output per hour differential.
[a] Decision to enter the formal market is instrumentalized.
[b] Other control variables were also included (see text).
***, **: statistically significant at 1% and 5% level respectively.
Standard errors are reported inside parentheses.

formality choice decision, the proportion of pensioners in 1987 is used as an instrument under the presumption that this variable generates a "demonstration effect", i.e., workers in 1987 were influenced by this proportion of formal pension beneficiaries when deciding between working in the formal or the informal sector. Hence, for the IV method, the following model is estimated:

$$ln(output\_hour) = \mu_0 + \mu_1 formal + \mu_2 schooling + \mu_3 controls + \epsilon, \tag{10}$$

where the additional control variables are the same ones used for the PS method. Here, the interest is on the estimate for μ_1, which indicates whether there is a difference between the production of formal and informal firms, controlling for skill level (measured by schooling). To deal with the problem of potential endogeneity in the determination of the variables *formal* and *schooling*, first-stage OLS regressions are performed for each of these variables on their instruments (*pensioners87* and *eduexp87* respectively) and controls to capture only the exogenous effects of these variables on productive efficiency. Results for the estimated production differential are shown in Table 2, column 3.[18]

Last, an alternative IV estimation is attempted based on the fact that the decision to enter the formal sector is a binary choice. Accordingly, now a probit regression is used in the first stage. The second stage is a standard IV regression to test whether self-employed workers in the formal sector (the "treated" group) are more productive. We call this approach the "treated" model. Again, in the second stage, the decision to enter the formal sector is instrumentalized, with *pensioners87* being used as the instrument for formality choice (the instrument is used in the first stage probit estimation). The resulting estimates of output differential using this approach are shown in column 4 in Table 2.[19]

Comparing the results presented in Table 2, the first column presents the average output differential between sectors, which is positive (0.916) and statistically significant (t-stat = 9.44). For the other estimations, the case where schooling is taken as exogenous is reported in row 2 (while row 3 presents the case where schooling is assumed to be endogenous – "I. Schooling" denotes the case where an instrument for schooling is used). In column 2, row 2, the result is generated with standard OLS procedure with the variable *schooling* included. For row 3, the exogenous proxy *eduexp87* is used alternatively. In column 3, standard OLS

Table 3: Formal Sector Entry Probability

Dependent variable: *formal*

	OLS	2SLS[a]	Probit	IV Probit[a]
tenure	0.005**	-0.011*	0.030**	-0.020**
	(0.003)	(0.006)	(0.014)	(0.010)
$tenure^2$	-0.000	0.001	-0.001	0.001
	(0.000)	(0.000)	(0.001)	(0.001)
$tenure^3$	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
age	-0.029***	-0.081***	-0.101	-0.139***
	(0.008)	(0.023)	(0.068)	(0.041)
age^2	0.001***	0.002***	0.004**	0.004***
	(0.000)	(0.001)	(0.002)	(0.001)
age^3	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
white	0.059***	-0.131**	0.323***	-0.257***
	(0.009)	(0.056)	(0.053)	(0.053)
otherincome	-0.018*	0.106***	-0.181***	0.177***
	(0.010)	(0.040)	(0.070)	(0.049)
children	-0.007*	0.031**	-0.064***	0.049***
	(0.004)	(0.013)	(0.025)	(0.017)
citynative	-0.006	-0.068***	-0.041	-0.136***
	(0.010)	(0.025)	(0.053)	(0.034)
urbanarea	0.000	-0.300***	0.120	-0.545***
	(0.013)	(0.089)	(0.118)	(0.073)
pensioners87	1.859***	1.444**	9.968***	3.313**
	(0.303)	(0.606)	(1.823)	(1.563)
schooling	0.018***	0.144***	0.091***	0.280***
	(0.001)	(0.035)	(0.007)	(0.004)
constant	0.052	-0.073	-2.612***	-1.234**
	(0.108)	(0.236)	(0.945)	(0.578)

Number of observations: 4,732

***, **, *: statistically significant at 1%, 5%, and 10% level respectively. Standard errors are reported inside parentheses.

[a] State government education expenditures in 1987 (*eduexp87*) is used as instrument for *schooling*.

estimations are used in the first stage for schooling and formality choice (and the predicted values are used in the second stage estimation). For row 2, however, there is no first stage estimation for schooling (i.e., the variable *schooling* is included directly in the second stage). Contrasting the output differentials shown in columns 2 and 3 in Table 2, observe that the OLS approach generates an attenuation bias compared to the IV approach, resulting in smaller estimated differential. In column 4, probit estimation is used in the first stage for formality choice. The difference between the results in rows 2 and 3 is that the variable *schooling* is used for the estimate in row 2, while *eduexp87* is considered alternatively for row 3. For the PS method (results presented in column 5), *schooling* and *eduexp87* are used alternatively in the first stage (i.e., in the probit regression, with *schooling* used for the result in row 2 and *eduexp87* for the result in row 3). As can be noticed, a positive output differential is found to be statistically significant for all methods used regardless of whether schooling is taken as exogenous or endogenous (except for the PS method with exogenous schooling). The largest of the estimates is for the IV estimation (column 3), with a differential of 4.6. Note that the result for the treated IV estimation is not too different than the one for the standard IV estimation. The use of the PS method (with endogenous schooling) generates a low estimated difference in production (only 0.415), but still statistically significant (t-stat = 2.00). In summary, while the estimated value of the output differential depends on the method used, the results indicate that in fact formal self-employed workers (firms) have greater productive efficiency, controlling for heterogeneous but observable characteristics of workers. These results, therefore, support the findings of the theoretical model presented in the paper.

Now, turn to the assessment of the second hypothesis, i.e., that a greater proportion of skilled workers are formal compared to unskilled ones. Recall that to estimate production differentials using the IV, the treated, and the PS methods, first stage regressions were performed to predict formal sector entry probability. According to the second hypothesis, higher skill of workers (measured by schooling) should increase that probability of formal operation. Hence, those first stage regressions can be used for testing the second hypothesis. Results using the different methods are presented in Table 3. In the table, the "OLS" results correspond to the standard OLS regression including schooling as exogenous variable. The "2SLS" results are generated by IV estimation with *eduexp87* taken as instrument for schooling. The "Probit" results come from probit estimation of formality probability, with exogenous schooling (i.e., *schooling* is used directly).[20] Last, the results for "IV Probit" come from an additional IV estimation attempted, using *eduexp87* as instrument for *schooling*. Note that, in the estimations, the variable *pensioners87* is included as "instrument" for the decision to enter the formal labor market. Results shown in Table 3 indicate that this variable is indeed statistically significant for all estimations. More importantly, the schooling variables (exogenous or endogenous) affects positively (at 1% significance level) the probability of entry into the formal sector, although the size of its impact depends on the estimation method, ranging from 2 to 28 percentage points. Nonetheless, the estimations confirm that more skilled workers tend to work in the formal sector as predicted by the theoretical model.

4 Conclusion

Previously existent empirical evidence indicates that high skilled workers are more predominant in the formal economy compared to low skilled counterparts. This paper studies

the reasons for this empirical fact and provides further supporting evidence.

In the model presented here, the allocation of high skilled workers to formal production results from profit maximizing decisions by firms, which can choose to operate formally or informally. To obtain the result suggested by empirical evidence, production (or something else correlated with labor productivity) is assumed to be costly concealed in the informal sector. Then, informal operation of firms employing skilled workers would be unattractive because the concealment cost would be too high. We also show (in the Appendix) that this result depends on the specification of the concealment cost function. In particular, if this cost is associated with the need to conceal workers (as assumed in FMS's paper), then skilled labor could end up being allocated to the informal sector instead. This could happen because high productive workers must be paid greater wage rates, thus the costly concealed labor use of firms employing high qualified workers might not become large enough to stimulate formal operation.

The paper also presents an empirical assessment of the main results of the theoretical model. Data on the Brazilian self-employed workers economy was used because formal and informal self-employed workers coexist in large number in Brazil, with available information on the characteristics and production of these workers. The econometric analysis indicates that indeed formal enterprises are more productive and that skilled workers tend to be found in the formal sector. As the last remark, because both formal and informal self-employed workers are commonly observed, it is clear that formal and informal firms employing the same number of workers can coexist. While this observation is not inconsistent with the theoretical model developed in this paper, it dissonates with the results of FMS's model, which suggest that only firms with large number of workers operate in the formal sector.

Appendix: The case of costly concealment of labor

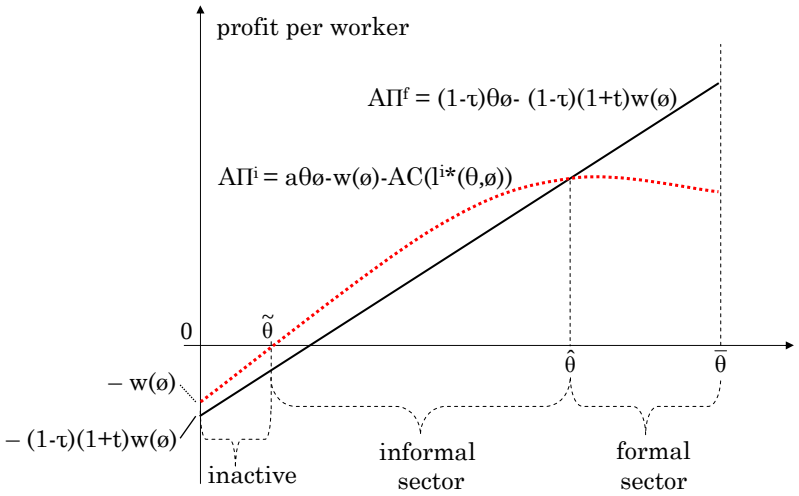
In this appendix, it is shown that if the cost of informal operation is a function of the number of workers employed by the firm, then it is possible that a greater proportion of high skilled workers might be in the informal sector rather than in the formal sector (as implied by the model in Section 2). The intuitive explanation for the ambiguity is that greater labor skill might imply that wages have to increase more than productivity (for less efficient firms). The resulting labor use decline would make concealment costs to go down too, thus inducing firms that employ high skilled workers to be more attracted to the informal sector. A complete explanation is provided below.

Modified FMS's model

Consider the original FMS model, modified to incorporate labor skill differences. The model is simplified with constant returns to scale in the production of the firms, although profit in the informal sector exhibits decreasing returns due to the concealment cost that increases with labor use at increasing rates. Formally, a firm operating in the informal sector incurs a cost $C(l^i)$, with $C(0) = 0$, $C'(.) > 0$ and $C''(.) < 0$. The profit per worker function in the formal sector continues to be given by equation (1). In the informal sector, firm profit is now given by

$$\pi^i_\phi = a\theta\phi l^i - w(\phi)l^i - C(l^i). \tag{11}$$

Figure 3: Potential firm profit in each sector - $C(l^i)$ case



thus, profit maximization implies that labor demand by an informal firm is given by

$$a\theta\phi - w(\phi) = C'(l^{i*}), \tag{12}$$

where l^{i*} is the optimal number of workers. Note from (10) that l^{i*} is an increasing function of θ:

$$\frac{dl^{i*}}{d\theta} = \frac{a\phi}{C''(l^{i*})} > 0. \tag{13}$$

In words, more efficient firms hire more workers because labor productivity will be greater.

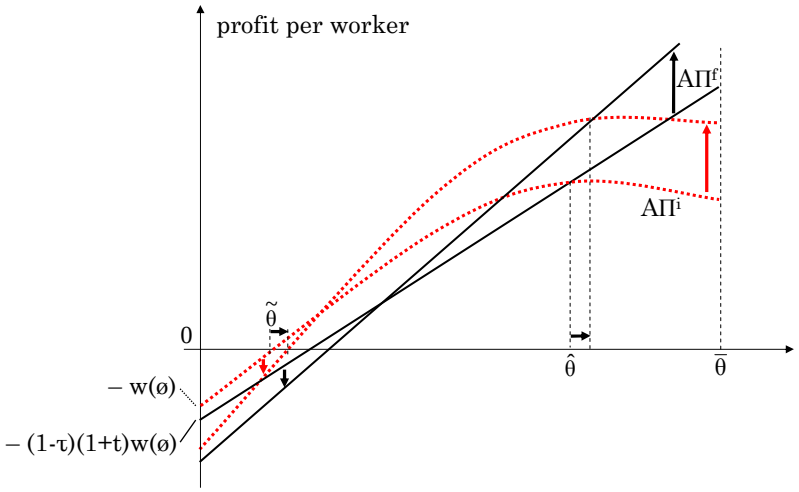
To figure out whether firms operate in the formal sector, consider again the more plausible cases where $a \geq (1-\tau)$ and $(1-\tau)(1+t) \geq 1$. Then, recall that formal firms' profit per worker increases with θ at a rate $(1-\tau)\phi$. By contrast, the informal firms' profit $(A\pi)$ per worker is:

$$A\pi^i_\phi = a\theta\phi - w(\phi) - AC(l^i), \tag{14}$$

where $AC(l^i) = \frac{C(l^i)}{l^i}$. Thus, noting that $AC'(l^i) = \frac{1}{l^i}\left[C'(l^i) - \frac{C(l^i)}{l^i}\right] > 0$, the growth rate (with respect to θ) of the informal firms' profit per worker is $\left(a\phi - AC'(l^{i*}(\theta))\frac{dl^{i*}(\theta)}{d\theta}\right)$. Hence, because $a \geq (1-\tau)$, increasing θ affects average production more in the informal sector, but on the other hand the average concealment cost is increasing (due to the convexity of the concealment cost function). Therefore, the informal sector may be more advantageous for low-θ firms, but eventually (for higher θ) the concealment cost becomes too expensive, inducing firms to operate formally. This situation is depicted in Figure 3, which represents the equilibrium case analogous to Figure 1.

Again, additional equilibrium cases exist depending on the parameters of the model. Analogously to the model in Section 2, all firms could be in the informal sector or in the formal sector. Double-crossing of the profit curves could also occur. The conditions for these special cases are analogous to the ones discussed in Section 2.

Figure 4: Changes in firm profit as ϕ increases - $C(l^i)$ case: $\hat{\theta}$ may increase



Skilled labor allocation

To analyze how profits in each sector are affected by increased labor skill ϕ, first note from (12) that labor demand by an informal firm may either increase or decrease as labor skill increases:

$$\frac{dl^{i*}}{d\phi} = \frac{a\theta - w'(\phi)}{C''(l^{i*})} \lessgtr 0. \tag{15}$$

Labor demand may go down because higher labor skill increases labor cost. If the wage rate increases more than the marginal productivity for informal firms, then these firms would hire fewer workers. Accordingly, because $C'(.) > 0$, the concealment cost $C(l^{i*}(\theta))$ could decrease.

To compare the resulting changes in potential profits in the formal and informal sectors, note that equations (1) and (11) yield, respectively,

$$\frac{dA\pi^f}{d\phi} = (1-\tau)\hat{\theta} - (1-\tau)(1+t)w'(\phi) \tag{16}$$

and

$$\frac{dA\pi^i}{d\phi} = a\hat{\theta} - w'(\phi) - AC'(l^{i*})\frac{dl^{i*}}{d\phi}. \tag{17}$$

Thus, if $\frac{dl^{i*}(\hat{\theta})}{d\phi} \leq 0$, then $\frac{dA\pi^i}{d\phi} > \frac{dA\pi^f}{d\phi}$ at $\hat{\theta}$ (because $a \geq (1-\tau)$, $(1-\tau)(1+t) \geq 1$, and $AC'(l^i) > 0$). In this case, the threshold $\hat{\theta}$ increases with ϕ, meaning that L^f_ϕ decreases with ϕ according to (4), holding everything else constant. This case is depicted in Figure 4. This result, however, contradicts empirical evidence that indicates a larger proportion of high-skilled workers in the formal sector.

The explanation for the contradicting result is the following. On the revenue side, higher labor skill increases productivity more in the informal sector than in the formal sector (because $a > 1-\tau$), stimulating informality in high-ϕ industries. On the cost side, the effective wage is increased more in the formal sector than in the informal sector (because

$(1-\tau)(1+t) \geq 1$), also stimulating informality. The concealment cost, however, may not be greater in high-ϕ industries because labor demand might not go up. Hence, high-ϕ firms have a greater incentive to be informal than low-ϕ firms.

The final question is whether it is plausible that labor use would not be significantly larger in higher-ϕ industries. To study this possibility, first analyze when $\frac{dl^{i*}(\hat{\theta})}{d\phi}$ would be positive. Equation (15) suggests that this is true when $\hat{\theta}$ is large enough. Suppose that $\hat{\theta}$ is high in low-ϕ industries (matching empirical evidence that indicates less formal employment among low-ϕ workers). Then, $\frac{dl^{i*}(\hat{\theta})}{d\phi}$ is likely positive. Consequently, the concealment cost could be raised enough to decrease $\hat{\theta}$ with higher ϕ. However, as $\hat{\theta}$ decreases, it would make $\frac{dl^{i*}(\hat{\theta})}{d\phi}$ smaller and smaller, then inducing $\hat{\theta}$ to converge to a certain value. Ultimately, this means that formal employment might increase with ϕ for low-ϕ levels, but eventually it may stop increasing. Notwithstanding, the proportion of formal employment could still be greater in high-ϕ industries despite increasing $\hat{\theta}$ if $g(\theta, \phi)$ is increasingly skewed toward high θ values as ϕ increases.

Concluding, the comparison of the case presented here (with concealment cost function based on labor use) with the case presented in Section 2 (where the concealment cost was based on production level) indicates that the proportion of formal employment for each labor skill level depends on how the concealment cost function is specified: unless the concealment cost does not increase significantly as skill level increases, theoretical results are less likely to match empirical evidence.

Notes

[1]See also Williams (2006) for a direct survey approach to measure the shadow economy.

[2]See Ingham (1989) and Nickell(1993) for the effects of education on unemployment and Ezeala-Harrison (1992) for an empirical test for the efficiency-wage model in Nigeria.

[3]For instance, if the concealment cost depended only on the need to hide workers, then a self-employed medical doctor could have a greater incentive to evade than a self-employed maid because the doctor's taxable income is greater. In contrast, if the concealment cost is a convex function of production, then the incentive to produce formally is greater for the more productive worker.

[4]While substitution could exist (i.e., a workers with skill level ϕ could work in an industry that requires a lower level), since getting qualification is costly, competition should make workers to supply their labor in the industry that corresponds to their ϕ level in equilibrium.

[5]An alternative interpretation for this assumption is that firms using higher skilled workers produce a higher valued product. In this case, however, relative prices must not be affected by changes in the supply of each product caused by formal or informal activity choices (the subsequent analysis assumes that production is lowered when a firm switches to informal operation, although the assumption does not affect the qualitative results).

[6]Instead of one worker per firm and heterogeneous firms, the same setup can be used if each industry has a representative producer facing 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 indicates that about 88% of informal firms employ only one worker (i.e., they are run by self-employed workers). This evidence is further discussed in the empirical part (next Section) of this paper.

[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 formal operation.

[8] Even though the paper does not analyze the labor market, it is assumed that the wage rate $w(\phi)$ makes the supply of each type of labor to exactly equal demand L_ϕ.

[9]Hamermesh (1993) discusses the existence of this complementarity.

[10]See Hamilton and Harper(2002) for a theory on entrepreneurship behavior.

[11]We measure efficiency differently from Uri (2002) but similar to Abdel-Rahman(1988).

[12]The need for taking more careful consideration of this endogeneity issue was suggested by an anonymous referee.

[13]For a more detailed explanation of this method, see Cameron and Trivedi (2005) and Rosenbaum and Rubin (1983).

[14]The equation above gives a general formula for the ATET - average treatment effect on the treated - estimator. See Cameron and Trivedi (2005), p. 863.

[15]In Brazil, state governments were the predominant providers of schooling in 1987. Also, the large majority of retired formal workers received some pension benefits from public pension funds.

[16]The choice of these variables is based on the existing literature on labor supply estimation. For instance, see Conway (1997).

[17]See Blundell et al. (2005) for a discussion on the endogeneity of schooling.

[18]Results for the first-stage estimations are available upon request.

[19]Even with these IV strategies, some endogeneity effects may still exist. For instance, one could argue that the instruments do not address the estimation bias due to the idiosyncratic characteristics omitted in cross-section samples.

[20]The "OLS", "2SLS", and "Probit" results in Table 3 correspond to the first stage regressions for the estimations of production differential using, respectively, the IV method with exogenous schooling, the IV method with endogenous schooling, and the treated method (and PS method) with exogenous schooling.

References

Abdel-Rahman, A-M. (1988). Optimality and Production Efficiency of Some Cropsharing Systems, Journal of Economic Studies, 15, 1, 53 - 70.

Amaral, P.S. and Quintin, E. (2006). A competitive model of the informal sector. Journal of Monetary Economics, 53: 1541-1553.

Antunes, A.R. and Cavalcanti, T.V.V. (2007). Start up costs, limited enforcement, and the hidden economy. European Economic Review, 51: 203-224.

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

Blundell, R., Dearden, L., and Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168: 473-512.

Botelho, F. and Ponczek, V. (2006). Segmentation in the Brazilian labor market. Proceedings of the 34th Meetings of the Brazilian Economics Association (ANPEC).

Cameron, A. and Trivedi, P. (2005). Microeconometrics: methods and applications. New York, NY: 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, MA: MIT Press.

Conway, E. (1997). Labor supply, taxes, and government spending: a microeconometric analysis. Review of Economics and Statistics, 79: 50-67.

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.

Ezeala-Harrison, F. (1992). An Empirical Framework for the Efficiency-wage Model: Use of Micro Data for Nigeria, Journal of Economic Studies, 19, 3.

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.

Gang, I. and Gangopadhyay S. (1990), A Model of the Informal Sector in Development, Journal of Economic Studies, 17, 5.

Hamermesh, D.S. (1993). Labor Demand. Princeton, NJ: Princeton University Press.

Hamilton, R. and Harper, D. (2002). The entrepreneur in theory and practice, Journal of Economic Studies, 21,6, p.3-18.

Ingham , M. (1989), Education and Youth Unemployment: A Reappraisal, Journal of Economic Studies, 16 , 3.

Leontaridi, M. (1998). Segmented labour markets: theory and evidence. Journal of Economic Surveys, 12: 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.

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

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: 235-248.

Nickell, S. (1993) Unemployment Revisited, Journal of Economic Studies, 20,1/2.

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.

Straub, S. (2005). Informal sector: the credit market channel. Journal of Development Economics, 78: 299-321.

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

Uri, N. (2002). Measuring the productive efficiency in telecommunications in the US, Journal of Economic Studies, 29, 2, p. 150-167.

Williams, C. (2006) Evaluating the magnitude of the shadow economy: a direct survey approach, Journal of Economic Studies, 33, 5, 369 - 385.