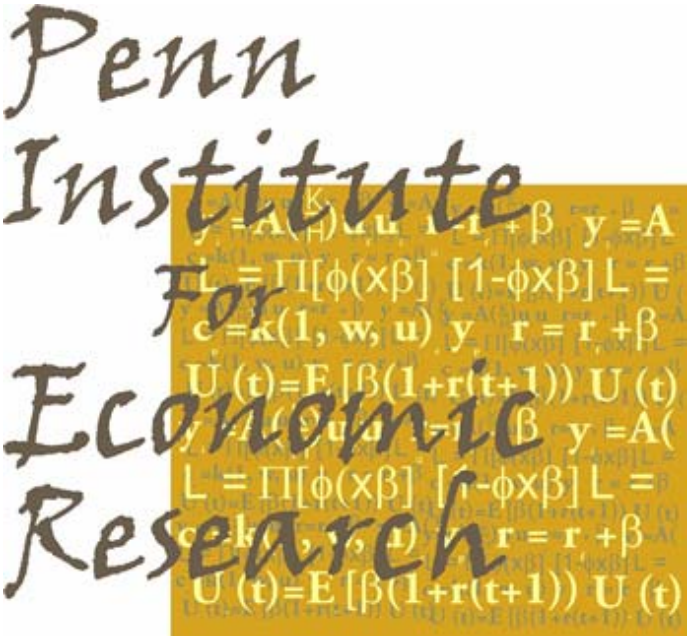

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 08-018

"The Informal Sector"
Third Version

by

Áureo de Paula and José A. Sheinkman

http://ssrn.com/abstract=1135724

The Informal Sector *

Áureo de Paula †
University of Pennsylvania

José A. Scheinkman ‡
Princeton University and NBER

*We thank Rita Almeida and Sandra Brandão for their help with the enforcement data, and Joana Monteiro and Juliano Assuncão for providing us with the ECINF dataset and clarifying its methodology. We also thank Julio Cacho, Paulo Natenzon, Justinas Pelenis and Glen Weyl for research assistance. We benefited from comments by seminar participants at several institutions and conferences and especially from conversations with William Maloney, Nicola Persico and Ken Wolpin. This material is based upon work supported by the National Science Foundation under Award Numbers SES0350770 and SES0718407. Portions of this work were done while José Scheinkman was a Fellow of the John Simon Guggenheim Memorial Foundation.

†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

The Informal Sector

Abstract

This paper investigates the determinants of informal economic activity. We present two equilibrium models of informality and test their implications using a survey of 48,000+ 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. When education is an imperfect proxy for ability, we show that the interaction of the manager's education and formality has a positive correlation with firm size. 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: H2, H3, K4.

1 Introduction

In this paper we investigate the determinants of informality. It is difficult to define informal activities unambiguously, but estimates indicate that in 1990-1993 approximately 10% of GDP in the United States was produced by individuals or firms that evaded taxes or engaged in illegal pursuits. It is also estimated that these activities produce 25 to 35% of aggregate output in Latin America, between 13 to 70% in Asian countries, around 15% in O.E.C.D. countries (see Table 2 in Schneider and Enste [23]). For the former Soviet Union states as well as other Eastern and Central European nations the underground economy is estimated to have been between 6 and 63.5% in the first half of the 90's (see Table 1 in Johnson *et al* [14]).

Informality creates a fiscal problem, but there is also growing evidence that informal firms are less efficient, 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. In many less developed countries, creating incentives for formalization is viewed as an important step to increase aggregate productivity.

We present two related equilibrium models of the determinants of informality and test their implications using a survey of 48,000+ small firms in Brazil. In both models informality is defined as tax avoidance. Firms in the informal sector avoid paying taxes but suffer from other limitations.

The first model can be seen as a variant of Rausch [19], who relied in the modeling strategy of Lucas [17] in which managerial ability differs across agents in the economy, and assumed a limitation on the size of informal firms. We make a modification that generates new testable implications. In addition to labor, the firms in our model use capital and informal firms face a higher cost of funds. This higher cost of capital for informal activities has been emphasized by DeSoto [6] who observed that because the right to assets held by the poor are not typically well documented "these assets cannot readily be turned into capital...[and] cannot be used as collateral for a loan...".[1] This difference in interest rates

[1]DeSoto [6], p.5-6. DeSoto [5] 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. We estimate a much smaller,

induces a higher capital-labor ratio in formal firms.[2] As in Rausch [19], agents with low managerial ability become workers and those with highest ability become formal managers, with an intermediate group running informal firms. 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 the scale limitations of informal firms. As a result, the marginal firm employs less capital and labor in the informal sector than it would employ if it joined the formal sector. Thus, as in Rausch [19], Fortin *et al.* [10] or Dabla-Norris *et al.* [4], a size gap develops. Managers that are slightly more efficient than the manager of the marginal informal firm employ discretely larger amounts of capital and labor.

In this class of models, entrepreneurs that operate in the informal sector are too inefficient to benefit from the lower capital costs and scale economies afforded to formal entrepreneurs. In this sense these models agree with the results from a survey of informal Mexican firms conducted by Mckenzie and Woodruff that is reported in Fajnzylber *et al* [9], where 75% of the respondents reported that they were too small to make it worth their while to become formal.[3]

Several implications of this model are supported by our empirical analysis on Brazilian data. 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 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 model predicts a correlation between manager's ability and the size of firm. Since a manager's ability predicts formality, formality should give no additional information

but still significant, difference in capital costs between informal and formal firms in our sample. Straub [24] develops a model in which a dual credit system arises in equilibrium.

[2]Informal firms may face lower labor costs, because their workers avoid some labor taxes. This would induce even larger differences in capital-labor ratio.

[3]...presumably relative to cost.

concerning size, once we condition on a manager's ability. However, ability is not observable and we study the implications of a model in which one can observe a variable, say educational achievement, that is correlated with a manager's ability. We show that a regression of the size of the firm on this observed variable and the interaction of the observed variable and formality should produce positive coefficients. This implication is supported by our empirical results.

The main focus of our theoretical analysis is a model that highlights the role of value added taxes (VAT) in transmitting informality. It exploits the idea that collecting value added taxes according to a credit scheme sets in motion a mechanism for the transmission of informality. The VAT is a prevalent form of indirect taxation: more than 120 nations had adopted it by 2000.[4] In the credit or invoice method that is often used, 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. Since purchases from informal suppliers do not generate tax credits and informal buyers cannot use tax payment from formal suppliers, there is an incentive for informal (formal) firms to deal with other informal (formal) firms.[5] Our 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. These findings survive when we use instrumental variables to control for possible simultaneity. 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 VAT is applied at some stage of production at a rate that is estimated by the State, this chain effect vanishes.

Since the mid 90's, following the lead of the Federal government, several Brazilian states introduced SIMPLES programs that simplified and lowered the VAT rates for small firms. The State of São Paulo, the largest and richest state in Brazil, introduced its SIMPLES in 1998. Rio Grande do Sul, another large and relatively rich southern state, started its own

[4]See Appendix 4 in Schenk and Oldman [22].

[5]To our knowledge, the only other study to investigate the informal sector in conjunction with a VAT structure is Emran and Stiglitz [8]. Their focus is on the consequences of informality for a revenue neutral tax reform involving value added and trade taxes.

program only in 2005. We use data on these two states and two rounds of the Brazilian survey of small firms to evaluate the impact of the introduction of these state programs. Our results point to a significantly positive impact of the program introduction, increasing the probability of formalization of small firms by approximately one-third.

The models in this paper ignore possible alternative reasons for informality, such as the fixed cost of complying with regulations, labor taxes or the existence of a minimum-wage. They also ignore benefits of formality that have been highlighted in the literature — such as access to participation in the legal system and other civil institutions. Considering these omitted costs and benefits should not change the qualitative implications of our models, which aim at providing additional — not necessarily exhaustive — explanations for informality. We focus on informality from the viewpoint of firms, not workers. There is a vast literature on labor informality, which is not addressed in the current paper. Finally, our models ignore partial compliance: firms either pay their taxes in full or not at all. This is a simplification that is also present in our data, which only provides us with binary information on formalization.

Other papers that investigate causes and determinants of informality include Loayza [16], Johnson *et al.* [14] and Friedman *et al.* [11] which provide 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). Junqueira and Monteiro [15] and Fajnzylber *et al.* [9] are recent papers that use an earlier (1997) wave of the the survey that we employ in this paper. They both explore the institution of the federal SIMPLES, which simplified and reduced rates for tax compliance for small firms in Brazil, to make inferences on the relation of taxes and informality. Although our empirical results speak to a somewhat different set of questions (for instance, the multi-stage transmission of informality captured by our second model), use data from a different year (2003 versus 1997) and refer to a different definition for formalization,[6] their empirical results are broadly in

[6]Junqueira and Monteiro [15] and Fajnzylber *et al.* [9] use municipal licensing as proxy for formalization instead tax registration, the measure we use. Junqueira and Monteiro recognize that tax registration would be a more appropriate indication of formalization, but opt for licensing because the question on tax registration

line with the implications of our models. Both papers find that the enactment of SIMPLES has increased formality through a smaller tax burden and cheaper formalization costs. In particular, Fajnzylber and co-authors find that the formalization is associated with more labor and capital stocks as well as higher productivity, which agrees with the predictions of our models. They fail to obtain significant effects on formalization of participation in government assisted programmes (about which our models are silent) and access to formal credit markets.[7]

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 parameterized by a scalar $\theta \geq 0$ that determines an agent's quality as an entrepreneur, and that is distributed according to a probability density function $g(\cdot)$. All agents are equally productive as workers. Each agent chooses between becoming a worker, operating a firm in the formal sector or in the informal sector. We assume that the production functions in the two sectors are identical. If an entrepreneur of quality θ employs l workers and k units of capital, output equals $y = \theta k^{\alpha} l^{\beta}$, with $\alpha, \beta > 0$

was only asked for those who indicated that their firm had been "legally constituted" — that is, a contract had been registered with the proper authorities. We do not view this as a problem, since according to Brazilian law only legally constituted firms are eligible for tax registration.

[7] In the preliminary version of Fajnzylber *et al.* [9] that we read, it is not clear how formal credit is defined, but we believe it refers to bank loans. In our empirical work we use a broader interpretation of credit — 40% of those who claimed to have obtained loans (25% of the formal entrepreneurs that claimed loans) did it from non-bank sources. In addition, Fajnzylber *et al.* [9] focuses on firms created around the time of the introduction of the SIMPLES in 1996, just after the implementation of the Real stabilization program, when Brazilian credit markets where much less developed than in 2002. The preliminary version also contains some omissions that prevent us to make more precise comparisons (for instance, which exogenous covariates they use and whether they control for sector of activity).

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 receive the same wage w.

An informal entrepreneur, if detected by the authorities, loses all profit. The probability of being detected depends monotonically on the size of the firm. Though there are several possibilities for measuring the size of a firm - output, capital stock or labor force - we choose here to use 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 a more general form for the function p can be adopted and our qualitative results are unchanged we assume here, for simplicity, that:

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

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

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

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 $r_i \geq r_f$, unconstrained informal firms have a lower capital-labor ratio than formal firms. In addition, constrained informal firms have a lower capital-labor ratio than 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 is the difference in capital costs between informal and formal firms $(r_i - r_f)$. In Section 4 we provide evidence in favor of the predicted difference in capital-labor ratios between formal and informal firms.

Each agent also has the choice of becoming a worker and receive the market wage w. Usual arguments in this class of models guarantee the existence of unique occupational

choice cutoff points. They are implicitly defined by:

$$\Pi_f(\overline{\theta}) = \Pi_i(\overline{\theta}) \tag{3}$$

$$\max\{\Pi_i(\hat{\theta}), \Pi_f(\hat{\theta})\} = w \tag{4}$$

and optimal choices are:

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

The 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 and cutoff points is straightforward. Workers, formal and informal firms will exist with positive probability as long as the support of g is large enough.

An implication of this model, which we explore empirically, 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 $\overline{k}$ 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 $k >> \overline{k}$. Furthermore, since we assumed that $r_i(1-\tau)^{\frac{1}{\alpha}} \leq r_f$ and $\Pi_i(\overline{\theta}) = \Pi_f(\overline{\theta})$ we know that

$$\Pi_i(\overline{\theta}) \leq \overline{\theta} l_f(\overline{\theta})^{\beta} k_f(\overline{\theta})^{\alpha}(1-\tau) - w l_f(\overline{\theta}) - r_i k_f(\overline{\theta})(1-\tau)^{1/\alpha}.$$

Hence $k_f(\overline{\theta})^{\alpha}(1-\tau) > \overline{k}^{\alpha}$, and, as a consequence:

$$\left(\frac{\overline{\theta}(1-\tau)\beta k_f(\overline{\theta})^{\alpha}}{w}\right)^{1/(1-\beta)} > \left(\frac{\overline{\theta}\beta\overline{k}^{\alpha}}{w}\right)^{1/(1-\beta)}. \tag{5}$$

The left (right) hand side of equation (5) is exactly the labor demand by a formal (informal) entrepreneur with quality $\overline{\theta}$. Hence labor demand also jumps up in the transition to formality. Thus our model predicts a "gap" in the capital and labor employed by firms near the the formalization threshold $\overline{\theta}$.

The empirical analysis of this gap is complicated because we do not observe an entrepreneur's ability θ and the data set we use has no information on interest rates paid. In order to account for these limitations we assume that entrepreneurial ability $\theta = x\exp(\epsilon)$ where ϵ is an unobserved determinant of entrepreneurial skill, independent of x and with zero expected value and x is some observed variable (or index of) that influences entrepreneurship. In our empirical application we take measures of education as proxies for x. In this case, one can use the expressions for optimal input level choices to obtain the expectation of the logarithm of employment l conditional on the $\log x$ and conditional on being in the formal or informal sector.

If one estimates a linear regression of $\ln l$ on $\ln x$ and an interaction between $\ln x$ and formalization ($\theta \geq \overline{\theta}$) as we do in our empirical section for a sample of entrepreneurs, the coefficient on the interaction term delivers the incremental sensitivity of $\ln l$ to $\ln x$ due to formalization. This is the sample counterpart of the best linear predictor of $\ln l$ conditional on $\ln x$ and $\mathbf{1}_{xe^{\epsilon}\geq\overline{\theta}}.\ln x$ in the population. We represent this object as

$$\mathbb{E}^{BLP}[\ln l|\ln x, \mathbf{1}_{xe^{\epsilon}\geq\overline{\theta}}.\ln x; xe^{\epsilon} \geq \hat{\theta}] = \xi_0 + \xi_1 \ln x + \xi_2 \mathbf{1}_{xe^{\epsilon}\geq\overline{\theta}}.\ln x$$

where the conditioning event $xe^{\epsilon} \geq \hat{\theta}$ reflects the fact that we use only observations on entrepreneurs. In the appendix we prove the following result:

Proposition 1 *Let x be a random variable that can only assume a finite number of values $\{x_i\}_{i=1}^n$. If $x_i \geq 0$, for $i = 1, \ldots, n$, with at least one non-zero element, then $\xi_2 > 0$.*

Proof. See Appendix.

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 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.

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. In the informal sector, only a limited amount of input 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 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

$$\Pi_f^u(\theta_u) = \pi p_f \theta_u \tag{6}$$

$$\Pi_i^u(\theta_u) = p_i \min\{\theta_u, \overline{y}\} \tag{7}$$

for the profit of an upstream firm with manager of quality θ_u if it produces in the formal (informal) sector. Downstream firms face a slightly more complicated problem, since they must also choose which intermediate good (formal or informal) to purchase.

Write

$$\Pi_f^d(\theta_d) = \max\{\max_x[\pi(q\theta_d x^\alpha - p_f x)], \max_x[q\pi\theta_d x^\alpha - p_i x]\}, \tag{8}$$

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

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

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

The demand of an informal entrepreneur of ability θ_d facing an input price p is:

$$x_i(\theta_d, p, q) = \min\left(\overline{x}, \left(\frac{q\alpha\theta_d}{p}\right)^{1/(1-\alpha)}\right). \tag{10}$$

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

$$x_f(\theta_d, p, q) = \left(\frac{q\alpha\theta_d}{p}\right)^{1/(1-\alpha)}, \tag{11}$$

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

The demand for the intermediate input will exhibit a discontinuity.

Proposition 2 *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's optimal choice is to buy the formal good and $p = \frac{p_i}{\pi}$ if the firm's 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: See Appendix.

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 $0 < \pi p_f \leq p_i \leq p_f$. In fact, if $\pi p_f > p_i$ profit maximization and equations (8) and (9) 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. Furthermore when $\pi p_f \leq p_i \leq p_f$ downstream informal (formal) entrepreneurs weakly prefer to buy from informal (formal) producers. If these inequalities are strict, preferences are also strict. 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).

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

Proposition 3 *(i) If* $\theta_u < \overline{\theta}_u(p_i, p_f) = \frac{p_i \overline{y}}{\pi p_f} \geq \overline{y}$ *then* $\Pi_i^u(\theta_u) \geq \Pi_f^u(\theta_u)$, *and if* $\theta_u > \overline{\theta}_u(p_i, p_f) = \frac{p_i \overline{y}}{\pi p_f}$ *then* $\Pi_i^u(\theta_u) < \Pi_f^u(\theta_u)$.

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

Proof: See Appendix.

As in the model with one stage, the size of firms will be discontinuous with respect to the quality of the entrepreneur.

Proposition 4 *(i) If* $p_f > \pi p_i$ *the output of the smallest upstream formal firm* $\frac{p_f \overline{y}}{\pi p_i} > \overline{y}$.

(ii) $\pi x_f(\overline{\theta}_d(p_i, p_f)) \geq \overline{x}$ *and, in particular, the output of the smallest downstream formal firm is strictly bigger than the output of the largest informal firm.*

Proof: See Appendix.

The existence of an equilibrium price vector $(p_i, p_f, 1)$ is established in the Appendix.

3.1 Comparative statics

Simulations of the model show that an increase in tolerance in the upstream sector increases the proportion of informal firms upstream and downstream. Figure 1 shows that as $\overline{y}$ increases, the proportion of upstream firms that are informal increases. As a result the price of the informal intermediate good p_i decreases and some of the downstream formal firms opt for informality. The fall in demand for the formal intermediate good causes a fall in its price p_f. A symmetric picture arises when we change the tolerance for informality in the downstream sector, $\overline{x}$.

[Figure 1 here]

4 Empirical Application

In this section we explore implications of our theoretical framework using a dataset on informal firms in Brazil. Tax noncompliance is an important phenomenon in Brazil. Schneider and Enste [23] estimate that informality represents more than one-quarter of the Brazilian economy. Its value added tax system was established in the sixties and value added taxes represent approximately 35% of total tax collection.

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 used the 2003 edition of that survey, collected in October 2003, which contains information on 48,701 entrepreneurs in urban regions from all states in the Brazilian federation. We also used the 1997 edition for the analysis present in subsection 4.7. The survey focused on units with five

or less employees.[8] The sampling strategy uses the demographic census as a frame. First, preliminary interviews screened households for the presence of at least one entrepreneur with a business employing five or less people, for possible inclusion in the survey. The sampling was done in two stages: in each state (of a total of 27) the primary sampling units (urban sectors) were 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 a probability proportional to the number of households in the sector. From each selected urban sector a total of 16 households was then randomly selected for interviews.[9]. Interviewees were told that the information collected for the survey was confidential and would only be utilized for statistical purposes and, in fact, a vast majority declared that their firm was informal.

An ideal dataset for testing our second model would contain information on the production chain associated with each firm. Although the ECINF contains certain characteristics of a firm's clientele (whether they were predominantly large or small companies, persons or governmental institutions), this information is quite 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 Brazilian national accounts.[10]

[8]The Brazilian small business administration SEBRAE defines small businesses as those with less than 10 workers in commerce or services or less than 20 workers in all other sectors. According to SEBRAE's Boletim Estatístico de Micros e Pequenas Empresas: Primeiro Semestre de 2005 (http://www.sebrae.com.br), in 2002 small businesses accounted for 93.6% of the total number of firms, employed 36.2% of the workers and responded for 10.3% of wages in 2002.

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

[10]Tables 1 and 2 under "Tabelas de Recursos e Usos" available under National Accounts on http://www.ibge.gov.br for 2003. The construction of technical coefficients follows the European System of Integrated Economic Accounts (ESA) specifications (see ten Raa [26]).

4.2 Description of Variables

We eliminated firms with owners who were less than 15 years old and the observations lacking education or gender information. Entrepreneurs who claimed that their main client was a governmental institution, which comprised less than 1% of the original data, were also discarded. This restricted our sample to around 48,000 observations.

Table 1 summarizes the main variables used in this paper. The first variable indicates formalization; it is a dummy variable that equals one if the firm is registered with the Brazilian tax authorities.[11] For firms in economic sectors that qualify for forward tax substitution (see subsection 4.8 for an explanation), *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*) (where large firms are those with more than five employees). Other alternatives are persons or ignored. *Outsidehouse* is a dummy that equals one when the activity is performed outside the home. The number of employees (*n_worker*) includes the owner. Even though the survey focused on firms with five or less employees, a few units (less than 0.1%) employ more than five people due to the lag between the screening and interviewing stages of the survey and the fact that firms may have multiple partners which are also counted as employees. The variables *revenue*, *otherjob* and *bankloan* are self-explanatory. *Education* is a categorical variable with values depicted in Table 2. *Age* of the owner is in years and *gender* equals 1 for male. The variable *ho_num* is a measure of wealth and is zero for non-homeowners and otherwise 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).[12] *Profit* equals revenue minus expenses in October/2003 (also in R$ 1,000). *Logwage* denotes the logarithm of the total expenditures in salaries (in R$1,000) divided by the

[11] The tax registry is the Cadastro Nacional de Pessoas Jurídicas, which replaced the previous system, the Cadastro Geral de Contribuintes (CGC), used in the 1997 survey. This variable is the most representative of formalization for our purposes, but we have nonetheless experimented with using "legally constituted firms" and obtained virtually identical results. This is not surprising, since, as we already mentioned, the latter is a prerequisite for tax registration and the correlation between the two measures of informality is 0.98.

[12] The value of installations refers to owned installations. Rented equipment is not included. Only 7% of formal firms and 7% of informal firms reported any rented equipment

number of employees in the firm.[13] The variables *clform* and *supform* measure formalization among customers and suppliers of a firm (see subsection 4.5 for the construction of these variables).

[Tables 1 and 2 here]

Each firm in the sample is classified into economic activities following the CNAE (Classificação Nacional de Atividades Econômicas) classification.[14] Using technical coefficients as well as sectoral output allocation coefficients from the National Accounts System (NAS) (using NAS sector classification) we are able to assign to each activity in the survey a vector with these coefficients. 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 (using NAS sector production value) of the coefficients in the corresponding NAS sectors. The ECINF survey also has its own aggregate sectoral characterization, displayed in Table 3.

We use these coefficients as a vector measure of sectoral allocation of output and sectoral input assignment by a firm. The last two variables in Table 1 are measures of formalization enforcement for suppliers and customers and were constructed as follows. 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.[15] We normalized the number of visits in each state and sector by the number of persons employed in that state and sector provided by the Brazilian Statistics Bureau (IBGE) (through the Cadastro Central de Empresas).[16] Assuming that a firm's clients were in the same state,[17] we generated an index of client formalization enforcement as a weighted average of these

[13]For comparison, annual GDP per capita in Brazil in 2003 was R$ 8,694.47.($\log(8.69447/12) = \log(0.72454) = -0.13$).

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

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

[16]Similar calculations were also performed using as normalizing variable the number of firms in the state-sector (also obtained through the Cadastro Central de Empresas).

[17]Only 5.23% of the firms in our sample sell in other states

variables, where the weights were the sectoral output allocation coefficients. We used sectoral input demand coefficients to obtain an analogous measure of supplier formalization enforcement.

The correlation matrix for our variables is in Table 4.

[Tables 4 and 3 here]

Table 5 contains probit estimates for the formalization variable *taxreg* using two different sets of controls. The signs obtained for each one of the regressors are as expected. The coefficient of the variable "working outside the home" is positive. In agreement with 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 likely to have substantial household duties, the sign on the gender variable is probably related to entrepreneurial input. These variables may also partially control for other determinants of informality, such as opportunities in the labor market.[18] The coefficients on all these variables are statistically significant.

[Table 5 here]

The two sets of estimates use different sectoral controls. In the first set we used dummies for state and sector (according to the specification in Table 3). In the second set of results we used the derived output coefficients obtained from the Brazilian National Accounts (similar results obtain using input coefficients). The National Accounts System in Brazil categorizes economic activity into forty-two sectors. The "use table" in the NAS allows one to obtain how much in a given year a sector required in terms of input from another sector in the economy. This can be used to derive technical coefficients for each NAS sector (see footnote 10). We were able to identify the NAS sector for each firm in the ECINF survey using equivalence tables among the different classification schemes that are

[18]With the available data it is not possible to control for business cycle variations another potentially relevant determinant of informality that is not captured in our model.

available from the Brazilian Statistics Bureau. The "make table" in the National Accounts provides the quantity of output destined to each sector of the economy and final demand. 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): $(oa_j)_{j=1,...,42}$. These controls, in additional to state dummies, were used in the second set of estimates presented in the table.

Our final descriptive table assembles estimations that focus on investments, capital and profits. Since an entrepreneur's true ability is not observable, it makes sense to measure the effect of formalization 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 0.51 for installations per worker regardless of the measure of formalization[19]. In other words, formalization is associated with an increase in investments (installations) per worker of 0.33 (0.51) standard deviations.

[Table 6 here]

We also examined the correlation of formalization with 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 association with profits. Formalization is associated with an increase in monthly profits of approximately 700 Reais.[20]

4.3 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. Our observable measure for entrepreneurial quality input, education, is

[19]For dummy variables, we define the economic significance as the regression coefficient divided by the standard deviation of the dependent variable.

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

an integer between 1 and 8. Hence $\ln x \geq 0$ and Proposition 1 guarantees that, if the model holds, the interaction coefficient should be positive.

Table 7 exhibits OLS estimates of the number of employees using education of the owner as the observable productivity enhancing feature and several additional controls. The coefficient of the interaction of education and formality is positive and significant. The result persists when we control for the level of wages within the firm. Since the number of employees is an integer, we also ran an ordered probit and a Poisson[21] regression and obtained very similar results.

[Table 7 here]

4.4 Cost of Capital

In the first model, the marginal product of capital of formal entrepreneurs is:

$$\frac{\alpha \times \theta(1-\tau)l^{\beta}k^{\alpha}}{k} = \frac{\alpha y(1-\tau)}{k}.$$

The marginal product of capital for *unconstrained* informal entrepreneurs is:

$$\frac{\alpha \times \theta l^{\beta}k^{\alpha}}{k} = \frac{\alpha y}{k}$$

These quantities should then equal the cost of capital: $\tilde{r}_f = \delta + r_f$ for formal and $\tilde{r}_i = \delta + r_i$ for unconstrained informal entrepreneurs, where δ is the common rate of depreciation. Since $\delta \geq 0$, $\frac{r_i}{r_f} \geq \frac{\tilde{r}_i}{\tilde{r}_f}$, and hence an estimate of $\frac{\tilde{r}_i}{\tilde{r}_f}$ is a lower bound for $\frac{r_i}{r_f}$. With the maintained assumption that α is the same for both formal and informal entrepreneurs, an estimator for $\frac{\tilde{r}_i}{\tilde{r}_f}$ would be:

$$\frac{y_i/k_i \text{ (for unconstrained informal firm)}}{(1-\tau)y_f/k_f \text{ (for formal firm)}}.$$

In practice, neither output nor capital are perfectly measured in the survey we used. Taking *revenue* (net of taxes) and the value of *installations* as imperfect measures of output

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

(net of taxes) and capital we would obtain:

$$\frac{\text{revenue}}{\text{installations}} = \frac{y + \epsilon_y}{k + \epsilon_k}$$

where ϵ_y and ϵ_k stand for the measurement errors in output and capital, which we assumed are on average zero and uncorrelated with output and capital. Under these assumptions, the average revenue and installation values converge in large samples to the expected output and capital in the population. Conventional application of the Central Limit Theorem and the Delta Method deliver:

$$\sqrt{N}\left(\frac{\text{avg revenue}}{\text{avg installation}} - \frac{\mathbb{E}(y)}{\mathbb{E}(k)}\right) = \sqrt{N}\left(\frac{\text{avg revenue}}{\text{avg installation}} - \frac{r}{\alpha}\right) \rightarrow_d \mathcal{N}(0, \Sigma)$$

where N is the number of observations and

$$\Sigma = \frac{\sigma^2_{revenue}}{\mathbb{E}(\text{installation})^2} - 2\frac{\mathbb{E}(\text{revenue})\sigma_{revenue,installations}}{\mathbb{E}(\text{installation})^3} + \frac{\mathbb{E}(\text{revenue})^2\sigma^2_{installations}}{\mathbb{E}(\text{installations})^4}$$

where σ^2 denote variances and $\sigma_{revenue,installations}$ the covariance between revenues and installations. Σ can be estimated consistently by its sample analog which we write as $\widehat{\Sigma}$. We append the subscript i or f to N, Σ and r when referring to unconstrained informal or formal entrepreneurs respectively. The estimator relies on the assumption that the measurement error is averaged out across many randomly sampled individual and is reminiscent of the strategy used by Milton Friedman in his classical study of consumption.

Assume now that one samples independently N_f formal entrepreneurs and N_i unconstrained informal entrepreneurs and that N_i/N_f converges to a positive value c as the sample size grows. An additional application of the usual asymptotic arguments shows that the distribution of the ratio of revenue per installation for unconstrained informal entrepreneurs and for formal entrepreneurs can be approximated in large samples by

$$\sqrt{N_f}\left(\frac{\frac{\text{avg revenue}}{\text{avg installations}}\text{for unconstrained informal firms}}{\frac{\text{avg revenue (net of taxes)}}{\text{avg installations}}\text{for formal firms}} - \frac{\tilde{r}_i}{\tilde{r}_f}\right) \rightarrow_d \mathcal{N}(0, V)$$

where

$$V = \frac{1}{(\tilde{r}_f/\alpha)^2}\Sigma_i + c\left(\frac{\tilde{r}_i/\alpha}{(\tilde{r}_f/\alpha)^2}\right)^2\Sigma_f$$

which again can be consistently estimated using the sample analogs for its components (for c use actual N_i/N_f).

Among the informal firms, the *unconstrained* entrepreneurs are those with lower skill parameter θ. Since more able entrepreneurs will employ more capital and more labor, we can use the number of workers as a sorting mechanism and focus on the group of entrepreneurs employing lower amounts of labor. Using informal employers with two or less workers leads to a point estimate of $\frac{\tilde{r}_i}{\tilde{r}_f}$ of 1.31 with a standard error of 0.0178. Using informal employers with only one worker yield similar estimates. Hence we estimate that, in our data set, informal firms face a rate of interest that is at least 1.3 times the interest rate faced by formal firms.

4.5 Chain Effects on Formalization

One initial approach to investigate the existence of cross-firms effects of formalization was to employ a characterization of a firm's clientele in the ECINF survey. Interviewees were asked to declare whether sales were principally to large firms (more than five employees), small firms, or persons. Sales to firms tend to increase the probability of formalization with a bigger effect being associated with large firms as depicted in Table 8. These correlations are supportive of the existence of a chain effect in formalization.

We also used a composite measure of formalization among a firm's suppliers to examine this 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

$$supplierformal_i = \frac{\sum_j tc_{ij} \times \text{formality}_j}{\sum_j tc_{ij}} \tag{12}$$

where formality_j is the percentage of firms in sector j that display tax registration[22] 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). Some caveats apply. This measure of supplier's formality only accounts for potential suppliers that are present in

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

the survey and, in particular, ignores all suppliers that are large firms. On the other hand, the technical coefficients are obtained from a sample of formal firms. The net impact on the measure of formalization is unclear. Nevertheless, 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 in Table 8. The marginal impact of supplier formalization on the probability of being formal is 0.358.

A similar strategy uses the sales of each firm, where a sectors' formalization is now weighted according to the output break up by sector:

$$clientformal_i = \frac{\sum_j oa_{ij} \times \text{formality}_j}{\sum_j oa_{ij}} \tag{13}$$

The results are depicted in Table 8. The coefficient on this composite measure of client formalization is positive and statistically significant, with a marginal impact of 0.618.

[Table 8 here]

To address the likely correlation of `largecl` with observable determinants of formality for a given firm, the regressions in Table 8 include sector and state controls and other covariates which we viewed as the most natural confounding variables. We have nonetheless tried additional specifications. First, we experimented with a narrower sector classification (the CNAE numerical activity designation) for controls instead of that displayed in Table 3. The results are unchanged. We also estimated the regressions of subsamples with different sectors (manufacturing and services) and the conclusions are basically unchanged. To handle potential correlation of capital intensity of production across various stages of a given production process we included a quadratic polynomial on capital (installations) in our regressions and the marginal effects are essentially unchanged for the main variables (`largecl` and `smallcl`). We also ran regressions including `largecl`, `smallcl` and the client and supplier formality proxies from regressions 2 and 3 together. The coefficients are all positive and only the supplier-formality coefficient is no longer significant.

While the degree of tax compliance among a firm's suppliers and customers seems to affect formalization, an endogeneity problem may arise since suppliers and customers of a firm

respond to the degree of tax compliance of that firm. This would tend to bias the estimator upwards. Since the variable we use as a proxy for formalization among clients is an imperfect measure of tax compliance, measurement error is also present. In our case, with mismeasured categorical variables, one cannot rule out the possibility of attenuation bias in the opposite direction of the simultaneity bias (see Bound *et al.* [3]). To address this potential endogeneity problem we ran instrumental variable versions for the estimation results displayed in Table 8 using the average education level in an entrepreneurs urban sector as an instrument for the formalization of his clients. The assumption is that the average education level in the urban sector only affects one's propensity to be formal through formalization of his or her clients and that the average education level in the urban sector does not respond to the formalization of a single individual. Similar strategies were used for instance in DiPasquale and Glaeser [7]. Since we used a single instrumental variable (and hence can only handle one endogenous variable), we consolidated the dummy variables indicating large firms and small firms as a single variable. Table 9 displays the results for the first set of estimates in Table 8 using the aggregate variable in place of `largecl` and `smallcl` and its IV version.[23] The coefficient on the consolidated variable, `lscl`, is positive and remains so in the IV version. In fact, the IV version displays an even larger coefficient, which we ascribe to the attenuation effect of measurement error in the non-instrumented estimation.

[Table 9 here]

We have also run instrumented and non-instrumented probit regressions using a subsample of firms having only large and small firm clients and using the latter as baseline. The coefficient for the large client dummy is also positive in the non-instrumented version of this estimation and it also increases when we use the instrumental variable.

[23]Since both our outcome of interest (`taxreg`) and endogenous variable (`lscl`) are dichotomous, standard procedures such as TSLS or Rivers and Vuong [21] are inadequate and we used a bivariate probit to generate our IV estimates as suggested in Heckman [12]. To achieve numerical convergence to a maximum, we had to drop `revenue`. We repeated the estimation using linearized TSLS and the Rivers-Vuong approach including `revenue` and the results are qualitatively unchanged — the coefficient on `lscl` is even larger.

4.6 The Effect of Enforcement

The previous results show evidence of correlation in the degree of informality across stages of production. Our second model suggests 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 should be accompanied by higher degree of tax avoidance in the upstream sector. We use the measures of formalization enforcement in the labor market described in subsection 4.2 as an indicator for monitoring within each state-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 in Table 10 show that enforcement in upstream or downstream sectors has a positive and significant effect on the probability of formalization.

[Table 10 here]

4.7 SIMPLES: São Paulo and Rio Grande do Sul

In 1996 the Brazilian federal government established the SIMPLES tax program, targeted at small firms - those with roughly less than R$1,000,000 in annual revenues. It consolidated taxes and social security contributions in a single payment and simplified verification and remittance procedures for tax collection. Although states and municipalities were allowed to join the system for the collection of value added taxes, very few did. More than 20 states eventually established instead their own state-level versions of the SIMPLES system for the collection of VAT and other state taxes. In 1998, the state of São Paulo established a local version of the SIMPLES program. The system exempted firms with less than R$ 120,000 annual revenues from the collection of state VAT and offered reduced rates to firms with at most R$1.2 million in annual revenues. The program provided firms with a significant reduction in VAT. For example, a firm with monthly sales of R$60,000 with input costs of R$20,000 would owe R$7,200 in VAT before the SIMPLES. Under the new program the VAT would total less than R$1,300.

We used the first round of the ECINF survey, collected in 1997, and its 2003 edition

to measure the effect of this reduction in VAT on formalization in the state of São Paulo. For comparison we used data from the state of Rio Grande do Sul, which established its state SIMPLES only by the end of 2005. Table 11 displays summary statistics on some key variables in 1997 for these two states. With the exception of the number of workers, the proportion of registered firms and whether the entrepreneur holds other jobs, the means for the variables are not significantly different at the 10% level.[24]

[Table 11 here]

Table 12 displays results from a probit model where dummy variables for the state and pre- and post-introduction of the state SIMPLES are used to assess the variation in the formalization in São Paulo. We applied the same controls we used in our previous formalization regressions.[25] The results point to a positive impact of the program's introduction with a marginal effect of 5.6 percentage points on formalization, an increase in the probability of formalization of approximately one-third.

[Table 12 here]

4.8 Robustness: Tax Substitution

Brazilian tax law imposes *forward tax substitution* ("substituição tributária para frente") in certain sectors.[26] Under this tax collection system, the value added tax is charged at some stage in the production chain at a rate estimated by the State. This method tends to be adopted for activities with a reduced set of initial producers and many smaller units at the subsequent stages of production. Since no extra value added tax is imposed one should not expect a chain effect in these sectors.

[24]The introduction of the SIMPLES seems to have had an effect on the number of firms per household, perhaps in response to the size-related eligibility for the program. In 1997, about 21% of the households in RS and about 16% in SP had more than one business. In 2003, this proportion grew by 8.6% percentage points in SP and decreased by 4 percentage points in RS.

[25]Standard errors are not clustered by urban sector since their definition varied between 1997 and 2003.

[26]Tax substitution is not peculiar to Brazil. See [20].

We ran probit estimates on activities where tax substitution is imposed. These activities (and their CNAE numerical activity designation) are automobile and auto-parts 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, 50020, 50030 and 54040), 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. In Table 13 we interact tax-substitution with our measure of formality of the clients. To facilitate comparisons with the results in Table 9 we again consolidate the dummy variables indicating large firm and small firm clients as a single variable. The coefficient of the interaction term is negative and significant. This is implied by the regression run on the subsample of firms eligible for tax substitution displayed in Table 13. The coefficient on the client formalization cease to be positive. If anything, there is evidence for a *negative* rather than positive coefficient.

[Table 13 here]

5 Conclusion

We presented two models of informality. An implication of the first model is that informal firms are smaller, less productive and with less capital per worker. The second model predicts that informality may be transmitted through vertical relationships when value added taxes are levied through the credit method. Using microdata from surveys conducted in Brazil, we confirmed implications of both models.

In our models, informal firms are less efficient than formal ones, but our analysis is silent concerning the optimal amount of informality. To discuss this question one must also model the cost of enforcement, presumably as a function of a firm's size, and the value of tax revenues.

Tax authorities in Brazil impose tax substitution hoping to increase compliance. Our second model predicts a decrease in the interaction effect when tax substitution is imposed,

but does not make any prediction concerning the effect of tax substitution on the level of informality. The firms in our sample that belong to tax substitution sectors tend to have more individuals as main clients and to be owned by less educated entrepreneurs; factors associated with less formality. Nonetheless, the difference in the rate of formalization between firms in tax substitution sectors and the other firms is 7.8 percentage points (with a standard error of .4), a very large difference when compared with the average level of 13.2% in our sample. This increased formalization probably reflects the criterium used by Brazilian tax authorities. Tax substitution is imposed when at some level in the chain the typical producer is a large firm which is then charged the estimated value added tax. If these large firms cannot afford to become informal and pay the estimated value added tax, smaller firms in the same production chain will face lower costs to formality than the typical small firm in Brazil and, for this reason, may more often opt to become formal.

Appendix A: Proofs

Proof of Proposition 1

The proof is by induction on the cardinality of $supp(x)$. The notation $supp$ denotes the support of a given random variable. For a set A, $\#A$ is the cardinality of that set. Recall that we assume that $\epsilon \sim G(\cdot)$ is independent of x and $supp(\epsilon) = \mathbb{R}$.

STEP 1: $(\#supp(x) = 1)$ In this case, $\ln x$ is a constant and we can focus on:

$$\mathbb{E}^{BLP}[\ln l | \ln x, \mathbf{1}_{xe^{\epsilon} \geq \bar{\theta}}. \ln x; xe^{\epsilon} \geq \hat{\theta}] = \varphi_0 + \varphi_1 \mathbf{1}_{xe^{\epsilon} \geq \bar{\theta}}$$

where $\varphi_0 = \xi_0 + \xi_1 \ln x$ (so that ξ_0 and ξ_1 are not separately identifiable) and $\varphi_1 = \xi_2 \ln x$. We will show that $\varphi_1 > 0$ and this in turn implies that $sgn(\xi_2) = sgn(\ln x)$. This being a best linear projection,

$$\varphi_1 = \frac{cov(\ln l(xe^{\epsilon}), \mathbf{1}_{xe^{\epsilon} \geq \bar{\theta}} | xe^{\epsilon} \geq \hat{\theta})}{var(\mathbf{1}_{xe^{\epsilon} \geq \bar{\theta}} | xe^{\epsilon} \geq \hat{\theta})} \Rightarrow sgn(\varphi_1) = sgn(cov(\ln l(xe^{\epsilon}), \mathbf{1}_{xe^{\epsilon} \geq \bar{\theta}} | xe^{\epsilon} \geq \hat{\theta}))$$

where we stress the point that the equilibrium demand for labor $l(xe^{\epsilon})$ is a function of x and ϵ. Let $\overline{\epsilon}$ solve

$$xe^{\overline{\epsilon}} = \overline{\theta} \Leftrightarrow \overline{\epsilon} = \ln \overline{\theta} - \ln x$$

and $\hat{\epsilon}$ solve

$$xe^{\hat{\epsilon}} = \hat{\theta} \Leftrightarrow \hat{\epsilon} = \ln \hat{\theta} - \ln x$$

The covariance can then be written as

$$\begin{aligned} cov(\ln l, \mathbf{1}_{xe^{\epsilon} \geq \overline{\theta}} | xe^{\epsilon} \geq \hat{\theta}) &= \int_{\epsilon \geq \hat{\epsilon}} \ln l(xe^{\epsilon}).\mathbf{1}_{xe^{\epsilon} \geq \overline{\theta}} dG(\epsilon | \epsilon \geq \hat{\epsilon}) \\ &\quad - \int_{\epsilon \geq \hat{\epsilon}} \ln l(xe^{\epsilon}) dG(\epsilon | \epsilon \geq \hat{\epsilon}). \int_{\epsilon \geq \hat{\epsilon}} \mathbf{1}_{xe^{\epsilon} \geq \overline{\theta}} dG(\epsilon | \epsilon \geq \hat{\epsilon}) \\ &= \int_{\epsilon \geq \overline{\epsilon}} \ln l(xe^{\epsilon}) dG(\epsilon | \epsilon \geq \hat{\epsilon}) - \int_{\epsilon \geq \hat{\epsilon}} \ln l(xe^{\epsilon}) dG(\epsilon | \epsilon \geq \hat{\epsilon}). \frac{1 - G(\overline{\epsilon})}{1 - G(\hat{\epsilon})} \\ &= \frac{G(\overline{\epsilon}) - G(\hat{\epsilon})}{1 - G(\hat{\epsilon})} \int_{\epsilon \geq \overline{\epsilon}} \ln l(xe^{\epsilon}) dG(\epsilon | \epsilon \geq \hat{\epsilon}) \\ &\quad - \int_{\hat{\epsilon} \leq \epsilon < \overline{\epsilon}} \ln l(xe^{\epsilon}) dG(\epsilon | \epsilon \geq \hat{\epsilon}). \frac{1 - G(\overline{\epsilon})}{1 - G(\hat{\epsilon})} \end{aligned}$$

Also notice that the optimal choice of labor input for an unconstrained firm is

$$\begin{aligned} \ln l(\theta, r, \tau) &= \frac{1}{1 - \beta} \ln \beta + \ln \theta + \frac{1}{1 - \alpha - \beta} \ln(1 - \tau) + \frac{\alpha}{1 - \alpha - \beta} \ln \alpha - \\ &\quad \frac{\alpha}{1 - \alpha - \beta} \ln r - \frac{1 - \alpha}{1 - \alpha - \beta} w. \end{aligned}$$

where $\tau = 0$ and $r = r_i$ if the entrepreneur is informal and $\tau > 0$ and $r = r_f$ otherwise. Remember that

$$l(\overline{\theta}, r_f, \tau) > \overline{l}$$

where $l(\overline{\theta}, r_f, \tau)$ is the optimal labor demand of a formal firm with skill parameter $\overline{\theta}$ and $\overline{l}$ is the labor demand for an informal firm with skill parameter $\overline{l}$ constrained to employ at most $k = \overline{k}$. This information is important because

$$xe^{\epsilon} \geq \overline{\theta} (\Leftrightarrow \epsilon \geq \overline{\epsilon}) \Rightarrow \ln l(xe^{\epsilon}) > l(\overline{\theta}, r_f, \tau)$$

and

$$xe^{\epsilon} < \overline{\theta} (\Leftrightarrow \epsilon < \overline{\epsilon}) \Rightarrow \ln l(xe^{\epsilon}) < \ln \overline{l}.$$

So the covariance should be

$$\begin{aligned} cov(\ln l, \mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}|xe^{\epsilon}\geq\hat{\theta}) &= \frac{G(\bar{\epsilon})-G(\hat{\epsilon})}{1-G(\hat{\epsilon})}\int_{\epsilon\geq\bar{\epsilon}}\ln l(xe^{\epsilon})dG(\epsilon|\epsilon\geq\hat{\epsilon}) \\ &\quad -\int_{\hat{\epsilon}\leq\epsilon<\bar{\epsilon}}\ln l(xe^{\epsilon})dG(\epsilon|\epsilon\geq\hat{\epsilon}).\frac{1-G(\bar{\epsilon})}{1-G(\hat{\epsilon})} \\ &> \frac{(G(\bar{\epsilon})-G(\hat{\epsilon}))\,(1-G(\bar{\epsilon}))}{(1-G(\hat{\epsilon}))^{2}}(\ln l(\bar{\theta},r_f,\tau)-\ln\bar{l}) \\ &\geq 0 \end{aligned}$$

STEP 2: $(\#supp(x)=n)$ Assume that $supp(\ln x)\subset\mathbb{R}_+$ and that the assertion in the proposition is valid for $\#supp(x)=n-1$.

Consider the following best linear projections:

$$\ln l = \alpha_0+\alpha_1\ln x+\eta$$

and

$$\mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}.\ln x = \beta_0+\beta_1\ln x+\nu.$$

These being best linear projections,

$$\eta = \ln l-\mathbb{E}(\ln l|xe^{\epsilon}\geq\hat{\theta})-\alpha_1[\ln x-\mathbb{E}(\ln x|xe^{\epsilon}\geq\hat{\theta})]$$

and

$$\nu = \mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}.\ln x-\mathbb{E}(\mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}.\ln x|xe^{\epsilon}\geq\hat{\theta})-\beta_1[\ln x-\mathbb{E}(\ln x|xe^{\epsilon}\geq\hat{\theta})]$$

where

$$\alpha_1=\frac{cov(\ln l,\ln x|xe^{\epsilon}\geq\hat{\theta})}{var(\ln x|xe^{\epsilon}\geq\hat{\theta})}\quad\text{and}\quad\beta_1=\frac{cov(\mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}.\ln x,\ln x|xe^{\epsilon}\geq\hat{\theta})}{var(\ln x|xe^{\epsilon}\geq\hat{\theta})}.$$

The Frisch-Waugh-Lowell Theorem then allows us to state that

$$\xi_2=\frac{cov(\eta,\nu|xe^{\epsilon}\geq\hat{\theta})}{var(\nu|xe^{\epsilon}\geq\hat{\theta})}.$$

The covariance in the numerator will determine the sign of ξ_2. This can be seen to be:

$$\begin{aligned} cov(\ln l,\mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}|xe^{\epsilon}\geq\hat{\theta}).\ln x &- \beta_1 cov(\ln l,\ln x|xe^{\epsilon}\geq\hat{\theta}) = \\ & cov(\ln l,(\mathbf{1}_{xe^{\epsilon}\geq\bar{\theta}}-\beta_1).\ln x|xe^{\epsilon}\geq\hat{\theta}). \end{aligned}$$

Let $\overline{x} = \max supp(x)$ and $K = supp(x) - \{\overline{x}\}$. We can view x as a mixture of two distributions: with probability $\mathbb{P}(x = \overline{x})$ we sample from a distribution that delivers $\overline{x}$ with certainty and with complementary probability we sample from the distribution of x conditional on the event $\{x \in K\}$. The first one has a support of size one and the second, a support of size $n - 1$.

An analysis of variance argument yields

$$\begin{aligned} cov(\ln l, \mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | xe^\epsilon \geq \hat{\theta}) &= \mathbb{E}\{cov(\ln l, \mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K; xe^\epsilon \geq \hat{\theta}) | xe^\epsilon \geq \hat{\theta}\} + \\ & \quad cov(\mathbb{E}(\ln l | \mathbf{1}_K), \mathbb{E}(\mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K; xe^\epsilon \geq \hat{\theta}) | xe^\epsilon \geq \hat{\theta}) \end{aligned}$$

where $\mathbf{1}_K = 1$ if the sample is taken from K and $= 0$, otherwise.

When $\mathbf{1}_K = 1$, the conditional covariance $cov(\ln l, \mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K = 1 | xe^\epsilon \geq \hat{\theta}) > 0$ because $\ln x > 0$ and $\#K = n - 1$. Alternatively, for $\mathbf{1}_K = 0$ the conditional covariance $cov(\ln l, \mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K = 0; xe^\epsilon \geq \hat{\theta}) = cov(\ln l, \mathbf{1}_{\overline{x}e^\epsilon \geq \overline{\theta}} \ln \overline{x} | xe^\epsilon \geq \hat{\theta})$ can be seen to be positive using an argument akin to the one on Step 1 and the fact that $\ln \overline{x} > 0$. The expectation of these conditional covariances is hence positive.

Notice as well that the $\mathbb{E}(\ln l | \mathbf{1}_K = 0; xe^\epsilon \geq \hat{\theta}) > \mathbb{E}(\ln l | \mathbf{1}_K = 1; xe^\epsilon \geq \hat{\theta})$ and $\mathbb{E}(\mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K = 0; xe^\epsilon \geq \hat{\theta}) > \mathbb{E}(\mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x | \mathbf{1}_K = 1; xe^\epsilon \geq \hat{\theta})$ since $\overline{x} > x, \forall x \in K$ and both $\ln l$ and $\mathbf{1}_{xe^\epsilon \geq \overline{\theta}} \ln x$ are increasing in x for every given ϵ. Consequently, the covariance of the conditional expectations is positive. By induction, the result holds. ■

Proof of Proposition 2

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. ■

Proof of Proposition 3

(i) is immediate from equations (6) and (7). To show that (ii) holds note that θ_d enters the definition of the profit function of formal firms exactly as an output price and hence, from the properties of profit functions with respect to output prices, we know that its derivative with respect to θ_d is proportional to $x_f(\theta_d, p)$ which goes to infinity as $\theta_d \to \infty$. Furthermore, the function $\Pi_i^d(\theta_d)$ 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, since $p_f \geq p_i$

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

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 $\Pi_i^d(\theta_d) > \Pi_f^d(\theta_d)$, provided $\theta_d > 0$. ■

Proof of Proposition 4

(i) is obvious. Furthermore, the entrepreneur $\overline{\theta}_d(p_i, p_f)$ must be indifferent between being formal or informal. Since informal (formal) entrepreneurs weakly prefer to buy from informal (formal) suppliers, we must have:

$$\overline{\theta}_d(p_i, p_f)\overline{x}^\alpha - p_i\overline{x} = \pi\left[\overline{\theta}_d(p_i, p_f)x_f^\alpha(\overline{\theta}_d(p_i, p_f)) - p_f x_f(\overline{\theta}_d(p_i, p_f))\right]. \tag{14}$$

Furthermore $F(\theta_d) = \theta_d\overline{x}^\alpha - p_i\overline{x} - \pi\left[\theta_d.x_f^\alpha(\theta_d) - p_f x_f((\theta_d))\right]$ must satisfy $F'(\theta_d(p_i, p_f)) \leq 0$.

Using the envelope theorem, it follows that

$$\overline{x}^{\alpha} \leq \pi x_f^{\alpha}(\overline{\theta}_d(p_i, p_f)). \tag{15}$$

Since $0 < \pi < 1$ and $0 < \alpha < 1$, $\overline{x} \leq \pi x_f(\overline{\theta}_d(p_i, p_f))$. ■

Equilibrium Existence (Second Model)

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

$$s_i = \int_0^{\frac{p_i \overline{y}}{\pi p_f}} \min\{\theta, \overline{y}\} g_u(\theta) d\theta \tag{16}$$

$$s_f = \int_{\frac{p_i \overline{y}}{\pi p_f}}^{\infty} \theta g_u(\theta) d\theta \tag{17}$$

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:[27]

$$s_i = \int_0^{\overline{\theta}_u} \theta g_u(\theta) d\theta \tag{18}$$

$$s_f = \int_{\overline{\theta}_u}^{\infty} \theta g_u(\theta) d\theta \tag{19}$$

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

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

the complement interval assigned to purchase in the formal sector.[28] In this case, a point $(x_i, x_f) \in X(p_i, p_f)$ if there exists a $\gamma \geq \overline{\theta}_d(p_i, p_f)$ such that:

$$x_i = \int_0^{\overline{\theta}_d(p_i,p_f)} x_i(\theta, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i,p_f)}^{\gamma} x_f(\theta, \frac{p_i}{\pi}) g_d(\theta) d\theta \tag{20}$$

$$x_f = \int_{\gamma}^{\infty} x_f(\theta, p_f) g_d(\theta) d\theta \tag{21}$$

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 :

$$x_i = \int_0^{\overline{\theta}_d(p_i,p_f,1)} x_i(\theta, p_i) g_d(\theta) d\theta \tag{22}$$

$$x_f = \int_{\overline{\theta}_d(p_i,p_f,1)}^{\infty} x_f(\theta, p_f) g_d(\theta) d\theta \tag{23}$$

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:

$$x_i = \int_0^{\gamma} x_i(\theta_d, p_i) g_d(\theta) d\theta \tag{24}$$

$$x_f = \int_{\gamma}^{\overline{\theta}_d(p_i,p_i,1)} x_i(\theta_d, p_i) g_d(\theta) d\theta + \int_{\overline{\theta}_d(p_i,p_i,1)}^{\infty} x_f(\theta_d, p_i) g_d(\theta) d\theta \tag{25}$$

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 present the proof of 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 the following preliminary result:

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

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

Proof: If $\pi p_f < p_i \leq p_f$ formal firms prefer to buy the formal good. Hence

$$\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} = -\pi x_f(\theta_d, p_f) \tag{26}$$

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

$$\frac{\partial \Pi_i^d(\theta_d)}{\partial p_i} = -x_i(\theta_d, p_i) \tag{27}$$

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 $\overline{\theta}_d(p_i, \frac{p_i}{\mu})$ we must establish the sign of:

$$\frac{1}{\mu}\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} - \frac{\partial \Pi_i^d(\theta_d)}{\partial p_i}. \tag{28}$$

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 $\pi p_i < p_f < p_i$,

$$\frac{1}{\mu}\frac{\partial \Pi_f^d(\theta_d)}{\partial p_f} - \frac{\partial \Pi_i^d(\theta_d)}{\partial p_i} = -\frac{\pi}{\mu}x_f(\theta_d, \frac{p_i}{\mu}) + x_i(\theta_d, p_i). \tag{29}$$

The marginal informal firm buys exactly $\overline{x}$. Hence, from Proposition 4

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

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

$$\int_0^\infty \theta g_u(\theta) d\theta. \tag{30}$$

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

$$\int_0^{\overline{\theta}_d(p_i,\frac{p_i}{\pi},1)} x_i(\theta,p_i)g_d(\theta)d\theta + \int_{\overline{\theta}_d(p_i,\frac{p_i}{\pi},1)}^{\infty} x_f(\theta,\frac{p_i}{\pi})g_d(\theta)d\theta \tag{31}$$

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 2 $x_f(\overline{\theta}_d, p_i/\pi) > x_i(\overline{\theta}_d, p_i)$, 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 (16) and (17) we obtain that the sum of the aggregate supplies is:

$$\int_0^{\frac{\mu\overline{y}}{\pi}} \max\{\theta,\overline{y}\}g_u(\theta)d\theta + \int_{\frac{\mu\overline{y}}{\pi}}^{\infty} \theta g_u(\theta)d\theta. \tag{32}$$

On the other hand, using equations (22) and (23), the sum of the aggregate demands equals:

$$\int_0^{\overline{\theta}_d(p_i,\frac{p_i}{\mu})} x_i(\theta_d,p_i)g_d(\theta)d\theta + \int_{\overline{\theta}_d(p_i,p_i/\mu)}^{\infty} x_f(\theta_d,\frac{p_i}{\mu})g_d(\theta)d\theta. \tag{33}$$

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.

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 $[\int_0^{\overline{\theta}_d(p_i,p_i/\pi)} x_i(\theta_d,p_i)g_d(\theta)d\theta, \int_0^{\overline{\theta}_d(p_i,p_i/\pi)} x_i(\theta_d,p_i)g_d(\theta)d\theta + \int_{\overline{\theta}_d(p_i,p_i/\pi)}^{\infty} x_f(\theta_d,p_i/\pi)g_d(\theta)d\theta]$. If these intervals overlap, at $p_i = p_i(\pi)/\pi$ then $(p_i(\pi), p_i(\pi)/\pi)$ 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 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 $[0, \int_0^{\overline{\theta}_d(p_i,p_f)} x_i(\theta_d, p_f) g_d(\theta) d\theta]$. Hence there always exists a unique μ^* such that $(p_i(\mu^*), p_i(\mu^*)/\mu^*, 1)$ is an equilibrium.

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] Bound, J., C. Brown and N. Mathiowetz, 2005, "Measurement Error in Survey Data" **in** *Handbook of Econometrics*, V.5., J. Heckman and E. Leamer (eds.), Amsterdam: North-Holland.

[4] Dabla-Norris, E., M. Gradstein and G. Inchauste, 2008, "What Causes Firms to Hide Output? The Determinants of Informality", *Journal of Development Economics*.

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

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

[7] DiPasquale, D. and E. Glaeser, 1999, "Incentives and Social Capital: Are Homeowners Better Citizens?", *Journal of Urban Economics*.

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

[9] Fajnzylber, P., W. Maloney and G. Montes Rojas, 2007, "Does Formality Improve Micro-Firm Performance? Quasi-Experimental Evidence from the Brazilian SIMPLES Program", World Bank Working Paper.

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

[11] 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.

[12] Heckman, J., 1978, "Dummy Endogenous Variable in a Simultaneous Equation System", *Econometrica*, **46**, pp.931-959.

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

[14] Johnson, S., D. Kaufmann and A. Shleifer, 1997, "The Unofficial Economy in Transition", *Brookings Papers on Economic Activity*, No.2.

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

[16] 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.

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

[18] 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.

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

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

[21] Rivers, D. and Q. Vuong, 1988, "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", *Journal of Econometrics*, **39**, pp.347-366.

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

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

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

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

[26] 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	48308	0.130	0.337
taxsub	1 = Tax Substitution	48314	0.179	0.384
largecl	1 = Large Client	48306	0.041	0.199
smallcl	1 = Small Client	48306	0.070	0.256
outsidehouse	1 = Outside Household	48310	0.640	0.480
n_worker	Number of Employees	48314	1.473	1.044
revenue	Revenue in Oct/2003 (R$ 1,000)	47570	2.077	6.276
otherjob	1 = Owner has Other Job	48288	0.125	0.330
bankloan	1 = Bank Loan	48292	0.062	0.241
education	Education Level (Owner)	48253	4.367	1.884
age	Age (Owner)	48314	41.026	12.313
gender	Gender (Owner)	48312	0.644	0.479
ho_num	Homeowner × Number of Rooms	48040	4.889	3.316
loginst	Log of Installations (R$)	39818	5.830	1.764
loginv	Log of Investments (R$)	8119	6.504	2.161
profit	Profit in Oct/2003 (R$ 1,000)	44707	0.771	4.514
sup_enf	Supplier Enforcement	47846	0.012	0.010
cl_enf	Client Enforcement	47846	0.010	0.010
logwage	Log of Mean Wage (R$ 1,000)	6491	-1.831	0.855
supplierformal	Formalization among Suppliers	47749	0.159	0.035
clientformal	Formalization among Clients	47846	0.127	0.039

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: Economic Sector

	Freq.	%	Description
1	5,130	10.62	Transformation and Mineral Extraction Industry
2	7,000	14.49	Construction
3	14,675	30.37	Retail and Repair Services
4	4,104	8.49	Lodging and Food Services
5	4,451	9.21	Transportation and Communications
6	3,125	6.47	Real Estate and Services
7	2,937	6.08	Education, Health and Social Services
8	4,693	9.71	Other Collective, Social and Personal Services
9	2,199	4.55	Other Activities

Table 4: Correlation Matrix

	taxreg	taxsub	largecl	smallcl	outsdhous	n_work	rev	otherjob	bkloan
taxsub	0.01	1.00							
largecl	0.12	-0.06	1.00						
smallcl	0.06	-0.16	-0.10	1.00					
outsidehouse	0.09	-0.05	0.01	-0.03	1.00				
n_worker	0.35	0.02	0.08	0.10	0.08	1.00			
revenue	0.29	0.05	0.17	0.03	0.09	0.30	1.00		
otherjob	-0.01	-0.06	-0.02	-0.06	0.01	0.04	-0.01	1.00	
bankloan	0.11	-0.02	0.01	0.00	-0.03	0.04	0.07	0.01	1.00
education	0.30	-0.16	0.08	0.10	0.06	0.17	0.13	0.24	0.06
age	0.03	0.02	0.01	0.01	0.00	0.04	0.05	0.01	0.00
gender	-0.07	0.05	0.06	0.08	0.10	-0.03	0.03	0.01	-0.06
ho_num	0.15	-0.06	0.05	0.03	0.05	0.08	0.10	0.06	-0.02
loginst	0.55	0.16	0.16	0.06	0.10	0.50	0.52	-0.05	0.15
loginv	0.38	0.01	0.12	0.09	0.07	0.29	0.27	0.06	0.16
profit	0.04	-0.08	0.07	0.04	0.05	0.06	0.37	-0.02	0.02
sup_enf	-0.02	0.03	-0.04	-0.03	-0.11	0.03	0.01	0.02	0.04
cl_enf	0.09	-0.06	0.03	0.16	-0.06	0.02	0.09	-0.01	0.04
logwage	0.33	-0.08	0.19	0.13	0.12	0.24	0.30	-0.08	0.05
clform	0.31	-0.04	0.10	0.15	0.05	0.08	0.20	-0.07	0.05
supform	0.04	-0.10	0.09	0.19	0.04	-0.01	0.03	-0.03	-0.05

Correlation Matrix (cont'd)

	education	age	gender	ho_num	loginst	loginv	profit	sup_enf	cl_enf
age	-0.12	1.00							
gender	-0.20	0.05	1.00						
ho_num	0.23	0.17	-0.06	1.00					
loginst	0.25	0.05	-0.02	0.12	1.00				
loginv	0.34	-0.02	-0.07	0.13	0.55	1.00			
profit	0.07	0.05	0.00	0.04	-0.04	0.07	1.00		
sup_enf	-0.11	-0.04	-0.03	-0.01	0.05	0.01	-0.06	1.00	
cl_enf	0.02	0.00	0.05	0.04	0.15	0.12	-0.02	0.63	1.00
logwage	0.21	0.09	0.08	0.09	0.55	0.37	0.05	-0.15	-0.03
clform	0.11	0.03	0.07	0.08	0.35	0.21	0.01	0.10	0.47
supform	0.10	0.01	0.13	0.07	0.01	0.04	0.05	-0.11	0.23

Correlation Matrix (cont'd)

	logwage	clform
clform	0.09	1.00
supform	0.08	0.54

Table 5: Probit Estimates

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**
outsidehouse	0.174**	0.021	0.178**	0.020
	(0.024)		(0.024)	
n_worker	0.407**	0.052	0.419**	0.050
	(0.012)		(0.012)	
revenue	0.051**	0.006	0.044**	0.005
	(0.005)		(0.004)	
bankloan	0.379**	0.062	0.348**	0.052
	(0.033)		(0.034)	
otherjob	-0.242**	-0.027	-0.275**	-0.028
	(0.033)		(0.033)	
education	0.192**	0.0246	0.175**	0.021
	(0.006)		(0.006)	
age	0.036**	0.005	0.041**	0.005
	(0.004)		(0.005)	
age^2	0.000**	0.000	0.000**	0.000
	(0.000)		(0.000)	
gender	0.148**	0.018	0.215**	0.025
	(0.020)		(0.021)	
ho_num	0.030**	0.004	0.028**	0.003
	(0.003)		(0.003)	
Sector Dummies	Yes		No	
Output Coeff.	No		Yes	
State Dummies	Yes		Yes	
N	47201		46749	
Pseudo-R^2	0.3634		0.3767	
$\chi^2_{(44)}$	5435.96			

1. Significance levels : † : 10% ∗ : 5% ∗∗ : 1%
2. Standard errors clustered by urban sector.

Table 6: Investment, Installations and Profits

Dep. Var. =	loginvperworker	loginstperworker	profit
	Coefficient	**Coefficient**	**Coefficient**
	(Std. Err.)	(Std. Err.)	(Std. Err.)
taxreg	0.649**	0.800**	0.680**
	(0.062)	(0.033)	(0.138)
outsidehouse	0.204**	0.289**	0.186**
	(0.045)	(0.017)	(0.053)
bankloan	0.737**	0.626**	0.061
	(0.059)	(0.026)	(0.116)
otherjob	-0.276**	-0.257**	-0.180†
	(0.058)	(0.022)	(0.099)
education	0.240**	0.127**	0.178**
	(0.013)	(0.005)	(0.016)
age	0.031**	0.067**	0.029**
	(0.010)	(0.003)	(0.008)
age^2	0.000**	-0.001**	0.000*
	(0.000)	(0.000)	(0.000)
gender	0.509**	0.355**	0.264**
	(0.044)	(0.015)	(0.044)
ho_num	0.030**	0.020**	0.017*
	(0.006)	(0.002)	(0.008)
revenue	0.018**	0.062**	
	(0.005)	(0.006)	
n_worker			0.408**
			(0.054)
N	7954	39176	44368
R^2	0.330	0.356	0.038
$F_{(44,\cdot)}$	68.51	300.16	20.82

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	0.007**	-0.013**
	(0.001)	(0.004)
taxreg × education	0.079**	0.040**
	(0.002)	(0.002)
outsidehouse	0.052**	0.040**
	(0.005)	(0.015)
revenue	0.017**	0.007**
	(0.002)	(0.001)
bankloan	0.108**	0.032*
	(0.010)	(0.017)
otherjob	0.013*	0.000
	(0.006)	(0.016)
age	0.004**	-0.001
	(0.001)	(0.003)
age^2	0.000**	0.000
	(0.000)	(0.000)
gender	0.023**	-0.017
	(0.004)	(0.011)
ho_num	0.005**	0.002
	(0.001)	(0.002)
logwage		0.006
		(0.010)
Sector Dummies	Yes	Yes
State Dummies	Yes	Yes
N	47201	6336
R^2	0.3	0.193
$F_{(\cdot,\cdot)}$	166.90	22.10

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.**
largecl	0.373**	0.061				
	(0.049)					
smallcl	0.168**	0.024				
	(0.035)					
supplierformal			2.803**	0.358		
			(0.294)			
clientformal					4.976**	0.618
					(0.296)	
outsidehouse	0.179**	0.022	0.167**	0.021	0.161**	0.02
	(0.024)		(0.024)		(0.024)	
n_worker	0.407**	0.052	0.407**	0.052	0.421**	0.052
	(0.012)		(0.012)		(0.012)	
revenue	0.049**	0.006	0.050**	0.006	0.046**	0.006
	(0.005)		(0.005)		(0.004)	
bankloan	0.381**	0.062	0.382**	0.062	0.361**	0.057
	(0.033)		(0.034)		(0.034)	
otherjob	-0.229**	-0.026	-0.238**	-0.026	-0.234**	-0.025
	(0.033)		(0.033)		(0.033)	
education	0.186**	0.024	0.184**	0.024	0.186**	0.023
	(0.006)		(0.006)		(0.006)	
age	0.035**	0.005	0.035**	0.005	0.037**	0.005
	(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	0.125**	0.015	0.114**	0.014	0.134**	0.017
	(0.020)		(0.021)		(0.020)	
ho_num	0.030**	0.004	0.029**	0.004	0.029**	0.004
	(0.003)		(0.003)		(0.003)	
N	47196		46654		46749	
Pseudo-R^2	0.3664		0.3657		0.3722	
χ^2	5491.36		5469.05		5597.23	

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

Table 9: IV Probit Estimates (Chain Effects)

Dep. Var. = taxreg	**Non-IV Coeff.** (Std. Err.)	**IV Coeff.** (Std. Err.)	**First Stage (IV)** Dep. Var. = lscl	**Coeff.** (Std. Err.)
lscl	0.296**	0.447**	educurbsec	0.096**
	(0.029)	(0.088)		(0.012)
outsidehouse	0.213**	0.213**	outsidehouse	0.007
	(0.024)	(0.024)		(0.021)
n_worker	0.482**	0.478**	n_worker	0.076**
	(0.010)	(0.010)		(0.012)
bankloan	0.423**	0.405**	bankloan	0.074*
	(0.033)	(0.033)		(0.034)
otherjob	-0.252**	-0.246*	otherjob	-0.174**
	(0.033)	(0.033)		(0.029)
education	0.204**	0.200**	education	0.089**
	(0.006)	(0.006)		(0.006)
age	0.038**	0.038**	age	0.006†
	(0.004)	(0.004)		(0.004)
age^2	-0.000**	-0.000**	age^2	0.000*
	(0.000)	0.000		(0.000)
gender	0.164**	0.150**	gender	0.498**
	(0.020)	(0.021)		(0.003)
ho_num	0.033**	0.033**	ho_num	-0.005†
	(0.003)	(0.003)		(0.003)
N	47,930	47,196		

1. Significance levels : † : 10% * : 5% ** : 1%
2. The regressions also control for state and sector.
3. The second regression uses the average level of education in the urban sector as an instrument.
4. IV results obtained as bivariate probit.
5. Standard errors clustered by urban sector.

Table 10: Probit Estimates (Enforcement)

Dep. Var. = taxreg	**Coeff.** (Std. Err.)	**Marg. Eff.**	**Coeff.** (Std. Err.)	**Marg. Eff.**
sup_enf	5.607**	0.724		
	(1.463)			
cl_enf			11.817**	1.510
			(1.294)	
outsidehouse	0.178**	0.022	0.177**	0.022
	(0.024)		(0.024)	
n_worker	0.407**	0.053	0.412**	0.053
	(0.012)		(0.012)	
revenue	0.051**	0.006	0.049**	0.006
	(0.005)		(0.004)	
bankloan	0.377**	0.062	0.373**	0.062
	(0.033)		(0.033)	
otherjob	-0.243**	-0.027	-0.238**	-0.027
	(0.033)		(0.033)	
education	0.192**	0.025	0.186**	0.024
	(0.006)		(0.006)	
age	0.035**	0.005	0.035**	0.004
	(0.004)		(0.004)	
age^2	0.000**	0.000	0.000**	0.000
	(0.000)		(0.000)	
gender	0.152**	0.019	0.141**	0.018
	(0.020)		(0.020)	
ho_num	0.030**	0.004	0.029**	0.004
	(0.003)		(0.003)	

N	46749	46749
Pseudo-R^2	0.3628	0.3649
$\chi^2_{(45)}$	5410.44	5482.02

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

Table 11: São Paulo (SP) × Rio Grande do Sul (RS) (1997)

	SP			RS		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.
outsidehouse	2726	0.66		2044	0.67	
n_worker**	2727	1.53	1.16	2045	1.61	1.18
revenue	2657	3.09	9.15	2022	2.94	6.90
bankloan	2722	0.07		2042	0.06	
taxreg**	2725	0.20		2045	0.27	
otherjob**	2723	0.10		2044	0.13	
education	2721	4.94	1.93	2043	4.83	1.90
age	2727	43.12	12.60	2045	42.87	12.31
gender	2727	0.64		2045	0.65	

** : Difference in means is significant at the 1% level.

Table 12: Probit Estimates (SIMPLES)

Dep. Var. = taxreg	**Coefficient** (Std. Err.)	**Marg. Eff.**
SP × 2003	0.221**	0.056
	(0.073)	
SP	-0.517**	-0.13
	(0.054)	
2003	-0.782**	-0.20
	(0.057)	
N	8576	
Pseudo-R^2	0.3057	
$\chi^2_{(21)}$	2761.74	

1. Significance levels : † : 10% * : 5% ** : 1%
2. Controls include `outsidehouse`, `n_employee`, `revenue`, `bankloan`, `otherjob`, `education`, `age`, `age`^{2}, `gender`, `homeown_numroom` and sector of activity.

Table 13: Probit Estimates (Tax Substitution)

	Full Sample		**Tax Sub = 1**	
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
	(Std. Err.)	(Std. Err.)	(Std. Err.)	(Std. Err.)
largecl	0.446**		0.01	
	(0.049)		(0.190)	
smallcl	0.265**		-0.425**	
	(0.036)		(0.121)	
lscl		0.336**		-0.282**
		(0.031)		(0.108)
taxsub_largecl	-0.385*			
	(0.187)			
taxsub_smallcl	-0.617**			
	(0.119)			
taxsub_lscl		-0.555**		
		(0.104)		
taxsub	0.345**	0.346**		
	(0.027)	(0.027)		
outsidehouse	0.202**	0.204**	0.215**	0.217**
	(0.024)	(0.024)	(0.046)	(0.046)
n_employee	0.398**	0.398**	0.336**	0.337**
	(0.012)	(0.012)	(0.022)	(0.022)
revenue	0.047**	0.048**	0.049**	0.049**
	(0.004)	(0.004)	(0.009)	(0.009)
bankloan	0.380**	0.378**	0.401**	0.397**
	(0.033)	(0.033)	(0.064)	(0.064)
otherjob	-0.223**	-0.224**	-0.248**	-0.248**
	(0.033)	(0.033)	(0.068)	(0.068)
education	0.196**	0.196**	0.174**	0.173**
	(0.006)	(0.006)	(0.013)	(0.013)
age	0.034**	0.034**	0.054**	0.054**
	(0.005)	(0.005)	(0.010)	(0.010)
age2	0.000**	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
gender	0.099**	0.097**	0.096*	0.094*
	(0.021)	(0.021)	(0.042)	(0.042)
homeown_numroom	0.030**	0.030**	0.046**	0.047**
	(0.003)	(0.003)	(0.007)	(0.007)
N	47196	47196	8440	8440
Pseudo-R^2	0.3717	0.3712	0.3261	0.3255
$\chi^2_{(47)}$	5800.83	5791.56	1420.92	1407.81

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

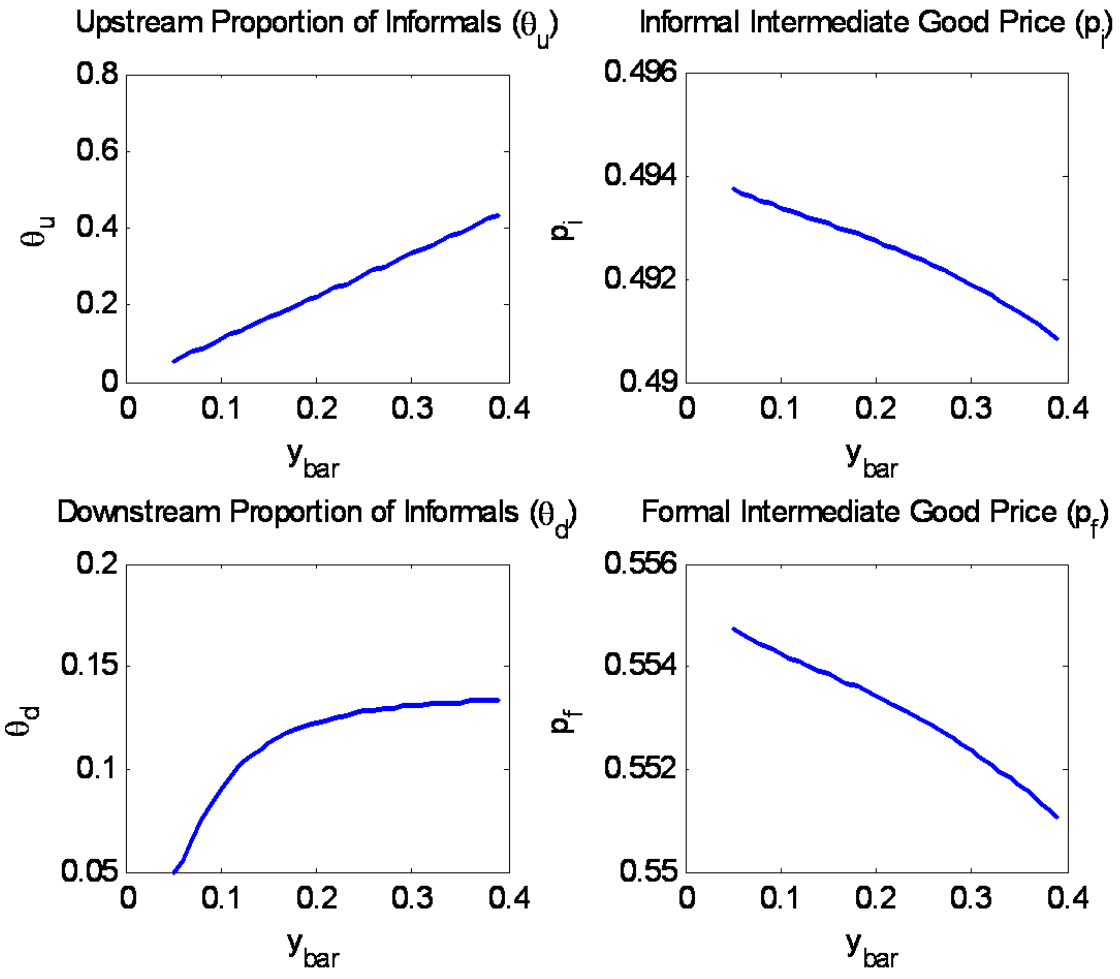

Figure 1: Comparative Statics