The Informal Sector *

Áureo de Paula †
University of Pennsylvania

José A. Scheinkman ‡
Princeton University and NBER

First version: January 11, 2006
This Version: July 12, 2006

*We thank Rita Almeida and Sandra Brandão for their help with the enforcement data, Joana Monteiro and Juliano Assunção for providing the ECINF dataset and for clarifying its content and methodology, and Julio Cacho for research assistance. We also benefited from comments by seminar participants at the 2006 NASMES. Scheinkman's research was supported by the National Science Foundation through grant 0350770.

†Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. E-mail: aureo@sas.upenn.edu

‡Department of Economics, Princeton University, Princeton, NJ 08544. E-mail: joses@princeton.edu

Abstract

This paper investigates the determinants of informal economic activity. We present two equilibrium models of the determinants of informality and test their implications using a survey of 50,000+ small firms in Brazil. We define informality as tax avoidance; firms in the informal sector avoid tax payments but suffer other limitations. In the first model there is a single industry and informal firms face a higher cost of capital and a limitation on size. As a result informal firms are smaller and have a lower capital-labor ratio. These implications are supported by our empirical analysis. A novel theoretical contribution in this paper is a model that highlights the role of value added taxes in transmitting informality. It predicts that the informality of a firm is correlated to the informality of firms from which it buys or sells. The model also implies that higher tolerance for informal firms in one production stage increases tax avoidance in downstream and upstream sectors. Empirical analysis shows that, in fact, various measures of formality of suppliers and purchasers (and its enforcement) are correlated with the formality of a firm. Even more interestingly, when we look at sectors where Brazilian firms are not subject to the credit system of value added tax, but instead the value added tax is applied at some stage of production at a rate that is estimated by the State, this chain effect vanishes.

JEL Codes:

1 Introduction

In this paper we investigate the determinants of informality. It is difficult to unambiguously define informal activities but estimates indicate that in 1990-1993 around 10% of GDP in the United States was produced by individuals or firms that evaded taxes or engaged in illegal pursuits. It is also estimated that these activities produce 25 to 35% of output in Latin America, between 13 to 70% in Asian countries, and around 15% in O.E.C.D. countries. (see Table 2 in Schneider and Enste [17]).

Informality creates a fiscal problem, but there is also growing evidence that informal firms are less efficient,[1] perhaps because of their necessarily small scale, perhaps because of their lack of access to credit or access to the infrastructure of legal protection provided by the State. For less developed countries, creating incentives for formalization is viewed as an important step to increase aggregate productivity.

We present two equilibrium models of the determinants of informality and test their implications using a survey of 50,000+ small firms in Brazil. In both models informality is defined as tax avoidance. Firms in the informal sector avoid tax payments but suffer other limitations.

The first model can be seen as a variant of Rausch [14], who relied on the modeling strategy of Lucas [11] in which managerial ability differs across agents in the economy, and assumed a limitation on the size of informal firms. We make a key modification that generates testable implications. The firms in our model use capital in addition to labor and informal firms face a higher cost of funds. This higher cost of capital for informal activities has been emphasized by DeSoto [4] who wrote that "Even in the poorest countries, the poor save. The value of savings among the poor is, in fact, immense − forty times all the foreign aid received throughout the world since 1945. (...) But they hold these resources in defective forms: houses built on land whose ownership rights are not adequately recorded, unincorporated businesses with undefined liability, industries located where financiers and investors cannot see them. Because the rights of these possessions are not adequately documented, these assets cannot readily be turned into capital, cannot be traded outside of narrow local circles where people know and trust each other, cannot be used as collateral for a loan, and cannot be used as a share against investment."[2] This difference in interest rates

[1] McKinsey [12] provides case study evidence on the impact of informality on firms' productivity. They estimate that the ratio of labor productivity between informal and formal firms is 39% in Turkey and 46% in Brazil.

[2] DeSoto [4], p.5-6. DeSoto [3] estimates that in June/85, informal firms in Lima (Peru) faced a nominal interest rate of 22% per-month, while formal firms paid only 4.9% per month. Straub [18]

induces a higher capital-labor ratio in formal firms.[3] As in Rausch [14], agents with lowest managerial ability become workers and the ones with highest ability become formal managers, with the intermediate group running informal firms. This is because managers with more ability would naturally run larger firms and employ more capital; for this reason they choose to join the formal sector, where they do not face limits on capital deployment and face a lower cost of capital. The marginal firm trades off the cost of paying taxes versus the higher cost of capital and scale limitations of informal firms. As a result, the marginal firm would employ in the informal sector less capital and labor than it would employ if it joined the formal sector. Thus, as in Rausch [14] or Fortin *et al.* [6], a size gap develops. Managers that are slightly more efficient than the manager of the marginal firm employ discretely larger amounts of capital and labor.

Several implications of this model are supported by our empirical analysis. Formalization is positively correlated with the size of firms and measures of the quality of the entrepreneurial input. Even after controlling for our measures of the quality of an entrepreneur, formalization is correlated with a firm's capital-labor ratio or investment per worker. In addition, after controlling for the quality of the entrepreneur, formalization is correlated with higher profits.

The main focus of our theoretical analysis is a model that highlights the role of value added taxes in transmitting informality. The value added tax is a prevalent form of indirect taxation: more than 120 nations had adopted it by 2000.[4] It exploits the idea that collecting value added taxes according to a credit scheme sets in motion a mechanism for the transmission of informality. In the credit or invoice method the value added tax applies to each sale and each establishment receives a credit for the amount of tax paid in the previous stages of the production chain. This credit is then used by the taxpayer against future liabilities with the tax authorities. The credit method is often used in practice because it offers advantages over other collection methods: (1) the tax liability is attached to the transaction and the invoice provides an important documentary evidence; (2) an audit trail is established and there is an important self-monitoring dimension across firms in different stages of the production chain; (3) multiple rates can be used and certain activities may be exempt for social or economic purposes (see [19]). Since purchases from informal suppliers are ineligible for tax credits, an incentive exists for the propagation of informality downstream in

develops a model in which a dual credit system arises in equilibrium.

[3]It is also probable that informal firms face lower labor costs, because their workers avoid some labor taxes. This would induce even larger differences in capital-labor ratio.

[4]See Appendix 4 in Schenk and Oldman [16].

the production chain. A similar mechanism also influences firms upstream in the chain: selling to informal firms increases the likelihood for a firm to be informal. Our empirical analysis shows that, in fact, various measures of formality of suppliers and (and its enforcement) purchasers are correlated with the formality of a firm. Even more interestingly, when we look at sectors where Brazilian firms are not subject to the credit system of value added tax, but instead the value added tax is applied after some stage of production at a rate that is estimated by the State, this chain effect vanishes. To our knowledge, the only study to investigate the informal sector in conjunction with a VAT structure is Emran and Stiglitz [5]. Their focus is on the consequences of informality for a revenue neutral tax reform involving value added and trade taxes.

Our models ignore some possible alternative reasons for informality, such as the fixed cost of complying with regulations or the existence of a minimum-wage. While we could accommodate the existence of a minimum wage, some empirical evidence actually points out that minimum wages may be as binding (if not more) in the informal sector than in the formal segment of the economy in Latin America (see Maloney and Mendez [13]).

Other papers that investigate causes and determinants of informality include Loayza [10], Friedman *et al.* [7], and Junqueira and Monteiro [9] who used the same dataset as we do. These authors have provided evidence of an association between the size of the underground economy and higher taxes, more labor market restrictions, and poorer institutions (bureaucracy, corruption and legal environment). The combination of the models we develop and the Brazilian microdata allows us to add novel insights to this literature.

The remainder of this paper is organized as follows. In the next section we develop a model of a single industry, while in Section 3 we treat the model with two stages of production. Section 4 contains the empirical results obtained using data on informal firms in Brazil and Section 5 concludes.

2 A Model with One Production Stage

We consider a continuum of agents; each characterized by a parameter $\theta \geq 0$ which indicates his quality as an entrepreneur and is distributed according to a probability density function $g(\cdot)$. An entrepreneur chooses between becoming a worker, operating a firm in the formal sector or in the informal sector. If an entrepreneur employs l workers and k units of capital, output equals $y = \theta k^{\alpha} l^{\beta}$, with $\alpha, \beta > 0$ and $\alpha + \beta < 1$.

A formal entrepreneur pays an *ad valorem* tax rate of τ and faces a capital cost of $r_f > 0$ per unit. An informal entrepreneur pays no taxes, but faces a capital cost of $r_i \geq r_f$. All workers are paid the same wage w.

An informal entrepreneur, if detected by the authorities, loses all his profits. The probability of being detected depends monotonically on the size of the firm. Though there are several possibilities for measuring the size of the firm - output, capital stock or labor force - we choose here the size of the capital stock (which we identify in the empirical work as the value of installations), because we imagine the probability of detection as a function of the "visibility" of the firm. We write $p(k)$ for the probability of detection. While in the appendix we discuss a more general form for the function p we will assume here that:

$$p(k) = 0, \text{ if } k \leq k^* \quad (1)$$

$$= 1, \text{ if } k > k^*, \quad (2)$$

that is an informal firm cannot employ more than k^* units of capital, but will not suffer any penalty when $k \leq k^*$.

Hence the profit for an entrepreneur of quality θ that chooses to be informal is given by

$$\Pi_i(\theta, r_i) = \max_{l,k \leq k^*} \{\theta l^\beta k^\alpha - wl - r_i k\}, \quad (3)$$

whereas if he chooses to enter the formal sector profits will be:

$$\Pi_f(\theta, r_f) = \max_{l,k} \{\theta(1-\tau) l^\beta k^\alpha - wl - r_f k\} \quad (4)$$

The capital-labor ratios of formal firms or informal firms that are unconstrained are proportional to the relative prices between labor and capital and independent of the entrepreneur's ability. Since $r_i \geq r_f$, the unconstrained informal firms have a lower capital-labor ratio than the formal firms. In addition, the constrained informal firms have a lower capital-labor ratio than the unconstrained informal firms. Hence the capital-labor ratios of informal firms are lower than that of the formal firms, the difference being bigger the larger the difference in capital costs between informal and formal firms $(r_i - r_f)$ is. In Section 4 we provide evidence in favor of the predicted difference in capital-labor ratios between formal and informal firms.

The usual properties of profit functions guarantee that both Π_i and Π_f are convex functions of θ, w and the respective cost of capital, r_i and r_f. Using the first order conditions and the envelope theorem one obtains :

$$\frac{d\Pi_f}{d\theta}(\theta) = \frac{\beta^{\beta/(1-\alpha-\beta)} \alpha^{\alpha/(1-\alpha-\beta)} (1-\tau)^{1/(1-\alpha-\beta)}}{r_f^{\alpha/(1-\alpha-\beta)} \times w^{\beta/(1-\alpha-\beta)}} \theta^{(\alpha+\beta)/(1-\alpha-\beta)}, \quad (5)$$

and that, for informal firms that are not constrained:

$$\frac{d\Pi_i}{d\theta}(\theta) = \frac{\beta^{\beta/(1-\alpha-\beta)}\alpha^{\alpha/(1-\alpha-\beta)}}{r_i^{\alpha/(1-\alpha-\beta)} \times w^{\beta/(1-\alpha-\beta)}}\theta^{(\alpha+\beta)/(1-\alpha-\beta)}, \tag{6}$$

A comparison of expressions (5) and (6) yields that, if $1-\tau \geq (\frac{r_f}{r_i})^\alpha$, taxes are too low with respect to the capital cost wedge and ever entrepreneur prefers to be formal. Since we are interested in the informal sector we assume from now on that $1-\tau < (\frac{r_f}{r_i})^\alpha$. In this case, every entrepreneur θ for which the optimal choice in the informal sector is unconstrained will prefer to be informal. Furthermore, for θ large enough the capital restriction is biding, and a simple calculation using the inequality, $1-\tau < (\frac{r_f}{r_i})^\alpha$, shows that $\frac{\partial \Pi_i(\theta)}{\partial \theta} - \frac{\partial \Pi_f(\theta)}{\partial \theta}$ decreases with θ. As a result, there exists a unique $\overline{\theta}$ such that $\Pi_i(\theta) < \Pi_f(\theta)$ if and only if $\theta > \overline{\theta}$.

Each agent also has the choice of becoming a worker and receive the market wage w. Hence the occupational choice cutoff points are implicitly defined by:

$$\Pi_f(\overline{\theta}) = \Pi_i(\overline{\theta}) \tag{7}$$

$$\max\{\Pi_i(\hat{\theta}), \Pi_f(\hat{\theta})\} = w \tag{8}$$

and optimal choices are:

$$\begin{aligned} \theta \leq \hat{\theta} &\implies \text{Worker;} \\ \theta \in (\hat{\theta}, \overline{\theta}] &\implies \text{Informal entrepreneur;} \\ \theta > \max\{\overline{\theta}, \hat{\theta}\} &\implies \text{Formal entrepreneur.} \end{aligned}$$

Since $\Pi_i(0) = 0$ and $\Pi_f(0) = 0$, $\hat{\theta} > 0$, whenever $w > 0$. However, if $\overline{\theta} < \hat{\theta}$ then no entrepreneur would choose informality. In any case, equilibrium in the labor market requires w to satisfy:

$$\underbrace{\int_{\hat{\theta}(w)}^{\max\{\overline{\theta}(w),\hat{\theta}(w)\}} l_i(\theta; w)g(\theta)d\theta + \int_{\max\{\overline{\theta}(w),\hat{\theta}(w)\}}^{\infty} l_f(\theta; w)g(\theta)d\theta}_{\text{Demand for Labor}} = \underbrace{\int_0^{\hat{\theta}(w)} g(\theta)d\theta}_{\text{Supply of Labor}}$$

where the arguments remind the reader of the dependence of the cutoffs and labor demand on the level of wages.

The existence of an equilibrium level of wages is straightforward. Also if k^* is large enough then $\overline{\theta} < \hat{\theta}$. Furthermore if θ is sufficiently large, an entrepreneur of quality θ would choose the formal sector.

Another implication of this model is the existence of a discontinuity in the level of capital and labor employed at levels of productivity around $\overline{\theta}$. This discontinuity

follows since an entrepreneur with ability just below $\overline{\theta}$ chooses the informal sector and employs exactly k^* units of capital, although the marginal product of capital exceeds his cost of capital. At a level just above $\overline{\theta}$, an entrepreneur chooses the formal sector and since he is now unconstrained, he would choose a level $k >> k^*$. The jump in the amount of capital employed also induces a jump in the marginal product of labor and hence employment also exhibits a discontinuity at $\overline{\theta}$. Thus our model predicts a "gap" in the quality distributions of firms.

3 A Model with Two Production Stages

In this section we introduce a model with two stages of production. Our goal is to illustrate the transmission of informality across sectors which results from the use of the value added tax. In Section 4 we document that this mechanism is relevant for the generation of informality in Brazil.

There are two stages of production: "upstream" and "downstream". All individuals in this model are entrepreneurs and, for simplicity, we assume that they are specialized in one of the stages. Each entrepreneur in the upstream sector is characterized by his ability $\theta_u > 0$. The density of θ_u is $g_u(\cdot)$. An entrepreneur of ability θ_u can produce θ_u units of the intermediate good in the formal sector, but only $\min(\overline{y}, \theta_u)$, where $\overline{y} > 0$, if in the informal sector.

The downstream entrepreneurs are characterized by an ability parameter θ_d with density $g_d(\cdot)$. An agent with ability θ_d, if in the formal sector, produces $\theta_d x^\alpha$ units of the formal good using x units of the intermediate good. However if in the informal sector he faces a limit on the quantity of input that can be used and the production function becomes $\theta_d \min(\overline{x}, x)^\alpha$, where $\overline{x} > 0$.

We assume that g_u and g_d are continuous and positive on their support. In addition we assume that there exists $\theta_u < \overline{y}$ for which $g_u(\theta_u) > 0$, and that $g_d(\theta_d) > 0$ for $\theta_d > 0$.

The final good is tradeable and has an exogenous price q. Firms in the formal sector pay an ad-valorem tax rate of τ and we write $\pi = 1 - \tau$. The value added tax is levied by the credit method: the tax rate applies to each sale and each establishment receives a credit for the amount of tax paid in the previous stages of production. Because of the tax credit, the prices paid for informal and formal goods may be distinct and we let p_f be the price of the intermediate good in the formal sector and p_i in the informal sector.

We write $\Pi_f^u(\theta_u) = \pi p_f \theta_u$ for the profit of an upstream firm with manager of

quality θ_u if it produces in the formal sector and $\Pi_i^u(\theta_u) = p_i \min\{\theta_u, \overline{y}\}$ if it produces in the informal sector. Downstream firms face a slightly more complicated problem, since they must also choose which intermediate good (formal or informal) to purchase.

Write

$$\Pi_f^d(\theta_d) = \max\{\max_x[\pi(q\theta_d x^\alpha - p_f x)], \max_x[q\pi\theta_d x^\alpha - p_i x]\}, \tag{9}$$

for the profit of a downstream firm with a manager with ability θ_d that chooses to operate in the formal sector. In an analogous manner, write

$$\Pi_i^d(\theta_d) = \max\{\max_x[q\theta_d \min(x, \overline{x})^\alpha - p_f x], \max_x[q\theta_d \min(x, \overline{x})^\alpha - p_i x]\}, \tag{10}$$

for the profit of a downstream firm with a manager of ability θ_d that chooses to operate in the informal sector.

If an informal entrepreneur of ability θ_d buys the input at a price p then he demands:

$$x_i(\theta_d, p, q) = \min\left(\overline{x}, \left(\frac{q\alpha\theta_d}{p}\right)^{1/(1-\alpha)}\right). \tag{11}$$

In turn, a formal entrepreneur demands, if he buys from the formal sector at a unit price p:

$$x_f(\theta_d, p, q) = \left(\frac{q\alpha\theta_d}{p}\right)^{1/(1-\alpha)}, \tag{12}$$

while if he buys from the informal sector he demands $x_f(\theta_d, \frac{p}{\pi})$, since the tax credit does not apply.

Similarly to the model with one stage, the demand for the intermediate input, as the following proposition shows, will exhibit a large enough "discontinuity".

Proposition 1 *If $\Pi_f^d(\theta_d) > \Pi_i^d(\theta_d)$ then the optimal choice of the firm with manager of quality θ_d, $x_f(\theta_d, p, q)$, where $p = p_f$ if the firm optimal choice is to buy the formal good and $p = \frac{p_i}{\pi}$ if the firm optimal choice is to buy the informal good, satisfies*

$$x_f(\theta_d, p, q) \geq \frac{\overline{x}}{\pi} > \overline{x} \geq x_i(\theta_d, p, q),$$

for any θ.

Proof: Suppose first that it is optimal for the firm with manager of quality θ_d to buy the formal good. If $\pi x_f(\theta_d, p_f, q) < \overline{x}$, since

$$q\theta_d(\pi x_f(\theta_d, p_f, q))^\alpha - \pi p_f x_f(\theta_d, p_f, q) \geq \pi(q\theta_d x_f^\alpha(\theta_d, p_f, q) - \pi p_f x_f(\theta_d, p_f, q)),$$

the firm would prefer to be in the informal sector and buy $\pi x_f(\theta_d, p_f, q)$ of formal inputs. If the firm bought the informal good and $\pi x_f(\theta_d, \frac{p_i}{\pi}, q) < \overline{x}$, since

$$q\theta_d(\pi x_f(\theta_d, \frac{p_i}{\pi}, q))^\alpha - \pi p_i x_f(\theta_d, \frac{p_i}{\pi}, q) \geq \pi q\theta_d x_f^\alpha(\theta_d, \frac{p_i}{\pi}, q) - \pi p_i x_f(\theta_d, \frac{p_i}{\pi}, q),$$

the firm would prefer to be in the informal sector and buy $\pi x_f(\theta_d, \frac{p_i}{\pi}, q)$ of informal inputs.

■

The next proposition shows the existence of cutoff points for each stage, $\overline{\theta}_u(p_i, p_f, q)$ and $\overline{\theta}_d(p_i, p_f, q)$ such that all managers with ability below the cutoff prefer informality and all those with ability above the cut-off points prefer to join the formal sector.

Proposition 2 *Suppose p_f, p_i and q are positive.*

(i) If $\theta_u < \overline{\theta}_u(p_i, p_f, q) = \frac{p_i\overline{y}}{\pi p_f}$ then $\Pi_i^u(\theta_u) > \Pi_f^u(\theta_u)$, and if $\theta_u > \overline{\theta}_u(p_i, p_f, q) = \frac{p_i\overline{y}}{\pi p_f}$ then $\Pi_i^u(\theta_u) < \Pi_f^u(\theta_u)$.

(ii) There exists a $\overline{\theta}_d(p_i, p_f, q)$ such that if $\theta_d < \overline{\theta}_d(p_i, p_f, q)$ then $\Pi_i^d(\theta_d) > \Pi_f^d(\theta_d)$ and if $\theta_d > \overline{\theta}_d(p_i, p_f, q)$ then $\Pi_i^d(\theta_d) < \Pi_f^d(\theta_u)$.

Proof: The function $\Pi_f^u(\theta_u)$ is linear, while the $\Pi_i^u(\theta_u)$ is linear if $\theta_u \leq \overline{y}$ and constant for $\theta_u \geq \overline{y}$. Hence there exists at most one positive $\overline{\theta}_u(p_i, p_f)$ where they cross. Since they cross at $\frac{p_i\overline{y}}{\pi p_f}$, the result follows.

To show that (ii) holds note that θ enters the definition of the profit function exactly as an output price and hence, from the properties of profit functions with respect to output prices, we know that if $q > 0$ the function $\Pi_f^d(\theta_d)$ is strongly convex. Furthermore, the function $\Pi_f^u(\theta_d)$ is convex and, since supply functions of firms must slope up, if the choice, conditional on informality, of a firm of ability θ satisfies $x_i(\theta) = \overline{x}$ then the optimal choice conditional on informality, $x_i(\theta_d) = \overline{x}$ for $\theta_d \geq \theta$, and as a consequence, $\Pi_i^d(\theta_d)$ is linear for $\theta_d \geq \theta$. In addition, whenever $x_i(\theta_d) < \overline{x}$, the informal firm's constraint is not binding. In this case, when $p_f > p_i$

$$\Pi_i^d = \varphi(p_i) > \varphi(p_f)$$

where $\varphi(p) = [\alpha^{\alpha/(1-\alpha)} - \alpha^{1/(1-\alpha)}]\left(\frac{q\theta_d}{p^\alpha}\right)^{1/(1-\alpha)}$. since

$$\Pi_f^d = \max\{\pi\varphi(p_f), \pi^{1/(1-\alpha)}\varphi(p_i)\}$$

then $\Pi_i^d(\theta_d) > \Pi_f^d(\theta_d)$. When $p_f \leq p_i$,

$$\Pi_i^d = \varphi(p_f) \geq \varphi(p_i)$$

and the definition of $\Pi_f^d(\theta_d)$ guarantees that $\Pi_i^d(\theta_d) > \Pi_f^d(\theta_d)$. In summary, whenever $x_i(\theta_d) < \overline{x}$, one obtains that $\Pi_i^d(\theta_d) > \Pi_f^d(\theta_d)$. Hence there is exactly one crossing point, $\overline{\theta}_u(p_i, p_f, q)$.

■

We now derive aggregate demand and supply of the intermediate good in the formal and informal sectors as a function of prevailing prices. Since we are interested in equilibrium prices we may restrict the range of prices to $\pi p_f \leq p_i \leq p_f$. In fact, if $\pi p_f > p_i$ profit maximization and equations (9) and (10) imply that both formal and informal entrepreneurs downstream would buy from informal upstream firms. However, every upstream entrepreneur will prefer to produce in the formal sector. Similarly, if $p_i > p_f$ every downstream entrepreneur would prefer to buy from formal firms. However, small θ_u agents would prefer to produce informally. In addition, the homogeneity of the system allows us to choose $q = 1$ (and hence we omit q as a function argument in what follows).

Because of the possibility of indifference, we have supply and demand correspondences as opposed to functions. We will write $S(p_i, p_f)$ for the set of possible aggregate supply vectors $(s_i(p_i, p_f), s_f(p_i, p_f))$ obtained from the choices of profit maximizing entrepreneurs in the upstream stage. If $p_i \neq \pi p_f$ the set $S(p_i, p_f)$ contains a single vector (s_i, s_f) given by

$$s_i = \int_0^{\frac{p_i\overline{y}}{\pi p_f}} \max\{\theta, \overline{y}\} g_u(\theta) d\theta \tag{13}$$

$$s_f = \int_{\frac{p_i\overline{y}}{\pi p_f}}^{\infty} \theta g_u(\theta) d\theta \tag{14}$$

If $\pi p_f = p_i = 0$ then $S(p_i, p_f) = \{0\}$. Finally when $\pi p_f = p_i \neq 0$ a point $(s_i, s_f) \in$

$S(p_i, p_f)$ if there exists a $\overline{\theta}_u \leq \overline{y}$ such that:[5]

$$s_i = \int_0^{\overline{\theta}_u} \theta g_u(\theta) d\theta \tag{15}$$

$$s_f = \int_{\overline{\theta}_u}^{\infty} \theta g_u(\theta) d\theta \tag{16}$$

Since we fixed $q = 1$ we write $X(p_i, p_f)$ for the set of possible aggregate demand vectors $(x_i(p_i, p_f), x_f(p_i, p_f))$ obtained from the choices of profit maximizing entrepreneurs in the downstream stage.

When $\pi p_f = p_i$ formal firms are indifferent between buying the formal or informal input, but informal firms prefer buying from informal firms. Hence we can allocate all formal firms with managers below a certain threshold to buying in the informal sector with the complement interval assigned to purchase in the formal sector.[6] In this case, a point $(x_i, x_f) \in X(p_i, p_f)$ if there exists a $\gamma \geq \overline{\theta}_d(p_i, p_f, 1)$ such that:

$$x_i = \int_0^{\overline{\theta}_d(p_i,p_f,1)} x_i(\theta_d, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i,p_f,1)}^{\gamma} x_f(\theta_d, \frac{p_i}{\pi}) g_d(\theta) d\theta \tag{17}$$

$$x_f = \int_{\gamma}^{\infty} x_f(\theta_d, p_f) g_d(\theta) d\theta \tag{18}$$

If $\pi p_f < p_i < p_f$ formal (informal) firms prefer to buy from formal (informal) firms. In this case, a point $(x_i, x_f) \in X(p_i, p_f)$ if :

$$x_i = \int_0^{\overline{\theta}_d(p_i,p_f,1)} x_i(\theta_d, p_i) g_d(\theta) d\theta \tag{19}$$

$$x_f = \int_{\overline{\theta}_d(p_i,p_f,1)}^{\infty} x_f(\theta_d, p_f) g_d(\theta) d\theta \tag{20}$$

If $p_f = p_i$ informal firms are indifferent, but formal firms prefer buying from formal firms. Hence we may assign informal firms arbitrarily to buying formal or informal inputs. In this case, a point $(x_i, x_f) \in X(p_i, p_f)$ if there exists $\gamma \leq \overline{\theta}_d(p_i, p_f, 1)$ such that:

$$x_i = \int_0^{\gamma} x_i(\theta_d, p_i) g_d(\theta) d\theta \tag{21}$$

$$x_f = \int_{\gamma}^{\overline{\theta}_d(p_i,p_i,1)} x_i(\theta_d, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i,p_i,1)}^{\infty} x_f(\theta_d, p_i) g_d(\theta) d\theta \tag{22}$$

[5] In principle we could assign any subset of the entrepreneurs with productivity below $\overline{y}$ to the informal sector, but there is always an interval containing the origin of these entrepreneurs that would produce exactly the same aggregate output.

[6] As before, these assignments can reproduce the demands realized by any arbitrary assignment of firms to each sector.

An equilibrium is a vector $(p_i, p_f, 1)$ such that $\exists z \in X(p_i, p_f) \bigcap S(p_i, p_f)$.

We will decompose the proof of the existence of an equilibrium price in two steps. First we will set $p_i = \mu p_f$ with $\pi \leq \mu \leq 1$. For each μ we will show that there exists a unique $p_i(\mu)$ such that if $(p_i, p_f) = (p_i(\mu), \frac{p_i(\mu)}{\mu})$ then the sum of aggregate supply of the formal and informal intermediate goods equals the sum of aggregate demands. We then show that there exits a unique μ^* such that $(p_i(\mu^*), \frac{p_i(\mu^*)}{\mu^*}, 1)$ is an equilibrium.

We will use a preliminary result:

Lemma 1 *If $\pi p_f < p_i < p_f$ then $\overline{\theta}_d(p_i, p_f, 1)$ decreases with p_i and it increases with p_f. Further, if $\pi \leq \mu \leq 1$ then, $\overline{\theta}_d(p_i, \frac{p_i}{\mu}, 1)$ increases with p_i.*

Proof: If $\pi p_f < p_i \leq p_f$ formal firms prefer to buy the formal good. Hence from equations (9) and the envelope theorem,

$$\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} = -\pi x_f(\theta_d, p_f) \tag{23}$$

Similarly, if $\pi p_f \leq p_i < p_f$, informal firms prefer to buy the informal good, and in an analogous fashion, using equation (10)

$$\frac{\partial \Pi_i^d(\theta_d)}{\partial p_i} = -x_i(\theta_d, p_i) \tag{24}$$

This establishes the first part of the lemma since increasing p_i reduces profits for informal firms and increasing p_f reduces profits for formal firms.

In order to sign the change in $\overline{\theta}_d(p_i, \frac{p_i}{\mu}, 1)$ we must establish the sign of:

$$\frac{1}{\mu}\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} - \frac{\partial \Pi_i^d(\theta_d)}{\partial p_i}. \tag{25}$$

for the marginal firm. If this is negative, the difference in profits in the formal and informal sectors for the marginal firm decreases and more firms will become informal. If $\pi p_i < p_f < p_i$, from equations (23) and (24)

$$\frac{1}{\mu}\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} - \frac{\partial \Pi_i^d(\theta_d)}{\partial p_i} = -\frac{\pi}{\mu} x_f(\theta_d, \frac{p_i}{\mu}) + x_i(\theta_d, p_i). \tag{26}$$

The marginal informal firm buys exactly $\overline{x}$. Hence, from Proposition 1

$$-\frac{\pi}{\mu} x_f(\overline{\theta}_d, \frac{p_i}{\mu}) + x_i(\overline{\theta}_d, p_i) \leq -\frac{\overline{x}}{\mu} + \overline{x} \leq 0$$

since we assume that $\mu \leq 1$ and the second part of the lemma follows.

The derivative $\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f}$ ($\frac{\partial \Pi_i^d(\theta_d)}{\partial p_i}$) is not well defined when $p_i = \pi p_f$ (resp. $p_i = p_f$), but it is easy to see that, in this case, the change in profit difference between formality and informality for the marginal firm still equals $-\frac{\pi}{\mu} x_f(\overline{\theta}_d, \frac{p_i}{\mu}) + x_i(\overline{\theta}_d, p_i)$.

■

We now turn to our equilibrium analysis. For $\mu = \pi$ ($p_i = \pi p_f$) the sum of the aggregate supply always equals

$$\int_0^\infty \theta g_u(\theta) d\theta. \tag{27}$$

On the other hand, the sum of aggregate demands always equals

$$\int_0^{\overline{\theta}_d(p_i, \frac{p_i}{\pi}, 1)} x_i(\theta, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i, \frac{p_i}{\pi}, 1)}^\infty x_f(\theta, \frac{p_i}{\pi}) g_d(\theta) d\theta \tag{28}$$

It is easy to check that this last expression goes to zero as $p_i \to \infty$ and to ∞ as $p_i \to 0$. Furthermore, since demand of any type decreases with the price of the input, and, from Proposition 1 $x_f(\overline{\theta}_d, p_i/\pi) > x_i(\overline{\theta}_d, p_i)$, using the Lemma above it is immediate that aggregate demand is monotonically decreasing with p_i. Hence there exists a unique $p_i(\pi)$ for which the sum of supplies equal the sum of demands.

For $\pi < \mu \leq 1$, using expressions (13) and (14) we obtain that the sum of the aggregate supplies is:

$$\int_0^{\frac{\mu \overline{y}}{\pi}} \max\{\theta, \overline{y}\} g_u(\theta) d\theta + \int_{\frac{\mu \overline{y}}{\pi}}^\infty \theta g_u(\theta) d\theta. \tag{29}$$

On the other hand, using equations (19) and (20), the sum of the aggregate demands equals:

$$\int_0^{\overline{\theta}_d(p_i, \frac{p_i}{\mu}, 1)} x_i(\theta_d, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i, p_i/\mu, 1)}^\infty x_f(\theta_d, \frac{p_i}{\mu}) g_d(\theta) d\theta. \tag{30}$$

Just as before, the result in the Lemma insures the monotonicity properties that yield the existence of a unique $p_i(\mu)$ that equates the sum of aggregate demands with that of aggregate supplies.

Notice that an increase in μ always decreases aggregate supply since it causes some firms in the upstream sector to switch from formal to informal. In addition, an increase in μ increases the demand by formal firms at each p_i and causes some firms to switch from informal to formal in the downstream sector. Thus, at each p_i, aggregate demand goes up. Hence $p_i(\mu)$ increases with μ.

The supply of the informal sector when $p_i = \pi p_f$ is some amount in the interval $[0, \int_0^{\overline{y}} \theta g_u(\theta) d\theta]$. The demand is some number in the interval $[\int_0^{\overline{\theta}_d(p_i, p_i/\pi, 1)} x_i(\theta_d, p_i) g_d(\theta) d\theta, \int_0^{\overline{\theta}_d(p_i, p_i/\pi, 1)} x_i(\theta_d, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i, p_i/\pi, 1)}^{\infty} x_f(\theta_d, p_i/\pi) g_d(\theta) d\theta]$. If these intervals overlap, at $p_i = p_i(\pi)/\pi$ then $(p_i(\pi), p_i(\pi)/\pi, 1)$ is an equilibrium. This will happen whenever the tolerance for informality in the upstream sector $(\overline{y})$ is high enough.

If these intervals do not overlap, notice that the informal supply of the intermediate good must necessarily go up with μ. On the other hand, the informal demand at $(p_i(\mu), \frac{p_i(\mu)}{\mu})$ will go down since $p_i(\mu)$ goes up and the relative price of the formal good goes down. At $\mu = 1$, the supply of the informal good is $\int_0^{\frac{\overline{y}}{\pi}} \max\{\theta, \overline{y}\} g_u(\theta) d\theta$ whereas the demand is any number in the interval $[0, \int_0^{\overline{\theta}_d(p_i, p_f, 1)} x_i(\theta_d, p_f) g_d(\theta) d\theta]$. Hence there always exists a unique μ^* such that $(p_i(\mu^*), \mu^* p_i(\mu^*), 1)$ is an equilibrium.

Remark: The equilibrium will be at $p_i = p_f$ whenever the tolerance for informality in the upstream sector, $\overline{y}$, is small enough.

3.1 Comparative statics

We now argue that, if in equilibrium $\pi p_f < p_i < p_f$, an increase in the tolerance towards informality in the upstream (downstream) sector will increase informality in downstream (upstream) sector.

To show that, again fix a μ and set $p_i = \mu p_f$. For each μ let $\tilde{p}(\mu, \overline{y})$ be such that $(p_i, p_f) = (\tilde{p}(\mu, \overline{y}), \frac{\tilde{p}(\mu, \overline{y})}{\mu})$ clears the market for the formal intermediate good. Notice that, once we fix μ the supply of the formal good $\int_{\frac{\mu \overline{y}}{\pi}}^{\infty} \theta g_u(\theta) d\theta$ is independent of p_i. The demand $\int_{\overline{\theta}_d(p_i, p_i/\mu, 1)}^{\infty} x_f(\theta_d, \frac{p_i}{\mu}) g_d(\theta) d\theta$, goes to ∞ as $p_i \to 0$, goes to zero as $p_i \to \infty$ and as a consequence of the Lemma, goes down with p_i. Hence $\tilde{p}(\mu, \overline{y})$ is well defined. Furthermore the supply goes down with μ whereas, for each p_i , again using the Lemma, the demand goes up with μ. Hence $\tilde{p}(\mu, \overline{y})$ increases with μ. Now suppose $(\overline{y})$ goes up. Then, at each μ, the supply of the formal goods decreases and the demand at each p_i is unchanged. Thus $\tilde{p}(\mu, \overline{y})$ increases with $\overline{y}$. We already know that an equilibrium exists. Write $\mu^*(\overline{y})$ for the equilibrium relative price between the informal and the formal good and $p^*(\overline{y})$ for the price of the informal good in equilibrium. Now suppose $\overline{y}$ increases to $\overline{y}'$, but in equilibrium $\theta_d(p_i^*(\overline{y}), p_i^*(\overline{y})/\mu^*(\overline{y}), 1) \geq \theta_d(p_i^*(\overline{y}'), p_i^*(\overline{y}')/\mu^*(\overline{y}'), 1)$, that is the measure of downstream informal firms does not go up. Since $p_i^*(\overline{y}') = \tilde{p}(\mu^*(\overline{y}'), \overline{y}')$, and, at a given μ,

$\tilde{p}(\mu, \overline{y}') > \tilde{p}(\mu, \overline{y})$, Lemma 1 requires that $\mu^*(\overline{y}') > \mu^*(\overline{y})$. Hence

$$\tilde{p}(\mu^*(\overline{y}), \overline{y}) < \tilde{p}(\mu^*(\overline{y}'), \overline{y}) < \tilde{p}(\mu^*(\overline{y}'), \overline{y}')$$

Thus, in equilibrium, the demand for the informal intermediate good decreases, since the number of informal downstream firms did not go up and the price they face goes up. Since $\mu^*(\overline{y}') > \mu^*(\overline{y})$, the equilibrium supply of the informal good goes up, a contradiction since in equilibrium demand equals supply. Hence, the number of informal downstream firms must go up when $\overline{y}$ increases.

A similar proof establishes that informality must go up in the upstream sector when $\overline{x}$, the tolerance in the downstream sector, is increased. These two implications will be examined in the empirical section.

4 Empirical Application

In this section we explore the implications of the theoretical framework laid out in the previous section using a dataset on the informal sector in Brazil.

4.1 Data

Our principal data source is the ECINF survey (Pesquisa de Economia Informal Urbana) on informal firms realized by the Brazilian statistics bureau (IBGE). We have used the 2003 edition of that survey,[7] from which we obtain information on 51,485 entrepreneurs in urban regions from all states in the Brazilian federation. The focus is on units with five or less employees and the sampling strategy uses the 2000 demographic census as a frame. Before the survey, preliminary interviews screened households for the presence of at least one entrepreneur with a business employing five or less people. Households without such an entrepreneur were not included in the frame for the survey. The sampling was designed in two stages: in each state (of a total of 27) the primary sampling units (urban sectors) are stratified geographically in three strata (capital, other urban sectors in the capital metropolitan area and remaining urban sectors). In a second step, the primary sampling units were stratified according to levels of income within the geographical stratum. Urban sectors were then randomly selected with probability proportional to the number of households in the sector in 2000. From each selected urban sector a total of 16 households was

[7]The survey was also conducted in 1997. We have reproduced the statistical exercises displayed in this section for that edition and the results are qualitatively unchanged.

then randomly selected for interviews. The survey was collected in October 2003[8]. Since the focus of the survey and the definition of informal economic unit adopted by the Brazilian statistics bureau were those firms with less than five employees and not those in irregular situation, we do believe answers were truthful even when individuals were inquired about their status with the Brazilian tax authorities. Interviewees were made aware that information collected for the survey was confidential and would only be utilized for statistical purposes.[9]

Given our second model, the ideal dataset would comprise information on the production chain in which a firm is immersed. Although the ECINF contains certain characteristics of a firm's clientele (whether they were predominantly large or small companies, persons or governmental institutions), the information is very limited. To complement these data we used the input-output matrix information available from the Brazilian Statistics Bureau (IBGE). We computed inter-sectoral technical input coefficients and measures of output sectoral destination using the 2003 make and use tables in the Brazilian national accounts.[10]

4.2 Description of Variables

In the statistical exercises performed, we have filtered out non-owners and individuals less than 15 years-old. Unavailability of education and gender variables also restricts our sample to circa 45,000 observations. It should also be mentioned that, even though the focus of the survey is on five or less employees, a few units (less than 3%) employ more than five people due to the existing lag between the screening and interviewing stages of the survey.

Table 1 summarizes the main variables used in the study. The first is indicative of formalization. It is a dummy variable equal to one if the firm is registered with the Brazilian tax authorities.[11] For firms in economic sectors that qualify for

[8]For more information on the sampling strategies employed, see Almeida and Bianchini [1].

[9]A disclaimer appears on top of the questionnaire stating that such information is confidential and protected by Law 5534 14/11/68.

[10]Tables 1 and 2 under "Tabelas de Recursos e Usos" available under National Accounts on http://www.ibge.gov.br for 2003. The information is at current 2003 prices (rather than the alternative: previous year monetary units). The construction of technical coefficients follows the European System of Integrated Economic Accounts (ESA) specifications (see ten Raa [20]).

[11]The tax registry is the Cadastro Nacional de Pessoas Jurídicas, which replaced the previous system, the Cadastro Geral de Contribuintes (CGC), in 1998. This variable is the most representative of formalization for our purposes, but we have nonetheless experimented with the existence of legal constitution for a firm and obtained virtually identical results. This is not surprising, since the latter

tax substitution (see subsection 4.8), *taxsub* takes the value one. The next two variables are dummies for firms that sell their products mainly to large firms (*largecl*) or small firms (*smallcl*). The variable *govcl* is a dummy for a firm that sells mainly to governmental institutions. Other alternatives are persons or ignored. The variable *outsidehouse* is a dummy that equals one when the activity is performed outside the home. The variable *education* is a categorical variable with values depicted on Table 2. The number of employees variable (*n_employee*) includes the owner. The variable *homeown_num* is a measure of wealth and is zero for those that do not own a home and positive otherwise. When an entrepreneur is a homeowner, the variable displays the number of rooms in the house. The variables *loginv* and *loginst* measure the logarithm of investments and capital installations in October/2003 (R$ 1,000). The variable profit equals revenue minus expenses in October/2003 (also in R$ 1,000).

Each firm in the sample is classified into economic activities following the CNAE (Classificação Nacional de Atividades Econômicas) categorization system.[12] Having obtained technical coefficients as well as sectoral output allocation coefficients from the National Accounts System (NAS) (by NAS sector classification) we are then able to assign to each activity in the survey a vector with those coefficients. Since the survey and National Accounts use different classification schemes we had to match the activities in both systems. Typically a CNAE activity corresponds to a single NAS sector, but there are a few exceptions. Whenever such a multiple match occurred, we assigned to a CNAE sector the weighted averages (by NAS sector production value) of the coefficients in the corresponding NAS sectors. The ECINF survey also has its own aggregate sectoral characterization, displayed on Table The ECINF survey also has its own sectoral characterization, displayed on Table 4.

These coefficients were used for each observation as a vector measure of sectoral allocation of output and sectoral input assignment by a firm, but were also utilized in the construction of a few indices we employed in the analysis. The last two variables on Table 1 are measures of formalization enforcement for suppliers and customers[13] and were constructed in this manner. To obtain these variables we used information available from the Brazilian Ministry of Labor on the number of firms visited in a given economic sector and state during 2002 to monitor labor regulation compliance by the unit under scrutiny. We normalized the number of visits in each state and sector

is a prerequisite for a firm to register with the tax authorities. The correlation between the two measures of informality is 0.9837.

[12]The Brazilian Bureau of Statistics website (http://www.ibge.gov.br) provides a description of this classification system as well as various matching tables with other classification schemes.

[13]The enforcement information was also used by Almeida and Carneiro [2].

by the number of firms in that state and sector provided by the Brazilian Statistics Bureau (IBGE) (through the Cadastro Central de Empresas).[14] For each observation we generated an index of supplier formalization enforcement as a weighted average of these variables where the weights were the sectoral input demand coefficients. A measure of client formalization enforcement was obtained analogously where the weights utilized were sectoral output allocation coefficients.

The correlation matrix for our variables is presented in Table 3.

4.3 Probability of Formalization

Table 5 contains probit estimates for the formalization variable *taxreg* using three different sets of controls. The signs obtained for each one of the regressors are to be expected. The coefficient of the variable "working outside the home" is positive in the regression for each of the formalization proxies. In accordance to the first model, the coefficients are also positive for variables related to the size of the firm (number of employees and revenue), credit (bankloan), or the quality of the entrepreneurial input (education, age or having no additional job). Since women in Brazil are more likely to have substantial household duties, the sign on the gender variable may also be related to entrepreneurial input. The coefficients on all these variables are statistically significant.

The three sets of estimates differ in the sectoral controls utilized. In the first battery of estimates, we utilize dummies for state and sector (according to the specification on Table 4). The second and third estimation results in the table use the derived input and output coefficients obtained from the Brazilian National Accounts. The National Accounts System in Brazil categorizes economic activity in forty-two sectors. The use table in the NAS allows one to obtain how much in a given year a certain sector required in terms of input from another given sector in the economy. This can be used to obtain the technical coefficients for each NAS sector (see footnote 10). We are able to identify the sector (according to the NAS) for each firm in the ECINF survey through equivalence tables among different classification schemes available from the Brazilian Statistics Bureau. For each observation we then assemble a vector of input requirements $(tc_j)_{j=1,...,42}$ and these are the controls utilized in the second column. The make table in the National Accounts provides the quantity of output destined to each sector of the economy (plus final demand, which comprises inventory,

[14] Similar calculations were also performed using as normalizing variable the number of employed in the state-sector. Results are qualitatively the same.

family consumption, exports and public administration). We used this information to assemble a vector of sectoral allocation for each monetary unit of output generated for each activity in our sample (and hence each observation in our sample): $(tc_j)_{j=1,...,42}$. These are the controls presented in the third set of estimations presented in the table. All three sets of results include state dummies.

4.4 Investment, Installations per Worker and Profits

Since an entrepreneur's true ability is not observable, it makes sense to measure the effect of formalization, even after controlling for characteristics of the manager and the firm. The model predicts that informal firms would choose a lower capital-labor ratio, and Table 6 depicts the effect of formalization on investments and installations per worker. The coefficient has the right sign and is statistically significant. Formalization has an economic significance of 0.33 for investments per worker and 0.52 for installations per worker regardless of the measure of formalization[15]. In other words, formalization is associated with an increase in investments (installations) per worker of 0.33 (0.52) standard deviations.

In a similar way we examined the impact of formalization on profits. The results are summarized in the same table. Again, after controlling for characteristics of the manager and the firm, formalization has a statistically significantly positive effect on profits. Formalization is associated with an increase in monthly profits of close to 700 Reais.[16]

4.5 Regression Regimes

In our regressions we used education as one of the measures of an entrepreneur's quality θ. Our model predicts a "gap" in the size distribution of firms as a function of the quality of the entrepreneur. Intuitively, informal entrepreneurs adjust their labor input less than formal entrepreneurs do, because of the constraints on capital they face. This suggests that employment reacts less to changes in the ability of the manager for an informal firm than to similar changes for a formal firm.

Table 7 exhibits OLS estimates of the number of employees on a series of controls and using education of the owner as the observable productivity enhancing feature. The coefficient of the interaction of education and formality is positive and

[15]For dummy variables, the economic significance is the regression coefficient divided by the standard deviation of the dependent variable.

[16]The figure is for October 2003, when 1 US dollar was worth 2.87 Reais.

significant. Since the number of employees is an integer, we also ran an ordered probit and a Poisson[17] regression, but the results are very similar.

4.6 Chain Effects on Formalization

One initial approach to investigate the existence of cross-industry effects of formalization was to employ a characterization of a firm's clientele as presented in the ECINF survey. Interviwees were asked to declare whether sales were principally to governmental institutions, large firms (more than five employees), small firms (five employees or less) or persons. Sales to governmental institutions, large firms and small firms tend to increase the probability of formalization with the largest effect being associated with governmental organizations and the lowest with small enterprises as depicted on Table 8. Since one can intuitively order these three categories according to formalization (with government being the most formal and large firms being more formal than small ones), we read these correlations as suggestive that there is a chain effect on formalization. We have calculated a measure of economic significance for these dummy variables as the difference in the predicted probability for the categorical variable at 1 and 0 in standard deviations of the formalization variable. The other variables were evaluated at their sample means. According to these measures, the economic significance of government, large and small clients is 0.3807, 0.1720 and 0.0810, respectively.

We also used a composite measure of formalization among a firm's suppliers to examine the chain effect. This measure consists of a weighted average of the formalization variable (*taxreg*) across supplying sectors using as weights the technical coefficients for input utilization from each sector. More precisely, the formality measure for the suppliers of firm i is given by

$$supplierformal_i = \frac{\sum_j tc_{ij} \times \text{formality}_j}{\sum_j tc_{ij}} \tag{31}$$

where formality_j is the percentage of firms in sector j that display tax registration[18] and tc_{ij} is the required amount of input from sector j per monetary unit of output produced by firm i (obtained from the technical coefficients for that firm's sector).

[17] A Poisson regression models the dependence of a countable random variable Y on covariates X. It postulates a poisson distribution for Y with expectation $\exp(\alpha + \beta' X)$.

[18] Four NAS sectors were excluded since they are not sampled in the ECINF survey: agriculture, mineral extraction, the sugar industry and other food products.

Although these variables are based on averages for all firms in a sector, as opposed to the actual firm's suppliers, the results of our analysis again favor the model: the coefficients attached to this variable are positive and statistically significant. The estimation results are displayed on Table 8. The marginal impact of supplier formalization on the probability of being formal is 0.230. The economic significance of the variable that measures formalization among suppliers is .0281. This means that an increase of one standard deviation in the formalization of suppliers increases the probability of formalization by 2.8% of a standard deviation.

A similar strategy was adopted for the sales of each firm, where a sectors' formalization is now weighted according to the output break up by sector obtainable as well from the NAS:

$$clientformal_i = \frac{\sum_j oa_{ij} \times \text{formality}_j}{\sum_j oa_{ij}} \tag{32}$$

The results are depicted on Table 8. The coefficient on this composite measure of client formalization is positive and statistically significant. The marginal impact for this coefficient is 0.36 and the economic significance is 0.0709 (one standard deviation increase in client formalization is associated with a seven percent of a standard deviation higher chance of formalization).

4.7 The Effect of Enforcement

The previous results show evidence of correlation in the degree of informality across stages of production. Our second model shows that increased tolerance towards informality in the upstream sector leads to a reduction in formalization in the downstream sector. Similarly, higher tolerance for informality among downstream firms is accompanied by higher degree of tax avoidance in the upstream sector. We use the measures of formalization enforcement in the labor market previously described as an indicator of monitoring within each state and economic sector from which a firm buys (using the technical coefficients as weights) and to which a firm sells (using the output allocations as weights). Our estimates on Table 9 show that enforcement in upstream or downstream sectors has a positive and significant effect on the probability of formalization for a given economic unit.[19]

[19]The seemingly low marginal impact is due mostly to the normalizing variable we use (number of firms in the state and sector). Using the number of employed persons, the effect is positive, significant and with a much higher marginal impact.

4.8 Robustness: Tax Substitution

Brazilian tax law imposes *forward tax substitution* ("substituição tributária para frente") for certain sectors in the economy.[20] Under this tax collection system, the value added tax is charged at the initial stage in the production chain at a rate estimated by the State. This scheme tends to be adopted for activities with a reduced set of initial producers and numerous smaller units at the subsequent stages of production. Since no extra value added tax is imposed one should not expect a chain effect within these particular activity sectors.

We ran probit estimates on activities where tax substitution is imposed. These activities (and their CNAE numerical activity designation) are automobile and autoparts manufacturing (34001, 34002, 35010, 35020, 35030, 35090), production of tires (25010), production and distribution of liquor (15050 and 53030), cigarettes (16000), commercialization of automobiles and tires (50010, 50030 and 540040), distribution of fuel (50050 and 53065), bars and similar establishments (55030) and oil refining (23010 and 23020).

The results concerning investment and installations, number of employees, and the entrepreneur's education level remain qualitatively as before. Table 10 shows that the the correlation of formalization across stages of production is substantially affected. The coefficient on sales to government clients remains positive, but decreases in significance. The coefficients on big and small firms have now the wrong sign and the sale to large companies is no longer statistically significant. There is no evidence for the chain effect in these activities.

5 Conclusion

We have presented two models of informality. In the first model, informal firms are shown to be smaller, less productive and with less capital per worker. Our second model showed that informality may be contagious across different industries when a vertical relationship exists and value added taxes are levied through the credit method. Using microdata from a survey in Brazil, we confirmed the implications of the models.

[20] Tax substitution is not peculiar to Brazil. In fact, records indicating its application by France, England and Germany date back to the 13th century. For a sample of the debate regarding this institution, see [15].

Appendix A: Non Degenerate Probability of Detection

We restrict ourselves to only one input: labor. In this model agents possess a degree of entrepreneurial ability quantified by the parameter θ, which is distributed according to a density function g. With an amount l of labor an entrepreneur can produce θl^{β} units of output, for which he or she pays a wage w taxed at a rate $\lambda - 1$ (if formally established). In case the entrepreneur chooses to be informal, he or she does not pay the tax but may be caught with a probability $p(l)$, which we assume to be a non-decreasing function of the number of employees a firm has. Once this happens, profits are ceased by the authorities and set to zero. A person can choose to be an entrepreneur in the formal or informal sectors or a worker.

The profit functions for formal and informal entrepreneurs are then given by:

$$\Pi_i(l) = (1 - p(l)) \times (\theta l^{\beta} - wl) \quad \text{and} \quad \Pi_f(l) = \theta l^{\beta} - \lambda wl$$

respectively.

Before proceeding we examine some conditions under which the maximization problem is concave. For the formal entrepreneur, $\beta < 1$ is sufficient. With respect to the informal manager, assume that $p(\cdot)$ is twice differentiable. Taking first derivatives one gets

$$\frac{d\Pi_i}{dl}(l) = -p'(l) \times (\theta l^{\beta} - wl) + (1 - p(l)) \times (\theta\beta l^{\beta-1} - w).$$

Differentiating this expression, we achieve

$$\frac{d^2\Pi_i}{dl^2}(l) = -p''(l) \times (\theta l^{\beta} - wl) - 2p'(l) \times (\theta\beta l^{\beta-1} - w) + (1 - p(l))\theta\beta(\beta - 1)l^{\beta-2}.$$

For the sake of illustration, take $p(l) = 1 - (1 + l)^{-\gamma}$. The parameter γ controls how fast the probability of capture goes to one as the number of employees increases as shown in Figure 1.

In a critical point,

$$\frac{d\Pi_i}{dl}(l) = -\gamma(1 + l)^{-1} \times (\theta l^{\beta} - wl) + (\theta\beta l^{\beta-1} - w) = 0.$$

The curvature of the function at this point is then given by

$$\begin{aligned}\frac{d^2\Pi_i}{dl^2}(l) &= \gamma(\gamma+1)(1+l)^{-\gamma-2}(\theta l^{\beta}-wl)-2\gamma(1+l)^{-\gamma-1}(\theta\beta l^{\beta-1}-w)+\\ &\quad +(1+l)^{-\gamma}\theta\beta(\beta-1)l^{\beta-2},\end{aligned}$$

which we want to be negative. Using the expression for the first derivative and simplifying terms, this requires that

$$(1-\gamma)(1+l)^{-1}(\theta\beta l^{\beta-1}-w)+\theta\beta(\beta-1)l^{\beta-2}\leq 0.$$

Noticing that the second term is negative and $\theta\beta l^{\beta-1}-w$, non-negative (from the first-derivative expression), a sufficient condition is that

$$\gamma\geq 1.$$

If, on the other hand, $\gamma\leq 1$, we may rewrite this expression as

$$(1-\gamma)(1+l)^{-1}(\theta\beta l^{\beta-1}-w)+\theta\beta(\beta-1)l^{\beta-2}\leq\frac{\theta\beta[(\beta-\gamma)l+(\beta-1)]}{l^{2-\beta}}$$

and the term in the right is negative if $\gamma\leq\beta$. So, a sufficient condition for the program to be concave is that $\gamma\leq\beta$. In other words, the probability of being caught rises fast enough with the employment size. This is by no means necessary, as $\gamma=0$ would still constitute a concave problem though. Examining the expressions above, one can see that, for reasonably small θs and/or large w the problem remains concave. Figure 2 depicts the profit function for different levels of γ and a specific set of parameters.

With a low enough probability of being caught it is intuitive that it pays for an entrepreneur to become and informal manager. In order to make it disadvantageous for higher ability entrepreneurs to become informal, the monitoring technology has to become informative sufficiently fast as employment rises.

Formally, we would need that there exists an entrepreneur that prefers formality:

$$\exists\theta:\Pi_f^*(\theta)=\frac{\beta^{\beta/(1-\beta)}(1-\beta^{\beta/(1-\beta)})}{(\lambda w)^{\beta/(1-\beta)}}\theta^{1/(1-\beta)}>(1-p(l))\times(\theta l^{\beta}-wl),\quad\forall l$$

Here, the left hand side expression is the optimal profit in the formal sector for the individual with ability θ. If we use $1-p(l)=(1+l)^{-\gamma}$, one observes that the above condition is easier to satisfy as γ increases. So, the faster the monitoring system improves with the number of employees, the more it favors formality.

References

[1] Almeida, R. and Z. Bianchini, 1998, "Aspectos de Amostragem da Pesquisa de Economia Informal Urbana", Texto para Discusso No.89, IBGE.

[2] Almeida, R. and P. Carneiro, 2005, "Enforcement of Regulation, Informal Labor and Firm Performance", IZA Discussion Paper No.1759.

[3] De Soto, H., 1989, *The Other Path: the Invisible Revolution in the Third World*, Harper and Row.

[4] De Soto, H., 2000, *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, Basic Books.

[5] Emran, M. and J. Stiglitz, 2005, "On Selective Indirect Tax Reform in Developin Countries", *Journal of Public Economics*, **89**, pp.599-623.

[6] Fortin, B., N. Marceau and L. Savard, 1997, "Taxation, Wage Controls and the Informal Sector", *Jounal of Public Economics*, **66**, pp.293-312.

[7] Friedman, E., S. Johnson, D. Kaufmann and P. Zoido-Lobaton, 2000, "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries", *Journal of Public Economics*, **76**, pp.459-493.

[8] Intituto Brasileiro de Geografia e Estatística (IBGE), 2003, *Sistema de Contas Nacionais - Brasil*, Série Relatórios Metodológicos, V.24.

[9] Junqueira, J. and J. Monteiro, 2005, "Outgoing the shadows: estimating the impact of bureaucracy and tax reduction on formality", Working Paper, PUC-Rio.

[10] Loayza, N., 1996, "The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America", *Carnegie-Rochester Conference Series on Public Policy*, **45**, pp.129-162.

[11] Lucas, R., 1978, "On the Size Distribution of Business Firms", *Bell Journal of Economics*, **9**, pp.508-523.

[12] McKinsey and Co., 2005, *Eliminando as barreiras ao crescimento econômico e à economia informal no Brasil.*

[13] Maloney, W. and J.N. Mendez, 2004, "Measuring the Impact of Minimum Wages: Evidence from Latin America", **in** J. Heckman and C. Pagés (eds.), *Law and Employment: Lessons from Latin America and the Caribbean*, Chicago: The University of Chicago Press.

[14] Rausch, J.E., 1991, "Modelling the Informal Sector Formally", *Journal of Development Economics*, **35**, pp.33-47.

[15] Ribeiro, A.P., 1997, "Substituição Tributária para Frente", *Revista CEJ*, Centro de Estudos Judiciários (Conselho de Justiça Federal), No.3, Brasília: CEJ.

[16] Schenk, A. and O. Oldman, 2001, *Value Added Taxes: A Comparative Approach in Theory and Practive*, Ardsley: Transnational Publishers.

[17] Schneider, F. and D. Enste, 2000, "Shadow Economies: Size, Causes and Consequences", *Journal of Economic Literature*, **38**, pp.77-114.

[18] Straub, S., 2005, "Informal Sector: The Credit Market Channel", *Journal of Development Economics*, **78**, pp.299-321.

[19] Tait, A., 1991, "Value-Added Tax: Administrative and Policy Issues", Occasional Paper 88, IMF.

[20] ten Raa, T., 2005, *The Economics of Input-Output Analysis*, Cambridge: Cambridge University Press.

Table 1: Variable Description

Variable	Description	Obs	Mean	Std. Dev.
taxreg	1 = Tax Registration	51474	0.161	0.367
taxsub	1 = Tax Substitution	51474	0.162	0.369
largecl	1 = Large Client	51474	0.049	0.215
smallcl	1 = Small Client	51474	0.079	0.269
govcl	1 = Government Client	51474	0.009	0.092
outsidehouse	1 = Outside Household	51474	0.645	0.478
n_employee	Number of Employees	51474	1.691	1.505
revenue	Revenue in Oct/2003 (R$ 1,000)	50604	2.630	8.140
otherjob	1 = Owner has Other Job	50340	0.110	0.313
bankloan	1 = Bank Loan	51449	0.066	0.248
education	Education Level (Owner)	45732	4.392	1.900
age	Age (Owner)	51457	40.840	12.312
gender	Gender (Owner)	45789	0.642	0.480
homeown_num	Homeowner × Number of Rooms	51442	0.786	0.410
loginst	Log of Installations	43004	5.977	1.805
loginv	Log of Investments	9325	6.744	2.208
profit	Profit in Oct/2003 (R$ 1,000)	47705	0.971	5.302
sup_enf	Supplier Enforcement	51474	10.678	13.992
cl_enf	Client Enforcement	51474	7.551	7.203

Table 2: Education

1 =	No education
2 =	Reads and writes
3 =	Some primary education
4 =	Graduated primary school
5 =	Some secondary education
6 =	Graduated secondary school
7 =	Some College education
8 =	Graduated College

Table 3: Correlation Matrix

	taxreg	taxsub	largecl	smallcl	govcl	outsdhous	n_empl	rev	otherjob
taxreg	1.000								
taxsub	0.054	1.000							
largecl	0.181	-0.085	1.000						
smallcl	0.092	-0.096	-0.073	1.000					
govcl	0.086	-0.035	-0.026	-0.032	1.000				
outsdhous	0.142	-0.036	0.054	0.002	-0.011	1.000			
n_empl	0.472	0.075	0.103	0.099	0.051	0.173	1.000		
revenue	0.359	0.065	0.195	0.056	0.085	0.125	0.371	1.000	
otherjob	0.008	-0.011	-0.025	-0.041	0.027	-0.028	0.058	0.003	1.000
bankloan	0.145	0.047	0.007	0.012	0.016	0.036	0.097	0.08	0.018
education	0.31	-0.157	0.111	0.138	0.077	0.034	0.222	0.172	0.149
age	0.03	0.068	0.016	-0.011	0.005	-0.021	0.024	0.031	-0.023
gender	0.056	0.008	0.07	0.06	0.027	0.233	0.044	0.067	-0.029
ho num	0.022	0.029	0.01	-0.022	-0.009	0.014	-0.008	0.018	-0.009
loginst	0.526	0.241	0.194	0.106	0.081	0.215	0.549	0.473	-0.027
loginv	0.413	0.042	0.224	0.12	0.085	0.208	0.366	0.305	-0.005
profit	0.092	-0.056	0.098	0.045	-0.003	0.049	0.086	0.418	0.003
sup_enf	-0.046	0.252	-0.042	-0.034	-0.007	-0.018	-0.019	0.001	-0.02
cl_enf	0.005	0.215	0.002	-0.004	0.009	-0.098	0.04	0.013	-0.012

Correlation Matrix (cont'd)

	bankloan	educ	age	gender	ho_num	loginst	loginv	profit	sup_enf
education	0.053	1.000							
age	0.005	-0.148	1.000						
gender	-0.01	-0.091	0.058	1.000					
ho num	-0.022	-0.025	0.146	0.003	1.000				
loginst	0.196	0.284	0.064	0.145	-0.011	1.000			
loginv	0.202	0.338	0.02	0.134	0.003	0.635	1.000		
profit	0.012	0.093	0.025	0.007	0.007	0.059	0.103	1.000	
sup_enf	-0.016	-0.138	-0.014	0.06	0.017	-0.007	-0.094	-0.016	1.000
cl_enf	0.076	-0.093	0.068	-0.005	0.019	0.115	0.068	-0.016	0.354

Table 4: Economic Sector

	Freq.	%	Description
1	5,639	10.96	Transformation and Mineral Extraction Industry
2	7,246	14.08	Construction
3	14,835	28.83	Commerce and Repair Services
4	4,679	9.09	Lodging and Food Services
5	4,636	9.01	Transportation and Communications
6	3,634	7.06	Real Estate and Services
7	3,453	6.71	Education, Health and Social Services
8	5,096	9.9	Other Collective, Social and Personal Services
9	2,246	4.36	Other Activities

Table 5: Probit Estimates

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**
outsidehouse	0.148**	0.018	0.151**	0.017	0.152**	0.017
	(0.025)		(0.025)		(0.025)	
n_employee	0.420**	0.0529	0.432**	0.050	0.431**	0.050
	(0.013)		(0.016)		(0.016)	
revenue	0.043**	0.005	0.042**	0.005	0.042**	0.005
	(0.005)		(0.009)		(0.009)	
bankloan	0.402**	0.065	0.366**	0.054	0.366**	0.054
	(0.036)		(0.036)		(0.036)	
otherjob	-0.198**	-0.022	-0.240**	-0.024	-0.239**	-0.024
	(0.037)		(0.037)		(0.037)	
education	0.202**	0.025	0.188**	0.022	0.188**	0.022
	(0.007)		(0.007)		(0.007)	
age	0.034**	0.004	0.036**	0.004	0.036**	0.004
	(0.005)		(0.005)		(0.005)	
age^2	0.000**	0.000	0.000**	-0.000	0.000**	-0.000
	(0.000)		(0.000)		(0.000)	
gender	0.204**	0.024	0.195**	0.022	0.197**	0.022
	(0.020)		(0.022)		(0.022)	
ho_num	0.120**	0.014	0.120**	0.013	0.120**	0.013
	(0.027)		(0.028)		(0.028)	
Sector Dummies	Yes		No		No	
Tech. Coeff.	No		Yes		No	
Output Coeff.	No		No		Yes	
State Dummies	Yes		Yes		Yes	
N	43925		43940		43940	
Log-likelihood	-10656.9		-10396.68		-10399.761	
$\chi^2_{(43)}$	4865.822					

1. Significance levels : † : 10% ∗ : 5% ∗∗ : 1%
2. Standard errors clustered by urban sector.
3. Prob > χ^2 is the p-value for the Wald χ^2 statistic displayed. Pseudo-R^2 = 1 - L_1/L_0, where L_0 and L_1 are the constant-onlyand full model log likelihoods.

Table 6: Investment, Installations and Profits

Dep. Var. =	loginvperworker	loginstperworker	profit
	Coefficient	**Coefficient**	**Coefficient**
	(Std. Err.)	(Std. Err.)	(Std. Err.)
taxreg	0.650**	0.815**	0.737**
	(0.057)	(0.021)	(0.171)
outsdhous	0.266**	0.333**	0.278**
	(0.044)	(0.015)	(0.063)
bankloan	0.752**	0.604**	-0.020
	(0.058)	(0.026)	(0.145)
otherjob	-0.285**	-0.241**	-0.174
	(0.058)	(0.022)	(0.128)
education	0.250**	0.129**	0.194**
	(0.012)	(0.004)	(0.018)
age	0.034**	0.062**	0.044**
	(0.010)	(0.003)	(0.013)
age^2	0.000**	-0.001**	0.000**
	(0.000)	(0.000)	(0.000)
gender	0.360**	0.316**	0.152**
	(0.043)	(0.015)	(0.041)
ho_num	0.034	-0.074**	-0.061
	(0.047)	(0.017)	(0.049)
revenue	0.019**	0.061**	
	(0.002)	(0.001)	
n_employee			0.410**
			(0.067)
N	7418	36512	41335
R^2	0.331	0.353	0.035
$F_{(43,\cdot)}$	85.022	462.267	19.697

1. Significance levels : † : 10% ∗ : 5% ∗∗ : 1%
2. The regressions also control for state and sector.
3. Standard errors are clustered by urban sector.

Table 7: Log of Number of Workers (= Dep. Var.)

Dep.Var. =	**Coefficient** (Std. Err.)	**Coefficient** (Std. Err.)
education	0.006**	-0.014**
	(0.001)	(0.004)
taxreg × education	0.081**	0.043**
	(0.001)	(0.002)
outsidehouse	0.055**	0.032*
	(0.004)	(0.014)
revenue	0.016**	0.007**
	(0.000)	(0.000)
bankloan	0.107**	0.041*
	(0.008)	(0.017)
otherjob	0.030**	0.018
	(0.006)	(0.017)
age	0.004**	0.000
	(0.001)	(0.003)
age^2	0.000**	0.000
	(0.000)	(0.000)
gender	0.012**	-0.023†
	(0.004)	(0.012)
ho_num	0.003	-0.004
	(0.005)	(0.014)
logwage		0.008
		(0.007)
Sector Dummies	Yes	Yes
State Dummies	Yes	Yes
N	43967	5783
R^2	0.299	0.206
$F_{(\cdot,\cdot)}$	436.398	33.782

Significance levels : † : 10% * : 5% ** : 1%

Table 8: Probit Estimates (Chain Effects)

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**
govcl	0.698**	.139				
	(0.095)					
largecl	0.412**	.067				
	(0.050)					
smallcl	0.195**	.028				
	(0.037)					
supplierformal			1.839**	0.230		
			(0.235)			
clientformal					2.998**	0.360
					(0.165)	
outsdhous	0.174**	0.021	0.157**	0.003	0.168**	0.020
	(0.025)		(0.024)		(0.024)	
n_empl	0.424**	0.053	0.424**	0.053	0.439**	0.053
	(0.012)		(0.012)		(0.013)	
revenue	0.043**	0.005	0.046**	0.006	0.041**	0.005
	(0.004)		(0.004)		(0.006)	
bankloan	0.389**	0.062	0.397**	0.064	0.365**	0.056
	(0.035)		(0.035)		(0.035)	
otherjob	-0.210**	-0.023	-0.219**	-0.025	-0.206**	-0.022
	(0.038)		(0.038)		(0.038)	
education	0.194**	0.024	0.197**	0.025	0.196**	0.024
	(0.006)		(0.006)		(0.007)	
age	0.030**	.004	0.031**	0.004	0.033**	0.004
	(0.005)		(0.005)		(0.005)	
age^2	0.000**	-0.000	0.000**	-0.000	0.000**	0.000
	(0.000)		(0.000)		(0.000)	
gender	0.154**	0.019	0.153**	0.018	0.170**	0.020
	(0.020)		(0.020)		(0.020)	
ho_num	0.131**	0.015	0.126**	0.015	0.117**	0.013
	(0.027)		(0.027)		(0.027)	
N	43967		43967		43967	
Log-likelihood	-10562.275		-10614.871		-10446.643	
χ^2	5003.921		5015.728		4960.043	

1. Significance levels : † : 10% ∗ : 5% ∗∗ : 1%
2. Standard errors clustered by urban sector.
3. The regressions also control for state and sector.
4. Prob > χ^2 is the p-value for the Wald χ^2 statistic displayed. Pseudo-R^2 = 1 - L_1/L_0, where L_0 and L_1 are the constant-only and full model log likelihoods.

Table 9: Probit Estimates (Enforcement)

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**
sup_enf	0.003**	0.0004		
	(0.001)			
cl_enf			0.012**	0.0015
			(0.002)	
outsidehouse	0.167**	0.0204	0.180**	0.0218
	(0.025)		(0.025)	
n_employee	0.425**	0.0537	0.423**	0.0532
	(0.012)		(0.012)	
revenue	0.045**	0.0057	0.045**	0.0057
	(0.004)		(0.004)	
bankloan	0.392**	0.0636	0.389**	0.06283
	(0.035)		(0.035)	
otherjob	-0.217**	-0.0241	-0.216**	-0.0238
	(0.038)		(0.038)	
education	0.204**	0.0258	0.204**	0.0256
	(0.006)		(0.006)	
age	0.030**	0.0038	0.030**	0.0038
	(0.005)		(0.005)	
age^2	0.000**	0.0000	0.000**	0.0000
	(0.000)		(0.000)	
gender	0.174**	0.0212	0.179**	0.0218
	(0.020)		(0.020)	
ho_num	0.123**	0.0147	0.125**	0.0149
	(0.027)		(0.027)	
N	43967		43967	
Log-likelihood	-10648.287		-10636.916	
$\chi^2_{(44)}$	4959.451		4984.399	

1. Significance levels : † : 10% ∗ : 5% ∗∗ : 1%
2. Standard errors clustered by urban sector.
3. The regressions also control for state and sector.
4. Prob > χ^2 is the p-value for the Wald χ^2 statistic displayed. Pseudo-R^2 = 1 - L_1/L_0, where L_0 and L_1 are the constant-only and full model log likelihoods.

Table 10: Probit Estimates (Tax Substitution)

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Coeff.** (Std. Err.)	**Coeff.** (Std. Err.)
govcl	0.815†		
	(0.439)		
largecl	-0.140		
	(0.242)		
smallcl	-0.568**		
	(0.136)		
comp_taxreg		-6.371*	
		(2.845)	
cl_taxreg			1.529**
			(0.332)
outsdhous	0.218**	0.214**	0.217**
	(0.050)	(0.050)	(0.050)
n_employee	0.362**	0.368**	0.376**
	(0.025)	(0.024)	(0.025)
revenue	0.043**	0.041**	0.039**
	(0.009)	(0.009)	(0.009)
bankloan	0.372**	0.349**	0.345**
	(0.070)	(0.071)	(0.071)
otherjob	-0.259**	-0.246**	-0.239**
	(0.082)	(0.081)	(0.081)
education	0.183**	0.178**	0.179**
	(0.015)	(0.015)	(0.015)
age	0.043**	0.043**	0.044**
	(0.011)	(0.011)	(0.011)
age^2	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
gender	0.206**	0.189**	0.177**
	(0.042)	(0.041)	(0.041)
ho_num	0.224**	0.227**	0.214**
	(0.064)	(0.063)	(0.063)
N	6997	6997	6997
Log-likelihood	-2350.366	-2359.251	-2350.645
$\chi^2_{(46)}$	1237.427	1240.023	1254.068

1. Significance levels : † : 10% * : 5% ** : 1%
2. Standard errors clustered by urban sector.
3. The regressions also control for state and sector.
4. Prob > χ^2 is the p-value for the Wald χ^2 statistic displayed. Pseudo-R^2 = 1 - L_1/L_0, where L_0 and L_1 are the constant-only and full model log likelihoods.

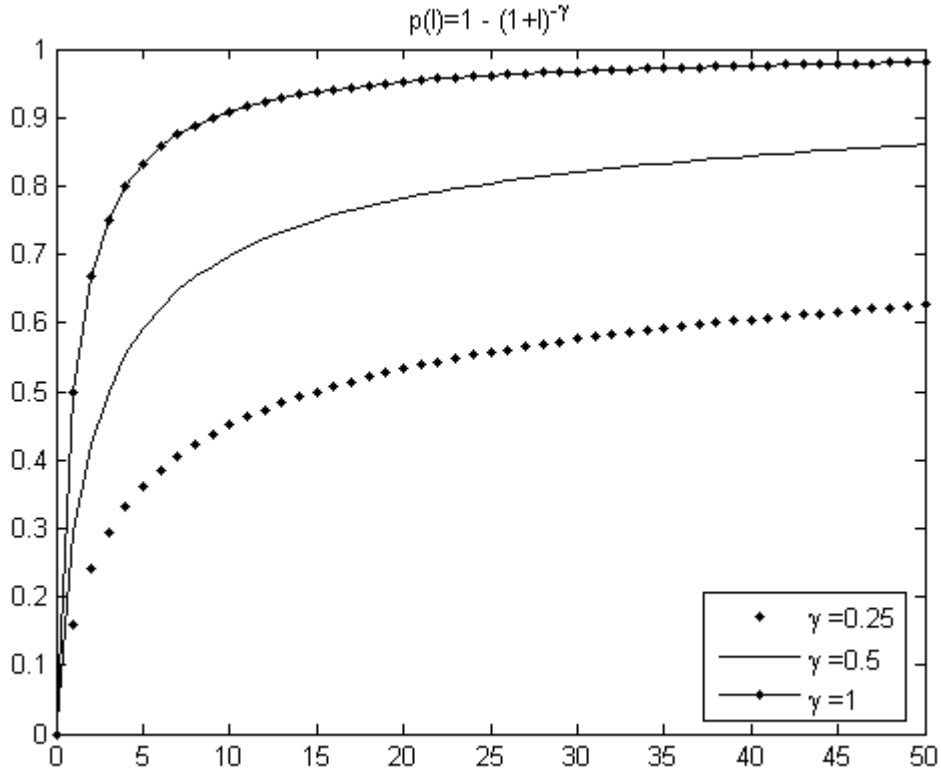

Figure 1: Probability of Being Caught

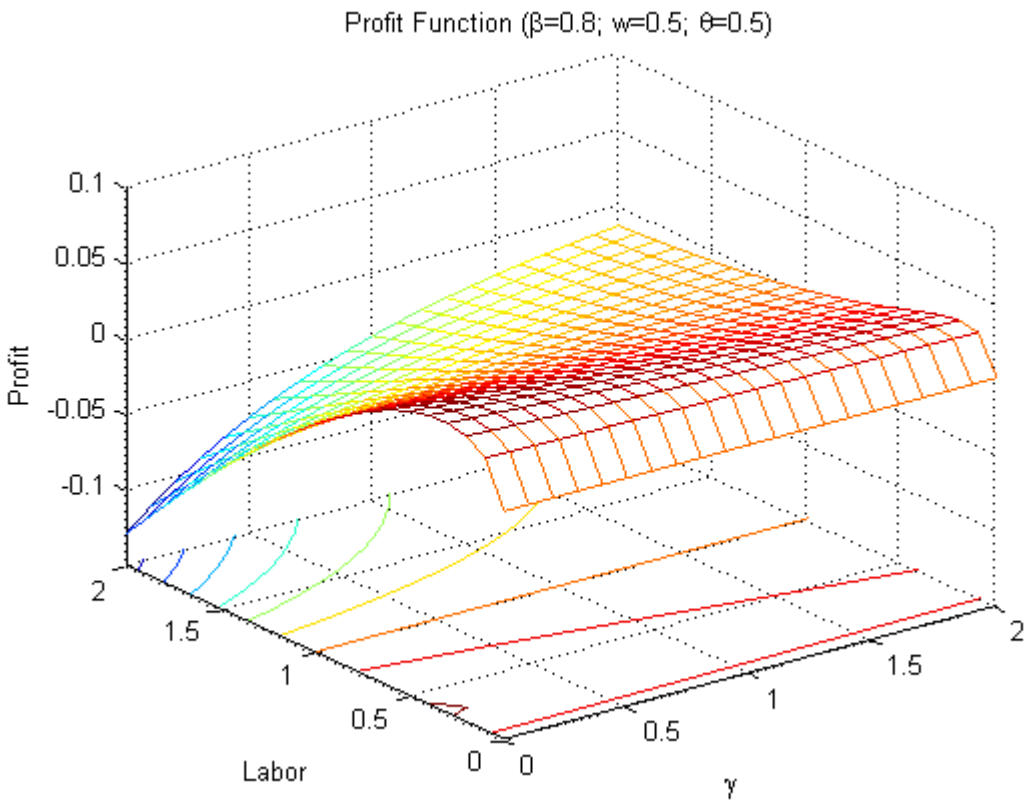

Figure 2: Profit Function for Informal Entrepreneur