the review of income and wealth

Review of Income and Wealth
Series 57, Special Issue, May 2011

THE INFORMAL SECTOR: AN EQUILIBRIUM MODEL AND SOME EMPIRICAL EVIDENCE FROM BRAZIL

BY ÁUREO DE PAULA*

University of Pennsylvania

AND

JOSÉ A. SCHEINKMAN

Princeton University and NBER

We test implications of a simple equilibrium model of informality using a survey of 48,000+ small firms in Brazil. In the model, agents' ability to manage production differs and informal firms face a higher cost of capital and limitation on size, although these informal firms avoid tax payments. As a result, informal firms are managed by less able entrepreneurs, are smaller, and employ a lower capital–labor ratio. The model predicts that the interaction of an index of observable inputs to entrepreneurial ability and formality is positively correlated with firm size, which we verify in the data. Using the model, we estimate that informal firms in our dataset faced at least 1.3 times the cost of capital of formal firms.

JEL Codes: H2, H3, K4

Keywords: informal sector, tax avoidance, Brazil

1. Introduction

In this paper, we construct a simple equilibrium model of informality and test the implications of this model using a survey of 48,000+ small firms in Brazil.

Our model is a variant of Lucas (1978) and Rauch (1991). In the model, informality is defined as tax avoidance. Firms in the informal sector avoid paying taxes but have a limit on size. Firms use capital and labor, and informal firms face a higher cost of funds. This higher cost of capital for informal activities has been emphasized by De Soto (2000), who observed that because the right to assets held by the poor is not typically well documented "these assets cannot readily be turned

Note: Most of this paper appeared previously as part of our unpublished manuscript *The Informal Sector*. We thank Joana Monteiro and Juliano Assuncão for helping us with the ECINF dataset and 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.

*Correspondence to: Áureo de Paula, Department of Economics, University of Pennsylvania, Philadelphia, PA 19104, USA (aureo@sas.upenn.edu).

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

into capital . . . [and] cannot be used as collateral for a loan. . . ."[1] Agents differ in their managerial abilities. As in Rauch (1991), 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 size limits and face a lower cost of capital. The lower cost of capital also leads formal managers to choose a higher capital–labor ratio than informal entrepreneurs.[2]

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 Rauch (1991), Fortin *et al.* (1997), or Dabla-Norris *et al.* (2008), 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.[3]

Although we do not formally test this equilibrium model, several of its implications 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 (imperfect) 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. Finally, although our model assumes that all workers receive the same wage, in our data, after controlling for characteristics of the firm, formalization is correlated with higher wages. The correlation of formalization with higher wages and profits is an indication that informal firms produce less value added even after controlling for the entrepreneur's quality, although it is possible that we are simply missing important aspects of quality.

The model predicts a correlation between a manager's ability and size of firm. If we could measure ability perfectly, formality should give no additional information concerning size, once we condition on a manager's ability. However, ability is not observable and only variables imperfectly correlated with a manager's ability, such as his educational achievement, are observable. We prove that a regression of the size of a firm on observed variables that are positively correlated with ability and the interaction of this variable and formality should produce positive coefficients.

[1]De Soto (2000, pp. 5–6). De Soto (1989) estimates that in June 1985, informal firms in Lima (Peru) faced a nominal interest rate of 22 percent per month, while formal firms paid only 4.9 percent per month. We estimate a much smaller, but still significant, difference in capital costs between informal and formal firms in our sample. Straub (2005) develops a model in which a dual credit system arises in equilibrium.

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

[3]This implication is supported by the results from a survey of informal Mexican firms conducted by David Mckenzie and Christopher Woodruff that is reported in Fajnzylber *et al.* (2011), where 75 percent of the respondents reported that they were too small to make it worth their while to become formal.

This implication is supported by our empirical results. We also use the model to estimate the relative capital cost of informal and formal firms.

The model in this paper does not aim at providing a complete explanation for informality. It overlooks many other reasons for choosing informality such as regulations, labor taxes, the minimum wage, or the informality of a firm's clients or suppliers.[4] It also ignores that formal firms have greater access to the legal system and other civic institutions. There is also a vast literature on labor informality, which is not addressed in this paper. Finally, our model ignores partial compliance: firms pay their taxes either in full or not at all. This matches our data, which only provides binary information on formalization.

Related papers on informality include Loayza (1996), Johnson, *et al.* (1997), and Friedman *et al.* (2000), who 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). Assunção and Monteiro (2005) and Fajnzylber *et al.* (2011) use an earlier (1997) version of the survey that we employ in this paper. Both papers explore the institution of the federal SIMPLES, which simplified and reduced rates for tax compliance for small firms in Brazil. Although our empirical results speak to a different set of questions, and use data from a different year (2003) and a different definition for formalization,[5] their results largely agree with ours. They both find that enactment of SIMPLES increased formality. Fajnzylber *et al.* (2011) show that formalization is associated with increased use of labor and capital, and with higher productivity, which agrees with the predictions of our model. They fail to obtain significant effects on formalization from access to formal credit markets.[6]

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

2. A Model of Informality

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.

[4]Some evidence actually indicates that minimum wages may be as binding (if not more) in the informal sector than in the formal segment of the economy in Latin America (see Maloney and Mendez, 2004). On the role of value added taxes in creating informality chains, see de Paula and Scheinkman (2010).

[5]Assunção and Monteiro (2005) and Fajnzylber *et al.* (2011) use municipal licensing as proxy for formalization instead of tax registration, the measure we use. Assuncão and Monteiro recognize that tax registration would be a more appropriate indication of formalization, but opt for licensing because the question on tax registration 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.

[6]In our empirical work we use a broad interpretation of credit—40 percent of those who claimed to have obtained loans (and 25 percent of the formal entrepreneurs that claimed loans) received their loans from non-bank sources. In addition, Fajnzylber *et al.* (2011) focus 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 were much less developed than in 2002.

Each agent chooses between becoming a worker, and 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 α, $\beta > 0$ and $\alpha + \beta < 1$.

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

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

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

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

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

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

whereas if he chooses to enter the formal sector profits are:

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

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 those of the formal firms, the difference being greater the larger is the difference in capital costs between informal and formal firms ($r_i - r_f$). In Section 3 we provide evidence in support of the predicted difference in capital–labor ratios between formal and informal firms.

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

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

and, for informal firms that are not constrained:

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

If $1-\tau \geq \left(\frac{r_f}{r_i}\right)^{\alpha}$, taxes are too low with respect to the capital cost wedge and every entrepreneur prefers to be formal. Since we are interested in the informal sector we assume from now on that $1-\tau < \left(\frac{r_f}{r_i}\right)^{\alpha}$. In this case, every entrepreneur θ for which the optimal choice in the informal sector is unconstrained will prefer to be informal. Let $\underline{\theta}$ be the lowest value of θ for which an informal entrepreneur would choose a capital stock $\bar{k}$. For $\theta > \underline{\theta}$ the informal entrepreneur would keep $k = \bar{k}$ and, as a consequence, in this range:

(7) $$\frac{d\Pi_i}{d\theta}(\theta) = c\theta^{\beta/(1-\beta)},$$

for some constant c. Comparison of this last expression with equation (5) above shows that there exists a unique $\bar{\theta}$ such that $\Pi_i(\theta) < \Pi_f(\theta)$ if and only if $\theta > \bar{\theta}$.

Each agent also has the choice of becoming a worker and receiving the market wage w. Usual arguments in this class of models guarantee the existence of unique occupational choice cutoff points ($\hat{\theta}$ and $\bar{\theta}$). They are implicitly defined by:

(8) $$\Pi_f(\bar{\theta}) = \Pi_i(\bar{\theta})$$

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

and optimal choices are:

$$\theta \leq \hat{\theta} \Rightarrow \text{worker};$$

$$\theta \in (\hat{\theta}, \bar{\theta}] \Rightarrow \text{informal entrepreneur};$$

$$\theta > \max\{\bar{\theta}, \hat{\theta}\} \Rightarrow \text{formal entrepreneur}.$$

The determination of the cutoff points is illustrated in Figure 1. In the graph, we plot the optimal profit functions for formal and informal entrepreneurs by

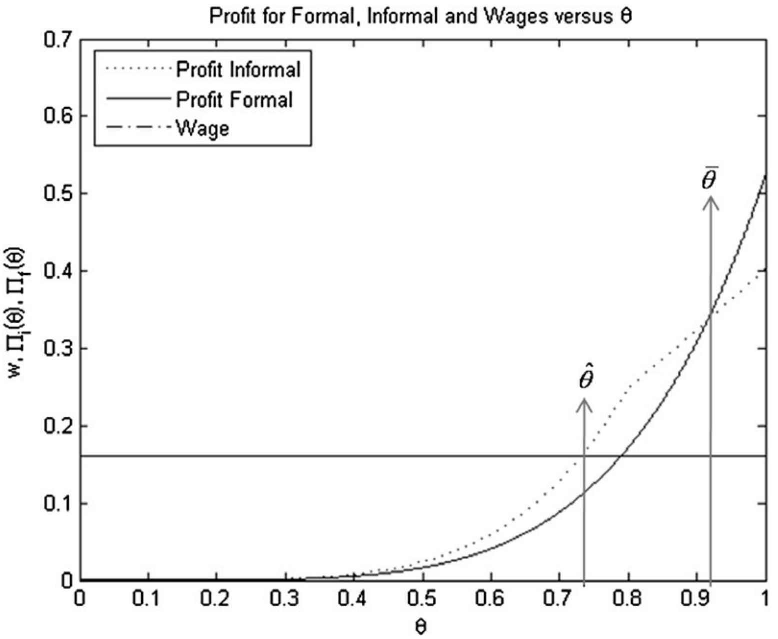

Figure 1. Equilibrium Cutoff Points

entrepreneurial ability and wage (at the equilibrium level). The cutoff between workers and informal entrepreneurs ($\hat{\theta}$) occurs where wage equals the informal profit function and the cutoff between informal and formal entrepreneurs ($\bar{\theta}$) occurs where the two profit functions intersect.

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

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

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 and formal firms will exist as long as the support of g is large enough. Informal firms exist as long as $\bar{\theta} > \hat{\theta}$. We assume this in what follows.

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 $\bar{\theta}$. This discontinuity follows because an entrepreneur with ability just below $\bar{\theta}$ chooses the informal sector and employs exactly $\bar{k}$ units of capital, although the marginal product of capital exceeds his cost of capital. At a level just above $\bar{\theta}$, an entrepreneur chooses the formal sector and since he is now unconstrained, he would choose a level $k >> \bar{k}$. Furthermore, since we assumed that $r_i(1-\tau)^{\frac{1}{\alpha}} < r_f$ and $\Pi_i(\bar{\theta}) = \Pi_f(\bar{\theta})$, we have

$$\begin{aligned}\Pi_i(\bar{\theta}) &= \bar{\theta} l_f(\bar{\theta})^{\beta} k_f(\bar{\theta})^{\alpha}(1-\tau) - w l_f(\bar{\theta}) - r_f k_f(\bar{\theta}) \quad (\equiv \Pi_f(\bar{\theta}))\\ &< \bar{\theta} l_f(\bar{\theta})^{\beta} k_f(\bar{\theta})^{\alpha}(1-\tau) - w l_f(\bar{\theta}) - r_i k_f(\bar{\theta})(1-\tau)^{1/\alpha}\end{aligned}$$

where the inequality follows because $r_i(1-\tau)^{\frac{1}{\alpha}} < r_f$. This suggests that an informal entrepreneur would attain higher profits if she were free to employ $l = l_f(\bar{\theta})$ and $k = k_f(\bar{\theta})(1-\tau)^{1/\alpha}$. Consequently, it should be the case that $k_f(\bar{\theta})(1-\tau)^{1/\alpha} > \bar{k}$ and

$$\left(\frac{\bar{\theta}(1-\tau)\beta k_f(\bar{\theta})^{\alpha}}{w}\right)^{1/(1-\beta)} > \left(\frac{\bar{\theta}\beta\bar{k}^{\alpha}}{w}\right)^{1/(1-\beta)}. \tag{10}$$

The left (right) hand side of equation (10) is exactly the labor demand by a formal (informal) entrepreneur with quality $\bar{\theta}$. 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 formalization threshold $\bar{\theta}$. This discontinuity is illustrated in Figure 2.

The empirical analysis of this gap is complicated because we do not observe an entrepreneur's ability θ and the dataset we use has no information on interest rates paid. In order to account for these limitations we assume that entrepreneurial ability $\theta = x\ \exp(\varepsilon)$ 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

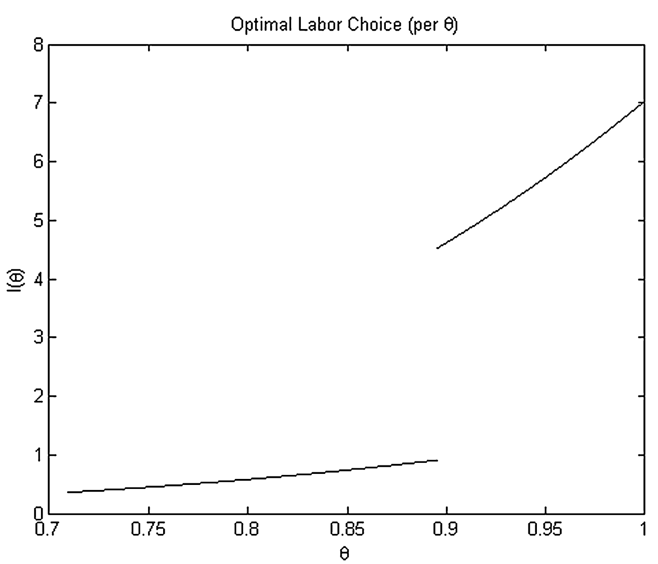

Figure 2. Discontinuity in Demanded Labor

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.

The left hand side of expression (10) provides the optimal labor demand for the marginal formal entrepreneur ($\theta = \bar{\theta}$). The same formula for the optimal labor demand applies for all formal entrepreneurs ($\theta > \bar{\theta}$). On the other hand, the right hand side provides the optimal labor demand for the marginal informal entrepreneur ($\theta = \bar{\theta}$). Similarly, if the optimal capital choice of an informal entrepreneur is $k_i(\theta)$, then the optimal labor choice is given by $\left(\frac{\theta\beta k_i(\theta)^{\alpha}}{w}\right)^{1/(1-\beta)}$. Taking logs of the optimality conditions for labor demand and letting $\theta = xe^{\varepsilon}$ vary, we get the following expression for ln l as a function of x and ε:

$$\text{(11)}\quad \ln l = \frac{1}{1-\beta}\ln\left[\frac{\beta}{w}\right] + \underbrace{\frac{1}{1-\beta}1_{xe^{\varepsilon}\geq\bar{\theta}}\ln(1-\tau)}_{\text{Formalization Effect}} + \underbrace{\frac{1}{1-\beta}(\ln x+\varepsilon)}_{\text{Direct Effect}} + \underbrace{\frac{\alpha}{1-\beta}\ln k(x,\varepsilon)}_{\text{Indirect Effect}}.$$

Managerial ability influences the demand for labor in three ways. A direct effect exists since this factor's marginal product is higher under better management. An indirect effect occurs because a better manager will also install more capital, driving up labor's marginal productivity. However, this indirect effect will not be present for the more skilled informal managers since these are capital constrained. A third effect, which we call "formalization effect" and is local to $\bar{\theta}$, occurs as entrepreneurs become formal and start paying taxes. This exerts a negative effect on the demand for labor which is nonetheless outweighed by the other two effects as we have shown above.

If one estimates a linear regression of ln l on ln x and an interaction between ln x and formalization ($\theta \geq \bar{\theta}$) 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 $1_{xe^{\varepsilon}\geq\bar{\theta}}\cdot$ ln x in the population. We represent this object as

$$\mathbb{E}^{BLP}\left[\ln l \middle| \ln x, \mathbf{1}_{xe^{\varepsilon}\geq\bar{\theta}}\cdot\ln x;\, xe^{\varepsilon}\geq\hat{\theta}\right] = \xi_0 + \xi_1\ln x + \xi_2\mathbf{1}_{xe^{\varepsilon}\geq\bar{\theta}}\cdot\ln x$$

where the conditioning event $xe^{\varepsilon} \geq \hat{\theta}$ 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* ln $x_i \geq 0$, *for $i = 1, \ldots, n$, with at least one non-zero element, then* $\xi_2 > 0$.

3. Empirical Application

In this section we explore implications of our theoretical framework using a dataset on informal firms in Brazil. Tax non-compliance is an important phenomenon in Brazil. Schneider and Enste (2000) estimate that informality represents more than one-quarter of the Brazilian economy.

3.1. *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 48,803 entrepreneurs in urban regions from all states in the Brazilian federation.[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 (census tracts) were stratified geographically in three strata (capital, other census tracts in the capital metropolitan area, and remaining census tracts). In a second step, the primary sampling units were stratified according to levels of income within the geographical stratum. Census tracts were then randomly selected with a probability proportional to the number of households in the sector. From each selected census tract 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; in fact, a vast majority declared that their firm was informal.

3.2. *Description of Variables*

We eliminated firms with owners who were less than 15 years old and the observations were lacking education or gender information. Entrepreneurs who claimed that their main client was a governmental institution, which comprised less than 1 percent 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.[10] *Outsidehouse* is a dummy that equals one when the activity is performed outside the home. The number of employees

[7]When an entrepreneur owns two firms, this corresponds to two observations in our sample. When a firm has two partners that live in the same household, this also corresponds to two observations.

[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 percent of the total number of firms, employed 36.2 percent of the workers, and responded for 10.3 percent of wages in 2002.

[9]For more information on the sampling strategies employed, see Almeida and Bianchini (1998).

[10]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 the correlation between the two measures of informality is 0.98.

TABLE 1

Variable Description

Variable	Description	Obs	Mean	Std. Dev.
taxreg	1 = tax registration	47,920	0.131	0.338
outsidehouse	1 = outside household	47,923	0.637	0.481
# employees	Number of employees	47,926	1.477	1.047
revenue	Revenue in Oct 2003 (R$1000)	47,188	2.091	6.299
otherjob	1 = owner has other job	48,290	0.125	0.330
education	Education level (owner)	48,255	4.367	1.884
age	Age (owner)	48,316	41.026	12.312
gender	Gender (owner)	48,314	0.644	0.479
ho_num	Homeowner × number of rooms	48,042	4.889	3.316
loginst	Log of installations (R$)	39,686	5.836	1.762
loginv	Log of investments (R$)	8,095	6.510	2.158
profit	Profit in Oct 2003 (R$1000)	44,406	0.775	4.529
logwage	Log of mean wage (R$1000)	6,485	–1.830	0.855

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

(*# employees*) includes the owner. Even though the survey focused on firms with five or less employees, a few units (less than 0.1 percent) 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* and *otherjob* 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; it 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$1000).[11] *Profit* equals revenue minus expenses in October 2003 (also in R$1000). *Logwage* denotes the logarithm of the total expenditures in salaries (in R$1000) divided by the number of employees in the firm.[12] The ECINF survey also has its own aggregate sectoral characterization, displayed in Table 3.

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

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

TABLE 3

Economic Sector

	Freq.	%	Description
1	5,106	10.65	Transformation and mineral extraction industry
2	6,916	14.43	Construction
3	14,605	30.47	Retail and repair services
4	4,092	8.54	Lodging and food services
5	4,427	9.24	Transportation and communications
6	3,106	6.48	Real estate and services
7	2,926	6.11	Education, health and social services
8	4,624	9.67	Other collective, social, and personal services
9	2,114	4.41	Other activities

TABLE 4

Probit Estimates

Dep. Var. = taxreg	Coeff. (Std. Err.)	Marg. Eff.
outside house	0.209**	0.026
	(0.024)	
# employees	0.492**	0.065
	(0.010)	
other job	–0.266**	–0.030
	(0.033)	
education	0.214**	0.028
	(0.006)	
age	0.040**	0.005
	(0.004)	
age^2	0.000**	0.000
	(0.000)	
gender	0.191**	0.024
	(0.020)	
homeowner × # rooms	0.004**	0.004
	(0.003)	
ECINF sector dummies	Yes	
State dummies	Yes	
N	47,572	
Pseudo-R^2	0.3333	
$\chi^2_{(42)}$	5,821.09	

Notes: Significance levels: †10%, *5%, **1%.
Standard errors clustered by census tract.

3.3. *Empirical Results*

Table 4 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 table suggests that higher ability entrepreneurs do tend to be formal. The coefficient of the variable "working outside the home" is positive. In agreement with our model, the coefficients are also positive for variables related to the size of the firm (number of employees), or the quality of the entrepreneurial input (education, age, or having no additional job). Since women in Brazil are

TABLE 5

Investment, Installations, and Profits

Dep. Var. =	loginvperworker	loginstperworker	profit
	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)
taxreg	0.851**	1.195**	0.682**
	(0.057)	(0.021)	(0.137)
outsidehouse	0.254**	0.353**	0.187**
	(0.046)	(0.018)	(0.053)
otherjob	–0.280**	–0.271**	–0.179†
	(0.058)	(0.023)	(0.100)
education	0.250**	0.152**	0.179**
	(0.013)	(0.005)	(0.016)
age	0.032**	0.072**	0.029**
	(0.010)	(0.003)	(0.008)
age^2	0.000**	–0.001**	0.000*
	(0.000)	(0.000)	(0.000)
gender	0.529**	0.401**	0.264**
	(0.043)	(0.015)	(0.044)
ho_num	0.030**	0.024**	0.017*
	(0.006)	(0.002)	(0.008)
# employees			0.409**
			(0.054)
N	8,038	39,412	44,087
R^2	0.301	0.284	0.038
$F_{(42,\cdot)}$	69.06	296.51	20.91

Notes: Significance levels: †10%, *5%, **1%.
The regressions also control for state and sector.
Standard errors are clustered by census tract.

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. The coefficients on all these variables are statistically significant.

Table 5 presents estimates 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 5 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 0.43 for investments per worker and 0.76 for installations per worker.[13] In other words, formalization is associated with an increase in investments (installations) per worker of 0.43 (0.76) standard deviations.

We also examined the correlation of formalization with profits. The results are summarized in Table 5. Again, after controlling for characteristics of the manager and the firm, formalization has a statistically significantly positive

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

TABLE 6

Log of Number of Workers (= Dep. Var.)

	Coefficient (Std. Err.)	Coefficient (Std. Err.)
ξ_2	2.033**	0.487**
	(0.091)	(0.045)
education	0.017**	0.015**
	(0.001)	(0.003)
other job	0.011*	–0.012
	(0.005)	(0.013)
age	0.009**	0.022**
	(0.001)	(0.002)
age^2	–0.000**	–0.000**
	(0.000)	(0.000)
gender	0.020**	0.160
	(0.004)	(0.010)
outsidehouse		0.002
		(0.001)
ho_num		0.004
		(0.008)
logwage		0.081**
		(0.014)
Sector dummies	Yes	Yes
State dummies	Yes	Yes
$\text{Min}(\mathbf{z}'\hat{\beta})$	0.013	0.015
N	47,841	6,430
R^2	0.4217	0.8597

Notes: Significance levels: †10%, *5%, **1%.

association with profits. Formalization is associated with an increase in monthly profits of approximately 700 Reais.[14]

Our model predicts that formalization is associated with a discrete jump in employment as a function of entrepreneurial ability. Table 6 displays supportive evidence for this "gap" (see Proposition 1). We assume that the observable measure of entrepreneurial quality is $x = \exp(\mathbf{z}'\beta)$ where $\mathbf{z}$ contains education, age, age squared, whether the entrepreneur has another job, and gender. Since these variables have finite support, the index also has finite support. Taking into account that, given our model for x, we can always choose $\xi_1 = 1$, we then estimate the regression

$$\ln(\text{number of workers}) = \mathbf{z}'\beta + \xi_2(\text{taxreg} \times \mathbf{z}'\beta) + \text{controls} + \varepsilon$$

which corresponds to the specification in Proposition 1. The controls include sector and state dummies and we also produce estimates including economic activity outside of the household, a wealth proxy (homeownership $\times$ number of rooms), and the log of the wage. Table 6 displays estimates using non-linear least squares. As Table 6 shows, $\mathbf{z}'\beta$ is positive. We thus apply Proposition 1 which implies that ξ_2 is positive as is indeed the case. The results remain if we introduce

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

additional controls (though gender and outsidehouse are now statistically insignificant).

3.4. *Cost of Capital*

In our 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 (net of taxes) and capital we would obtain:

$$\frac{\text{revenue}}{\text{installations}} = \frac{y+\varepsilon_y}{k+\varepsilon_k}$$

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:

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

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 $\hat{\Sigma}$. 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 individuals 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

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

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 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 yields similar estimates. Hence we estimate that, in our dataset, informal firms face a rate of interest that is at least 1.3 times the interest rate faced by formal firms.

4. Conclusion

In many developing countries, policies aimed at increasing incentives for formalization are viewed as an important step in augmenting aggregate productivity. This paper contributes to the growing body of evidence backing these policies. Predictions of our model—that informal firms are smaller, less productive, and employ less capital per worker—are supported by data from Brazil.

Programs of microcredit, directed at facilitating loans to predominantly informal entrepreneurs, are motivated by the perception that informal firms face an excessive cost of capital. In fact, using our model we estimate that informal firms in our sample faced at least 1.3 times the cost of capital of formal firms. Closing this gap would no doubt increase the use of capital by informal firms and augment the income of informal entrepreneurs. However, another implication of our model (cf. Section 2) is that these subsidies are also accompanied by an increased attraction of informality and associated losses of productivity.

Appendix: 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 $\varepsilon \sim G(\cdot)$ is independent of x and $supp(\varepsilon) = \mathbb{R}$.

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

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

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,

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

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

$$xe^{\bar{\varepsilon}} = \bar{\theta} \Leftrightarrow \bar{\varepsilon} = \ln \bar{\theta} - \ln x$$

and $\bar{\varepsilon}$ solve

$$xe^{\bar{\varepsilon}} = \hat{\theta} \Leftrightarrow \hat{\varepsilon} = \ln \hat{\theta} - \ln x.$$

The covariance can then be written as

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

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

$$\ln l(\theta, r, \tau) = \frac{1}{1-\beta}\ln\beta + \ln\theta + \frac{1}{1-\alpha-\beta}\ln(1-\tau) + \frac{\alpha}{1-\alpha-\beta}\ln\alpha - \frac{\alpha}{1-\alpha-\beta}\ln r - \frac{1-\alpha}{1-\alpha-\beta}w$$

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

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

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

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

and

$$xe^{\varepsilon} < \bar{\theta}\,(\Leftrightarrow \varepsilon < \bar{\varepsilon}) \Rightarrow \ln l(xe^{\varepsilon}) < \ln \bar{l}.$$

So the covariance should be

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

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

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

These being best linear projections,

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

and

$$\nu = \mathbf{1}_{xe^{\varepsilon} \geq \bar{\theta}} \cdot \ln x - \mathbb{E}\left(\mathbf{1}_{xe^{\varepsilon} \geq \bar{\theta}} \cdot \ln x \middle| xe^{\varepsilon} \geq \hat{\theta}\right) - \beta_1 \left[\ln x - \mathbb{E}\left(\ln x \middle| xe^{\varepsilon} \geq \hat{\theta}\right)\right]$$

where

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

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

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

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

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

Let $\bar{x} = \max supp(x)$ and $K = supp(x) - \{\bar{x}\}$. We can view x as a mixture of two distributions: with probability $\mathbb{P}(x = \bar{x})$ we sample from a distribution that delivers $\bar{x}$ 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

$$\begin{aligned} cov\left(\ln l, \left(\mathbf{1}_{xe^{\varepsilon} \geq \bar{\theta}} \ln x \middle| xe^{\varepsilon} \geq \hat{\theta}\right)\right) = \mathbb{E}\left\{cov\left(\ln l, \mathbf{1}_{xe^{\varepsilon} \geq \bar{\theta}} \ln x \middle| \mathbf{1}_K; xe^{\varepsilon} \geq \hat{\theta}\right) \middle| xe^{\varepsilon} \geq \hat{\theta}\right\} \\ + cov\left(\mathbb{E}(\ln l | \mathbf{1}_K), \mathbb{E}\left(\mathbf{1}_{xe^{\varepsilon} \geq \bar{\theta}} \ln x \middle| \mathbf{1}_K; xe^{\varepsilon} \geq \hat{\theta}\right) \middle| xe^{\varepsilon} \geq \hat{\theta}\right) \end{aligned}$$

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

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

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

References

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

Assunção, J. and J. Monteiro, "Coming Out of the Shadows: Estimating the Impact of Bureaucracy Simplification and Tax Cut on Formality and Investment," Working Paper, PUC-Rio, 2005.

Dabla-Norris, E., M. Gradstein, and G. Inchauste, "What Causes Firms to Hide Output? The Determinants of Informality," *Journal of Development Economics*, 85, 1–27, 2008.

de Paula, A. and J. Scheinkman, "Value Added Taxes, Chain Effects and Informality," *American Economic Journal: Macroeconomics*, 2, 195–221, 2010.

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

———, *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, Basic Books, New York, 2000.

Fajnzylber, P., W. Maloney, and G. Montes Rojas, "Does Formality Improve Micro-Firm Performance? Quasi-Experimental Evidence from the Brazilian SIMPLES Program," *Journal of Development Economics*, 94, 262–76, 2011.

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

Friedman, E., S. Johnson, D. Kaufmann, and P. Zoido-Lobaton, "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries," *Journal of Public Economics*, 76, 459–93, 2000.

Johnson, S., D. Kaufmann, and A. Shleifer, "The Unofficial Economy in Transition," Brookings Papers on Economic Activity: 2, 1997.

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

Lucas, R., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–23, 1978.

Maloney, W. and J. Mendez, "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*, The University of Chicago Press, Chicago, 2004.

Rauch, J., "Modelling the Informal Sector Formally," *Journal of Development Economics*, 35, 33–47, 1991.

Schneider, F. and D. Enste, "Shadow Economies: Size, Causes and Consequences," *Journal of Economic Literature*, 38, 77–114, 2000.

Straub, S., "Informal Sector: The Credit Market Channel," *Journal of Development Economics*, 78, 299–321, 2005.