Formal and Informal Firm Dynamics*

Gabriel Ulyssea
PUC-Rio

Preliminary, for discussion purposes only

Abstract

This paper develops a tractable framework to analyze firms' dynamics and life cycle behavior in developing countries. I build a stochastic model of industry dynamics that features burdensome regulations and taxes, imperfect enforcement and where heterogeneous firms can exploit two margins of informality: (i) whether or not to register their business (extensive margin); and (ii) whether formal firms hire informal workers (intensive margin). The model generates a broad array o new predictions regarding formal and informal firm dynamics. These predictions are tested and confirmed using micro data from formal and informal firms in Brazil.

*I am indebted to James Heckman, Steven Durlauf and Chang-Tai Hsieh for their advice, comments and constant encouragement. I would like to thank Azeem Shaik, Rafael Lopes de Melo, Ben Moll, Carlos Henrique Corseuil and Miguel Foguel for their comments and helpful conversations. I am also thankful to participants of the North American Meeting of the Econometric Society (Denver), the Capital Theory and Econometrics Working Groups at the University of Chicago, IPEA and LACEA-Labor (Cartagena) for their comments. Finally, I would like to thank Ricardo Paes de Barros for first encouraging me to pursue this project. Of course, all errors are mine. Financial support from CAPES, IPEA and The University of Chicago is gratefully acknowledged.

1 Introduction

The theoretical literature on firm dynamics is quite extensive and has been expanded in different fundamental directions [e.g. Jovanovic (1982), Hopenhayn (1992), Cooley and Quadrini (2001) and Asplund and Nocke (2006)]. Nonetheless, the existing frameworks do not seem readily applicable to the analysis of firm dynamics in developing economies. These countries display some distinguishing features that can substantially affect firms' behavior over the life cycle, in particular: restrictive and burdensome institutions, low enforcement, and a large informal sector.[1] These differences seem to be empirically and quantitatively relevant. Hsieh and Klenow (2012) present evidence that firms in India and Mexico exhibit striking differences in their life cycle patterns when compared to firms in the US, and these can explain a large fraction of the TFP differentials between these countries.

The aim of this paper is to develop a tractable framework to analyze firms' dynamics and life cycle behavior in developing countries. I build a stochastic model of industry dynamics that features burdensome regulations and taxes, imperfect enforcement and where heterogeneous firms can exploit two margins of informality: (i) the extensive margin – to register or not their business; and (ii) the intensive margin – whether registered firms hire informal workers to evade labor costs [Ulyssea (2013)]. The latter has been largely overlooked in the literature but is empirically relevant [e.g. Kumler et al. (2012)]. The model contains a novel entry structure that generates sorting between sectors based on productivity, and overlapping productivity distributions among entrants in both sectors, which is consistent with the data.

The combination of burdensome regulations, selection on productivity and imperfect enforcement drives the existence of both margins of informality. Because there is imperfect enforcement and larger firms are more visible to the government, the costs of both margins of informality are increasing in firms' size. This simple structure is in line with the available evidence on government's enforcement technology [e.g. Almeida and Carneiro (2012)] and immediately implies the following: (i) there is a cap to informal firms growth; and (ii) there are strong incentives for formal firms to remain small, as they can more easily evade labor costs. The model can thus offer a rationale for the fact that developing countries are often characterized by high (bureaucratic) entry costs, firm size distributions that are too skewed to the left and high turnover rates [see Tybout (2000)].

I show that there exists an unique stationary competitive equilibrium where both

[1] The informal sector comprises the majority of firms and accounts for 30-50% of GDP in developing countries. See, for example, Perry et al. (2007), Levy (2008) and de Mel et al. (2013).

sectors exist and have positive entry and exit. I then use the model to characterize the life cycle behavior of some key outcomes of formal and informal firms. Among these, the model predicts that the productivity and thus size distribution within a given cohort should be increasing in the cohort's age in both sectors. However, formal sector's age-size profile is everywhere above informal sector's. Similarly, the extensive and intensive informality margins within a given cohort are predicted to be decreasing with the cohort's age. These predictions are tested and confirmed using two unique, firm-level, matched employer-employee data sets from formal and informal firms in Brazil.

This paper builds primarily on two distinct literatures. First, it builds on Hopenhayn (1992) and it is thus related to the literature on firm dynamics that it has spanned [e.g. Melitz (2003) and Asplund and Nocke (2006)]. Second, on the (static) literature that sees informality as the result of heterogeneous firms responding optimally to the incentives provided by the institutional design they face. This view dates back to the seminal paper of Rauch (1991) [who builts on Lucas (1978)], and the literature that followed [among others, Fortin et al. (1997), Auriol and Warlters (2005), Dabla-Norris et al. (2008), de Paula and Scheinkman (2010) and de Paula and Scheinkman (2011)]. Additionally, the model developed here is related to D'Erasmo and Boedo (2012), who use a dynamic model to analyze the impact of capital market imperfections and formal sector's entry costs on total factor productivity.

The remaining of the paper is organized as follows. Section 2 presents the model and its main implications. Section 3 empirically analyzes the model's implications. Section 4 concludes.

2 The model

2.1 Static problem

The static problem is similar to other models available in the literature [e.g. Rauch (1991) and de Paula and Scheinkman (2011)]. The main modification is the introduction of the intensive margin of informality, i.e., that formal firms may hire workers without a formal contract.

There is a continuum of potential entrepreneurs that are indexed by their idiosyncratic productivity, $\theta \in \Theta \equiv [\theta_l, \theta_h]$, $\theta_l \geq 0$, which has a c.d.f. $G(\theta)$ defined over this support. These potential entrepreneurs must decide between starting a (formal or informal) firm, or being a worker and taking the common equilibrium wage (formal and informal firms hire workers in a competitive labor market). The informal firm avoids taxes but there is the possibility of being detected by government officials. If caught, it faces a penalty that can assume the form of a bribe, a fine or even being obliged to shut down. This expected cost assumes a very general form of a labor distortion term that increases the unit cost of labor. Informal firm's profit function is thus given by

$$\Pi_i(\theta, w) = \max_{\ell} \{\theta q(\ell) - \tau_i(\ell) w\} \tag{1}$$

where the function $q(\cdot)$ is a assumed to be increasing, concave, and twice continuously differentiable.

This profit function is a standard one except for the term $1 \leq \tau_i() < \infty$, which is the labor distortion mentioned above. It is assumed to be increasing and convex in firm's size: $\tau_i', \tau_i'' > 0$. Regardless of how one motivates the existence of this labor distortion,[2] the central assumption is that government's audits and enforcement technology imply a cap on informal firms size, above which they cannot grow.

Formal firms can choose how many workers to hire with and without a formal contract. Even though workers are homogeneous, hiring costs of formal and informal workers differ due to labor regulations: firms must pay a payroll tax on formal workers, while there is an expected cost to hire informal workers, which is increasing and convex on the number of informal workers hired. This cost is given by the function $\tau_{fi}(\cdot)$, $\tau_{fi}', \tau_{fi}'' > 0$ and $1 \leq \tau_{fi}(\cdot) < \infty$. The cost for formal firms of hiring informal workers is given by $\tau_{fi}(\ell)w$, while the cost of hiring formal workers is $(1+\tau_w) w\ell$, where τ_w is the labor tax.

Since formal and informal workers are perfect substitutes, on the margin firms hire the cheapest one. The marginal cost of hiring informal workers is strictly increasing, $C_{fi}'(\ell) = w\tau_{fi}'(\ell)$, and the marginal cost of hiring formal workers is constant, $C_{ff}' = (1+\tau_w) w$. Hence, there is an unique value of ℓ such that $C_{fi}'(\tilde{\ell}) = C_{ff}'$, above which formal firms only hire formal workers (on the margin). If the labor quantity that maximizes formal firm's profit is such that $\ell^* \leq \tilde{\ell}$, then the formal firm will only hire informal workers. If $\ell^* > \tilde{\ell}$, then the firm will hire $\tilde{\ell}$ informal workers and $\ell^* - \tilde{\ell}$

[2] A model commonly used, for example, is to assume that the firm has to shut down its operations if caught and the probability of detection is increasing and convex in firm's size (but bounded between zero and one, see Appendix A.1). This probability of detection is typically modeled as the analogous to a revenue or profit tax. I choose the current parametrization for clarity, as it allows a direct comparison with the distortions faced by formal firms. This is without loss of generality, since labor is the only factor used and thus these two different formulations are isomorphic (see Appendix A.1).

formal workers. Formal firms' profit function can thus be written as follows:

$$\Pi_f\left(\theta,w\right)=\max_{\ell}\left\{\left(1-\tau_y\right)\theta q(\ell)-C\left(\ell\right)\right\} \tag{2}$$

where τ_y denotes the revenue tax and

$$C\left(\ell\right) = \begin{cases} \tau_{fi}\left(\ell\right)w, & \text{for } \ell\leq\tilde{\ell} \\ \left[\tau_{fi}\left(\tilde{\ell}\right)+\left(1+\tau_w\right)\left(\ell-\tilde{\ell}\right)\right]w, & \text{for } \ell>\tilde{\ell} \end{cases} \tag{3}$$

2.1.1 Equilibrium

The solution for informal firm's problem is straightforward. Conditional on sector choice, the profit function and the demand for labor, $\Pi_s(\theta,w)$ and $\ell_s^*\equiv\ell_s\left(\theta,w\right)$, $s=i,f$, respectively, are increasing in θ.

As it is usual in this class of models, the equilibrium is characterized by a cutoff solution. Allowing for the intensive margin of informality thus has important qualitative implications, but none from the point of view of equilibrium determination.[3] As long as the productivity space is large enough, formal and informal firms will exist with positive probability. The cutoffs that determine occupational choice are given by $\underline{\theta}$ and $\overline{\theta}$ such that $\Pi_i(\underline{\theta})=w$ and $\Pi_i(\overline{\theta})=\Pi_f(\overline{\theta})$. The optimal choice rule is thus given by:

1. If $\theta<\underline{\theta}$, the entrepreneur decides to be a worker.
2. If $\theta\in\left[\underline{\theta},\overline{\theta}\right)$, the entrepreneur produces in the informal sector.
3. If $\theta\in\left[\overline{\theta},\theta_h\right]$, the entrepreneur produces in the formal sector.

Finally, the equilibrium wage, w^*, must be such that given the above cuttoffs, the labor market clears:

$$\int_{\underline{\theta}}^{\overline{\theta}}\ell_i^*\left(\theta,w^*\right)dG(\theta)+\int_{\overline{\theta}}^{\theta_h}\ell_f^*\left(\theta,w^*\right)dG(\theta)=\int_{\theta_l}^{\underline{\theta}}dG(\theta) \tag{4}$$

[3]The only consequence of having the intensive margin is to have a point $\left(\tilde{\ell}\right)$ where the profit function is not differentiable (although it is continuous everywhere).

2.2 Dynamics

2.2.1 Set up

The general features of the static set up are carried over to the dynamic model: firms produce an homogeneous good; product and labor markets are competitive and formal and informal firms face the same prices.[4] I do not explicitly model the demand side of this economy nor the factor supply. Instead, I assume the existence of well-behaved aggregate demand and labor supply functions, which are characterized as follows:

Assumption (A.1): (i) D is continuous, strictly decreasing aggregate demand function, and $\lim_{x\rightarrow\infty} D(x) = 0$; (ii) There is an aggregate (inverse) supply function $W(L)$ that is continuous, nondecreasing and strictly bounded above zero.

The dynamics at the firm level is driven by the evolution of the idiosyncratic productivity parameter, which is a source of uncertainty to the firm. The stochastic process that describes the evolution of the individual productivity is assumed to be a first order Markov process, and it is the same in both sectors and independent across firms, with conditional distribution $F(\theta'|\theta)$. I make the following assumptions regarding the conditional distribution of productivity shocks:

Assumption (A.2): (i) F is continuous in θ and θ'; (ii)F is strictly decreasing in θ; (iii) for any $\epsilon > 0$, there exists an integer n such that $F^n(\epsilon|\theta) > 0$, where $F^n(\cdot|\theta)$ gives the distribution of θ_{t+n} given $\theta_t = \theta$.

Assumption (A.2.ii) implies that higher productivity in the current period increases the probability of higher productivity in the following period. Assumption (A.2.iii) implies that there is always a positive probability of a firm receiving a bad shock in the future, regardless of its present productivity. This assumption, together with the optimal exit policy, implies that no firm lives forever in probability.

Analogously to the static model, I assume that all entrepreneurs can opt to get the equilibrium wage forever if they choose not to be entrepreneurs and thus their outside option is given by the market wage. The most common assumption in the literature is to treat this outside option as a exogenous per-period fixed cost. This set up,

[4]Some authors argue that the formal and informal sectors operate in completely separate markets. However, as argued by Ulyssea (2013), formal and informal firms coexist even within narrowly defined markets (4-digit industries). Hence, to assume that firms face the same output price seems like a reasonable approximation.

however, creates another channel through which market conditions can affect firms exit decisions. The period return functions in both sectors are the profit functions described in 1 and 2 net of the outside option: $\pi_s(\theta, w) = \Pi_s(\theta, w) - w$, $s = i, f$.

Entrants in both sectors must pay a fixed cost of entry, which is assumed to be higher in the formal sector, $c_f^e >> c_i^e$. For informal to formal transitions and the initial sorting between sectors, the relevant parameter is the difference between these entry costs, thus one could normalize c_i^e to zero. However, the difference between these entry costs is interpreted here to be exclusively a consequence of the regulation of entry into the formal sector (e.g. red tape bureaucracy). Thus, informal sector's entry cost can be seen as the initial investment or minimum scale required to operate in the given industry.

2.2.2 Incumbent's problem

Formal incumbent's problem is similar to the one in Hopenhayn (1992), as it faces a simple stopping-time problem. If it stays active, the firm must solve the static problem described in 2. Informal incumbents have an additional option, as they can either shut down, stay informal or make the transition to the formal sector. To make the transition to the formal sector, firms must pay the differential entry cost, $\tilde{c}^e = c_f^e - c_i^e$, and will also face the regulatory burden implied by formality in the following periods.

As there is no industry-wide shock and there is a continuum of (atomless) firms, uncertainty washes out at the aggregate level. Hence, when making their decisions firms take the sequence of future wages as given and the only source of heterogeneity is firm's own future productivity level. Let the sequence of future wages prices be denoted by $\mathbf{w}$. The value function of formal and informal *incumbents*, respectively, can be written in recursive form as follows:

$$V_f(\theta, \mathbf{w}) = \pi_f(\theta, w) + \beta \max\left\{0, E\left[V_f(\theta', \mathbf{w}) \middle| \theta\right]\right\} \quad (5)$$

$$V_i(\theta, \mathbf{w}) = \pi_i(\theta, w) + \beta \max\left\{0, E\left[V_i(\theta', \mathbf{w}) \middle| \theta\right], E\left[V_f(\theta', \mathbf{w}) \middle| \theta\right] - \tilde{c}^e\right\} \quad (6)$$

where β denotes the discount factor and $E\left[V_s(\theta', \mathbf{w}) \middle| \theta\right] = \int V_s(\theta', \mathbf{w})\, dF(\theta' | \theta)$, $s = i, f$.

It is worth noting that 5 and 6 give the firm's value function after the current productivity (θ) is realized. The exit decision, however, is made before future productivity is revealed, and it follows a cut-off rule: as soon as productivity falls below the threshold, $\theta < \underline{\alpha}_s$, firm θ leaves the industry. This threshold is determined as follows:

$$\underline{\alpha}_s = \inf \left\{ \theta \in \Theta : \int V_s\left(\theta', \mathbf{w}\right) dF\left(\theta' \middle| \theta\right) \geq 0 \right\} \tag{7}$$

and $\underline{\alpha}_s = \theta_h$, $s = i, f$, if the above set is empty.

The informal to formal transition is determined by an analogous cut-off rule, where the threshold is determined as follows:

$$\overline{\alpha}_i = \inf \left\{ \theta \in \Theta : \int \left[V_f\left(\theta', \mathbf{w}\right) - V_i\left(\theta', \mathbf{w}\right) \right] dF\left(\theta' \middle| \theta\right) \geq \tilde{c}^e \right\} \tag{8}$$

and $\overline{\alpha}_i = \theta_h$ if the above set is empty.

The *timing* of informal incumbents' decision process can thus be described as follows:

1. In the beginning of period they draw their productivity shock, θ, and decide how much to produce (i.e. how much labor to hire).

2. If $\theta \in \left[\underline{\alpha}_i, \overline{\alpha}_i\right)$, the firm will start next period in the informal sector again; if $\theta \geq \overline{\alpha}_i$, the firm will start next period in the formal sector and pay $\tilde{c}^e$; and if $\theta < \underline{\alpha}_i$, the firm will exit.

The timing for formal firms is exactly the same, except that they only have one option, which is to stay or exit the industry in the following period.

2.2.3 Entrants

Every period there is a mass M of potential entrants that only observe a pre-entry productivity parameter, denoted by $\nu \sim G$, where G is the c.d.f. defined over the support Θ. The ν is assumed to be i.i.d. and thus the mass of entrants in one period does not affect the composition of the potential entrants in the following period. If a potential entrant decides to pay the fixed cost of entry into either sector, she then draws her initial productivity from the conditional distribution $F(\theta|\nu)$. The expected value of entry for a firm with pre-entry signal ν is given by

$$V_s^e\left(\nu, \mathbf{w}\right) = \int V_s\left(\theta, \mathbf{w}\right) dF\left(\theta \middle| \nu\right), \qquad s = i, f \tag{9}$$

Entry is also characterized by threshold rule relative to the pre-entry signal parameter. These thresholds are determined as follows:

$$V_i^e\left(\overline{\nu}_i, \mathbf{w}\right) = c_i^e \tag{10}$$

$$V_f^e\left(\overline{\nu}_f, \mathbf{w}\right) = V_i^e\left(\overline{\nu}_f, \mathbf{w}\right) + \left(c_f^e - c_i^e\right) \tag{11}$$

where $\overline{\nu}_s$ is the pre-entry productivity of the last firm to enter sector $s = i, f$.

2.2.4 Stationary equilibrium

Total industry's size in any given period is described by a measure $\mu \equiv \mu(A)$, where $A \subseteq \Theta$ is the set of active firms in that period. Similarly, the size of sector s is given by $\mu_s \equiv \mu(A_s)$, where (A_i, A_f) constitute a disjoint partition of A and hence one can write $\mu = \mu_i + \mu_f$. These two measures evolve acording to the following laws of motions:

$$\mu_i'(\theta') = \int_{\theta \in I_1} F(\theta'|\theta)\, d\mu_i(\theta) + \underbrace{\int_{\nu \in [\underline{\nu}_i, \underline{\nu}_f)} F(\theta'|\nu)\, dG(\nu)\, M}_{\Lambda_i(\theta')} \tag{12}$$

$$\mu_f'(\theta') = \int_{\theta \geq \underline{\alpha}_f} F(\theta'|\theta)\, d\mu_f(\theta) + \underbrace{\int_{\nu \geq \underline{\nu}_f} F(\theta'|\nu)\, dG(\nu)\, M}_{\Lambda_f(\theta')} + \int_{\theta \geq \overline{\alpha}_i} F(\theta'|\theta)\, d\mu_i(\theta) \tag{13}$$

where $I_1 = [\underline{\alpha}_i, \overline{\alpha}_i)$ and Λ_s is the mass of entrants in sector $s = i, f$.

The aggregate production and labor demand, respectively, are determined by the following expressions:

$$Y = Y_i + Y_f = \int y_i(\theta, w)\, d\mu_i(\theta) + \int y_f(\theta, w)\, d\mu_f(\theta)$$
$$L^d = L_i^d + L_f^d = \int l_i(\theta, w)\, d\mu_i(\theta) + \int l_f(\theta, w)\, d\mu_f(\theta)$$

where $y_s(\theta, w)$ and $l_s(\theta, w)$ denote the optimal production and labor demand choices of a firm with productivity θ in sector $s = i, f$. With this notation in hand, I can go ahead and define the competitive stationary equilibrium:

Definition (D.1): A *stationary competitive equilibrium* in this model is a set of wage, allocations, cutoffs and sector and entry measures – $(w, Y_s, L_s, \underline{\alpha}_s, \overline{\alpha}_i, \underline{\nu}_s, \Lambda_s, \mu_s)$, $s = i, f$ – such that they remain constant over time and the following conditions hold in every period:

1. Labor market clears. [5]
2. The cutoffs $(\underline{\alpha}_s, \overline{\alpha}_i, \underline{\nu}_s)$, $s = i, f$, are determined according to (7), (8), (10) and (11), respectively.
3. Entry conditions (10) and (11) hold, with equality if $\Lambda_s > 0$.
4. μ_i and μ_f are determined by the laws of motion described in 23 and 24 and have a fixed point.

These equilibrium conditions are straightforward: condition 2 states that the cutoffs are chosen optimally, while condition 3 states that there are no further gains from entry.

Existence and uniqueness

The first main result of this section establishes the existence and uniqueness of a stationary equilibrium with entry and exit in the two sectors (Proposition 1). Although this result is obtained in the context of fairly simple model, it can be easily generalized to models that allow for many factors of production.

Proposition 1 (existence and uniqueness of the stationary equilibrium): *Assume the following: (i) (A.1) and (A.2) hold; and (ii) the profit function is such that q_s and l_s are continuous, single-valued and strictly increasing in θ (as in 1 and 2). Then there exists an unique stationary competitive equilibrium where both the formal and informal sector exist, and both sectors have positive entry and exit.*

Proof: See appendix.

In the following section I characterize the stationary equilibrium. For that, I focus on the analysis of the model's implications regarding the behavior of some key outcome variables over the firms' life cycle.

2.2.5 The life cycle o firms: size, productivity and informality

This model can generate predictions for the life-cycle behavior of a broad array of variables. In particular, in this section I focus on the behavior of productivity and size distributions by cohort in both sectors, as well as the evolution of both margins of informality in a given cohort.

[5] I normalize the price of the final good to one throught the analysis, but this is clearer way of stating the labor market conditions.

Let the productivity distribution among firms of age n in sector s be denoted by the probability measure λ_s^n. This measure is defined over the set of active firms with a given age, which is denoted by A_s^n. The evolution of the productivity distribution in the informal and formal sectors within a given cohort is given by:

$$\begin{aligned}\lambda_i^{n+1}\left(\theta'\right) &= \int_{X_1} F\left(\theta'\middle|\theta\right) d\lambda_i^n\left(\theta\right)\\ \lambda_f^{n+1}\left(\theta'\right) &= \int_{X_2} F\left(\theta'\middle|\theta\right) d\lambda_f^n\left(\theta\right) + \int_{X_3} F\left(\theta'\middle|\theta\right) d\lambda_i^n\left(\theta\right)\end{aligned}$$

for all $\theta' \in \Theta$; $X_1 \equiv A_i^n \cap \left[\underline{\alpha}_i, \overline{\alpha}_i\right)$, $X_2 \equiv A_f^n \cap \left[\underline{\alpha}_f, \theta_h\right)$, and , $X_3 \equiv A_i^n \cap \left[\overline{\alpha}_i, \theta_h\right)$.

Now consider the ordering on measures given by the the first order stochastic dominance criterion, so that $\lambda_s^{n+1} \succeq \lambda_s^n$ means that the productivity distribution at age $n+1$ first order stochastically dominates the one at age n. Then the first result of this section can be stated as follows:

Proposition 2: *Assume that (A.2) holds and that the pre-entry signal parameter has a continuous distribution, $G(\nu)$. Then the productivity distribution among active firms within a sector and a cohort is increasing in the cohorts age. That is, $\lambda_s^{n+1} \succeq \lambda_s^n$* for all n and $s = i, f$.

Proof: See appendix.

Corollary 1: *As the productivity distribution among active firms within a sector and a cohort is increasing in the cohorts age, so is the integral of any function that is increasing in θ. In particular, this is true for the survival rate, average size (in number of workers) and average revenue.*

Proof: Given the result in Proposition 2, this corollary follows mechanically and no proof is provided.

Corollary 1 is central, as it implies that for any given function that is increasing in productivity in the static model, (that is, in the cross-section), then it should also be increasing (within a cohort) in the cohort's age. Hence, it is possible to characterize the life cycle behavior of any outcome that depends on firms productivity based on the analysis of the static model. This result was already available in standard models of firm dynamics, but Proposition 2 and Corollary 1 show that it extends for a context where firms can also exploit the extensive and intensive margins of informality. Moreover, using the results established thus far, it is straightforward to prove the following results regarding the two margins of informality:

Corollary 2 (intensive margin of informality): *As a consequence of Proposition 2 and Corollary 1, the average informality rate within formal firms in a given cohort is strictly decreasing with cohort's age.*

Proof: See appendix.

In the static model it is straightforward to see that the share of informal workers is decreasing in firm's size and therefore productivity. Thus, by Proposition 1 and Corollary 1 one gets Corollary 2.

Finally, let the informality rate for a given cohort of age n be expressed as $\gamma_b^n = \frac{\mu(A_i^n)}{\mu(A_i^n)+\mu(A_f^n)}$, where $\mu(A_s^n)$ denotes the measure of the set of active firms in sector s with age n. With this notation in hand, the last result of this section can be stated as follows:

Corollary 3 (extensive margin of informality): *Under the conditions of Proposition 2, the informality rate γ_b^n within a given cohort is weakly decreasing in cohort's age. That is,* $\gamma_b^{n+1} \leq \gamma_b^n$ for all n.

Proof: See appendix.

The results presented by Propositions 1 and 2 and Corollaries 1-3 provide empirical predictions that can be tested in the data. In the next section I go ahead and assess whether these predictions are in fact observed in the data.

3 Empirical examination of the model's implications

In this section I compare the model's predictions with the data. The focus lies on the life cycle behavior of the key outcomes analyzed in Section 2.2: the behavior of firms' productivity, size, and both margins of informality over their life cycle. I start, however, by presenting the data used in Subsection 3.1 and analyzing some basic static implications from the model's mechanics, in Subsection 3.2. I then proceed to the core of this Section, which is the analysis of the model's dynamic implications, in Subsection 3.3.

3.1 Data

I use two data sets. The first is the ECINF survey (*Pesquisa de Economia Informal Urbana*), a repeated cross-section of small Brazilian firms (up to five employees) collected by the Brazilian Bureau of Statistics (IBGE) in 1997 and 2003. This is a matched employer-employee data set that contains information on the entrepreneurs, their business, and employees. The ECINF is representative at the national level for firms with at most five employees.[6]

Particularly relevant for the purposes of this paper, it includes information regarding the business registration at different government levels and also about the formalization of the firms' employees. I define as informal the firms not registered with the tax authorities (i.e. the extensive margin). The worker is classified as informal if she does not have the mandatory work permit (formal contract), which entitles her to all the benefits mandated by the labor legislation (this is the standard measure for formal labor contracts in Brazil). The informality rate within the formal firm (i.e. intensive margin) is defined as the share of wage workers in a given firm who do not have a formal contract.

Table 1 contains the descriptives statistics of the main variables of interest for formal and informal firms separately. The data confirm the standard regularities found in the literature [e.g. de Paula and Scheinkman (2010), de Paula and Scheinkman (2011) and La Porta and Shleifer (2008)]: (i) informal entrepreneurs are less educated; (ii) informal firms are smaller both in terms of employees and revenues; and (iii) wages and profits are much lower for informal firms. Interestingly, formal and informal firms have the same average age (measured in months), although one should keep in mind that this is a crude tenure measure, as it is obtained in the cross-section and cannot account for censoring.

The second data set to be used to complement the ECINF is the *Relacao Anual de Informacoes Sociais* (RAIS). It is an administrative data set collected by the Brazilian Ministry of Labor that contains the universe of formal firms and workers. It is available at an annual frequency, and it is a matched employer-employee panel of both workers and firms. Thus, the RAIS is useful not only for its panel dimension but also because it contains the entire size distribution of formal firms, whereas the ECINF is truncated at 5 employees. This is important, because although this size cap is not very much binding for informal firms, it certainly is for formal firms,

3.2 Static Implications

A central characteristic of the model discussed in Section 2 is that firms sort into both sectors based on productivity right upon entry. Measuring productivity is of course an extremely difficult task and there is an entire body of literature devoted to it. I use a very crude proxy for productivity that is readily computable from the data, which is simply value-added per worker. To obtain a slightly cleaner

[6] See de Paula and Scheinkman (2010) for a more detailed description of the ECINF data set.

measure, I regress the log of value-added on a set of industry dummies to purge inter-industry variation. The computed log-residuals are the productivity measure used. Figure 1 indicates that indeed firms sort based on productivity right upon entry. The figure shows that entrants in the formal sector are on average larger and the entire distribution is significantly shifted to the right. Nonetheless, the sorting is not perfect, as the two distributions overlap.

I now turn to the relationship between the informality margins and productivity. The model implies that both margins are decreasing in firms' productivity. To examine this implication in the data, I compute the share of informal firms and the average share of informal workers within formal firms in each percentile of the productivity distribution (using the measure defined above). Figure 2 shows that there is a strong negative association between the extensive margins and productivity, which is not so pronounced for the intensive margin. The latter result is likely to be driven by the size cap imposed by the ECINF sample, which is binding for formal firms.

The negative relationship between productivity and the extensive margin of informality has been largely documented in the literature. However, the same is not true for the intensive margin, which has been largely overlooked. Since the intensive margin accounts for a large fraction of total informal employment [Ulyssea (2013)], any model that aims at describing firms' choices regarding informality should be able to conform with these facts.

3.3 Dynamic Implications

Before discussing the results, it is worth highlighting that, since the ECINF is a cross-section, the estimates for informal firms inevitably confound true age effects with cohort effects. To minimize this problem, I restrict my analysis in the ECINF data to firms that are at most 10 years old, as these firms were born after the crucial economic reforms that took place in late 1980's and early 1990's (trade liberalization, constitutional reform and price stabilization, among the most important). This problem is not present in the RAIS data set, as it is a panel of formal firms.

Prediction 1: Size distribution is increasing in cohort's age

I first examine the prediction that the productivity distribution and thus the firm size distribution is stochastically increasing with cohort's age (Proposition 2). If the selection mechanism highlighted in the model plays a significant role in a given industry, the productivity distribution within a given cohort should improve as the cohort ages, as the least productive firms exit the industry and the more productive survive.

To test these predictions, I examine the size distribution in both sectors. For informal firms, I use log-revenue as a size measure, while for formal firms I use log-employment (the data set for formal firms does not contain information on sales). For informal firms, I compute the size distribution for different cohorts within the cross-section. For formal firms, I plot the size distributions of a single cohort (born in 1997), which is followed for 10 years (the results hold for any cohort that one might choose).

The distributions in Figure 3 are similar but statistically different from each other, according to the Kolmogorov-Smirnov test of equality of distributions. The graphs show that the cdf of older cohorts first order stochastically dominates the cdf of younger ones, which corroborates the predictions of Proposition 2.

Prediction 2: Average size within a cohort is increasing in cohort's age

As a consequence of Proposition 2 and Corollary 1, one should observe positive age effects on average. I first examine this claim by computing the age-size profile in the cross-section using the ECINF for informal firms and the RAIS for formal firms. I do so controlling for aggregate industry dummies (e.g. manufacturing and services), so these are within-industry profiles. Again, for informal firms I use log-revenues as the size measure and for formal firms I use the log of firm's employees.

As predicted in the model, firms' average size within the cohort is increasing (and concave) in cohort's age, both in the informal and formal sectors (Figure 4). The growth pattern for formal firms is somewhat similar to that observed in the US, although with a less pronounced growth rate: in the US, Hsieh and Klenow (2012) show in an analogous graph that firms' that are 15 years old are on average two times larger than firms that are 5 years old or younger, while in Brazil this ratio is around 1.6. Using the panel structure of the RAIS data set, I estimate a fixed-effects model to estimate age effects using within-firm variation. As Figure 6 shows, the same pattern arises but the growth rate is a little smaller than before, of around 1.49.

Finally, in order to have a comparable size measure in both sectors, I turn to the PNAD data set to compare the log-earnings of formal and informal entrepreneurs in the cross-section and across cohorts. The same increasing and concave pattern arises, but the age-size profile in the informal sector is everywhere below formal sector's, as also predicted in the model.

Prediction 3 (*Intensive Margin*): Average within-firm informality is decreasing in cohort's and firm's age

Corollary 2 implies that the average informality rate within-firms in a given cohort should be decreasing in cohort's age. To examine this claim, I plot the average informality rate within firms against the cohort's age. As Figure 7 shows, the average within-firm informality rate is decreasing in cohorts' age but not monotonically. It is worth highlighting however, that this graph is obtained using only data from the ECINF, which truncates the size distribution. This is likely to have a severe impact on the sample of formal firms, as formal firms grow much larger than informal ones as they age, which implies that this graph only captures the least productive older formal firms, which are those that did not grow above 5 employees as they aged. This size truncation thus goes against the prediction that Figure 7 aims to corroborate.

Looking at evidence from regressions, Table **??** shows conditional correlations between the intensive margin and firm's age. These are obtained by regressing individual formal firm's share of informal employees on its age, industry dummies and additional controls, which are a dummy for whether the entrepreneur conducts her business outside her home, if she needed capital to start her business and wether she used own capital to start her business. I do not control for entrepreneurs' characteristics and other business characteristics to stay as close to the theoretical set up as possible. These age effects appear exactly because entrepreneurs become more productive as they age, so the desired correlation is not conditional on entrepreneurs' quality. As the table shows, age effects are strongly negative and nonlinear (as already suggested by Figure 7).

Prediction 4 (*Extensive Margin*): Firm informality rate within the cohort is decreasing in cohort's age

Corollary 3 predicts that the share of informal firms within a cohort should be decreasing in the cohort's age. To examine this prediction, I plot the share of informal firms within each one-year cohort (up to 10 years old). As Figure 8 shows, the share of informal firms is decreasing in cohort's age, but not sharply decreasing (total variation is of 6 percentage points). Again, this effect should be seen as a lower bound of the age effect on the extensive margin, as the sample used is limited to firms with at most 5 employees. Since on average older firms are larger than younger ones, the sample design is likely to be excluding many formal firms that grew bigger as they aged and became more productive.

Finally, I use firm level data to run a similar regression as above. I estimate a probit with a dummy of whether the firm is informal as dependent variable, and using as controls firms' age, industry dummies and dummies for wether the entrepreneur conducts her business outside her home, if she needed capital to start her business and wether she used own capital to start her business. Figure 9 shows the marginal effects of firms' age obtained from this regression, which are negative and statistically significant, as predicted in the model.

4 Final remarks

This paper develops a tractable framework to analyze firms' dynamics and life cycle behavior in developing countries. I develop a model of industry dynamics with heterogeneous firms, endogenous entry and exit, and where firms can exploit the extensive and intensive margins of informality, respectively: (i) whether or not to register their business; and (ii) whether formal firms hire informal workers. The model aims at capturing the main trade-offs between these informality margins and being formal, with an emphasis on the institutional and dynamic aspects of these trade-offs.

The model delivers predictions regarding a variety of key outcomes, such as the life-cycle behavior of both margins of informality, firm's productivity, and size. Among these, the model predicts that productivity and thus size distributions within a given cohort should be increasing in the cohort's age in both sectors, but formal sector's age-size profile is everywhere above informal sector's. Similarly, both informality margins within a given cohort are decreasing with the cohort's age. These predictions are tested and confirmed using micro data from formal and informal firms in Brazil.

References

Almeida, R. and P. Carneiro (2012). Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics 4*(3).

Asplund, M. and V. Nocke (2006, 04). Firm turnover in imperfectly competitive markets-super-1. *Review of Economic Studies 73*(2), 295–327.

Auriol, E. and M. Warlters (2005, April). Taxation base in developing countries. *Journal of Public Economics 89*(4), 625–646.

Cooley, T. F. and V. Quadrini (2001). Financial markets and firm dynamics. *The American Economic Review 91*(5), pp. 1286–1310.

Dabla-Norris, E., M. Gradstein, and G. Inchauste (2008). What causes firms to hide output? the determinants of informality. *Journal of Development Economics 85*, 1–27.

de Mel, S., D. McKenzie, and C. Woodruff (2013). The demand for, and consequences of, formalization among informal firms in sri lank. *American Economic Journal: Applied Economics*. Forthcoming.

de Paula, A. and J. A. Scheinkman (2010). Value-added taxes, chain effects, and informalit. *American Economic Journal: Macroeconomics 2*(4), 195–221.

de Paula, A. and J. A. Scheinkman (2011). The informal sector: An equilibrium model and some empirical evidence. *Review of Income and Wealth 57*, S8–S26.

D'Erasmo, P. and H. Boedo (2012). Financial structure, informality and development. *Journal of Monetary Economics 59*(3), pp. 286–302.

Fortin, B., N. Marceau, and L. Savard (1997). Taxation, wage controls and the informal sector. *Journal of Public Economics 66*(2), 293–312.

Hopenhayn, H. A. (1992, September). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Hsieh, C.-T. and P. J. Klenow (2012). The life cycle of plants in india and mexico. Technical report, NBER Working Paper No. 18133.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrics 50*(3), pp. 649–670.

Kumler, T., E. Verhoogen, and J. Frias (2012). Enlisting workers in monitoring firms: Payroll tax compliance in mexico. Mimeo.

La Porta, R. and A. Shleifer (2008). The unofficial economy and economic development. *Brooking Papers on Economic Activity 105*(3), 473–522.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico.* Brookings Institution Press.

Lucas, R. (1978). On the size distribution of business firms. *Bell Journal of Economics 9*(2), 508–523.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), pp. 1695–1725.

Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi (2007). *Informality: Exit or exclusion.* The World Bank. Washington, D.C.

Rauch, J. E. (1991). Modeling the informal sector formally. *Journal of Development Economics 35*(1), 33–47.

Tybout, J. R. (2000). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic literature 38*(1), 11–44.

Ulyssea, G. (2013). Firms, informality and the welfare effects of formalization. Unpublished manuscript.

A Additional Results and Proofs

A.1 Correspondence between different formulations of the profit function

Instead of the profit function described in 1, consider the following representation:

$$\Pi_i(\theta) = \max_{l} \{[1 - p(\ell)]\, \theta q(\ell) - w\ell\}$$

where $0 < p(\cdot) \leq 1$ is strictly increasing and convex.

The $p(\ell)$ can be interpreted as the probability of being caught by government's officials, in which case the informal firm looses all of its production. To obtain a direct correspondence between this formulation and expression 1, one could parametrize the labor distortion as $\tau_i(\ell) = \frac{\ell}{1-p(\ell)}$, so that as $\lim_{p(\ell)\to 1} \tau_i(\ell) = \infty$ and $\lim_{p(\ell)\to 0} \tau_i(\ell) = \ell$.

A.2 Proof to Proposition 1

The proof of this proposition is organized in several steps.

Step 1: Properties of the profit and value functions

The assumption about the production and cost functions in both sectors implies that the profit functions are continuous and that q_j^* and l_j^* are continuous, single valued, and strictly increasing in θ. Write the corresponding equilibrium prices as

$$w^*(\mu) = W\Big(L^*(\mu, w(\mu))\Big) \tag{14}$$

It follows directly from Hopenhayn (1992), Lemma 3, that the function $w^*(\mu)$ is well defined and continuous. Using 14, one can re-write the value functions in each sector as follows:

$$V_f(\theta, \mu) = \tilde{\pi}_f(\theta, \mu) + \beta \max\left\{0, E\left[V_f(\theta', \mu)\,|\theta\right]\right\} \tag{15}$$

$$V_i(\theta, \mu) = \tilde{\pi}_i(\theta, \mu) + \beta \max\left\{0, E\left[V_i(\theta', \mu)\,|\theta\right], E\left[V_f(\theta', \mu)\,|\theta\right] - \tilde{c}^e\right\} \tag{16}$$

where $\tilde{\pi}_s(\theta, \mu) \equiv \pi_s(\theta, w^*(\mu))$.

Lemma 1: *The functions $\tilde{\pi}_s$ are continuous in both arguments, strictly increasing in θ and decreasing in μ.*

Proof: The proof follows directly from the proof in Hopenhayn (1992). Continuity follows from the continuity of the profit functions in θ and w, and from continuity of $w^*(\mu)$. Since $p^*(\mu) > 0$, $\forall\mu$, it also follows from the properties of the profit function that $\tilde{\pi}_s$ is increasing in θ.

It remains to show that $\tilde{\pi}_j$ is decreasing in μ. Let $\mu_2 > \mu_1$ with the corresponding equilibrium prices w_1^* and w_2^*. Suppose by way of contradiction that $w_2^* < w_1^*$. This implies that $q_s(\theta, w_2^*) > q_s(\theta, w_1^*)$ $\forall\theta$ and therefore $Q_s^*(\mu_2) = \int q_s(\theta, w_2^*)\, d\mu_2^j(\theta) > \int q_s(\theta, w_1^*)\, d\mu_1^j(\theta) = Q_s^*(\mu_1)$ (the last inequality follows because q_j is strictly increasing in θ), $s = i, f$. But if $Q^*(\mu_2) > Q^*(\mu_1)$, then $L^*(\mu_2) > L^*(\mu_1)$ and therefore $w_2^* > w_1^*$ [due to assumption (A.1)], which contradicts the initial assumption that $w_2^* < w_1^*$. Hence, if $\mu_2 > \mu_1$ then $w_2^* > w_1^*$. Hence, one can easily verify that $\tilde{\pi}_s(\theta, \mu_2) \equiv \tilde{\pi}_s(\theta, w_2^*) < \tilde{\pi}_s(\theta, w_1^*) \equiv \tilde{\pi}_s(\theta, \mu_1)$, which establishes the desired result. □

Given these properties of the profit functions, I can go ahead and establish the properties of the value functions:

Lemma 2: *The functions, V_s, that solve 15 and 16 are unique and have the following properties: (i) they are continuous functions; (ii) strictly increasing in θ and decreasing in μ; and (iii) the option value in both sectors (the integral terms in 15 and 16) is strictly increasing in θ.*

Proof: Given the assumption (A.2) and the properties of $\tilde{\pi}_j$ established in Lemma 1, the properties (i) and (ii) follow directly from standard dynamic programming arguments. The same can be said about (iii), except that it also relies on (ii). Note that the presence of V_f in the option value of the informal firm does not alter the argument, as the value functions in both sectors share the same properties. □

Step 3: Uniqueness of the cut-offs for entry, exit and informal to formal transitions.

Due to the properties of the value functions established in Lemma 2 (continuity and monotonicity), it follows that the exit and entry thresholds in both sectors, $\underline{\alpha}_s$ and $\underline{\nu}_s$, and the informal to formal transition cut-off, $\overline{\alpha}_i$, are uniquely determined by the following expressions:

$$\int V_s(\theta', \mu)\, dF(\theta' | \underline{\alpha}_s) = 0, \qquad s = i, f \tag{17}$$

$$\int V_i(\theta, z)\, dF(\theta | \overline{\nu}_i) = c_i^e \tag{18}$$

$$\int V_f(\theta, z)\, dF(\theta | \overline{\nu}_f) = V_i^e(\overline{\nu}_f, w) - (c_i^e - c_f^e) \tag{19}$$

$$\int [V_f(\theta', \mu) - V_i(\theta', \mu)]\, dF(\theta' | \overline{\alpha}_i) = \tilde{c}^e \tag{20}$$

Using assumption (A.2.ii) and the functional form assumed for the production and cost functions in each sector, one can verify that $\underline{\alpha}_f > \underline{\alpha}_i$. This will be the case even if at $\theta = \underline{\alpha}_f$ the formal firm hires all of its employees informally. Using a similar reasoning, $c_f^e > c_i^e$ implies that $\underline{\nu}_f > \underline{\nu}_i$.

Finally, condition 20 requires a more careful explanation. The analysis of the static model showed that there is a productivity threshold $\overline{\theta}$ above which $\pi_f(\theta) > \pi_i(\theta)$, for any $\theta > \overline{\theta}$. Hence, from the results proved in Lemmas 1 and 2 it follows that, if there is informal to formal transition, then the threshold $\overline{\alpha}_i$ defined by 20 is unique.

Step 4: The industry measures

Following Hopenhayn (1992), I define the operator $\hat{P}_k$ as follows:

$$\hat{P}_k \equiv \hat{P}_k(\theta, B) = \begin{cases} \int_B dF(x|\theta) & \text{if} \quad \theta \in I_k \\ 0 & \text{otherwise} \end{cases}$$

for all Borel sets $B \subset \Theta$; where $I_1 = [\underline{\alpha}_i, \overline{\alpha}_i)$, $I_2 = [\overline{\alpha}_i, \theta^H)$, and $I_3 = [\underline{\alpha}_f, \theta^H)$. This is a bounded $\left(\left\|\hat{P}_k\right\| \leq 1\right)$, linear operator on the space of positive bounded measures: $\hat{P}_k\mu(B) = \int \hat{P}_k(\theta, B)\, d\mu(\theta)$, for all Borel sets $B \subset \Theta$ [see Hopenhayn (1992)]. Using this notation, one can rewrite the law of motion of both sectors' measure as follows:

$$\mu_i' = \hat{P}_1\mu_i + \Lambda_i'\tilde{G}_i' \tag{21}$$

$$\mu_f' = \hat{P}_3\mu_f + \Lambda_f'\tilde{G}_f' + \hat{P}_2\mu_i \tag{22}$$

Assumption (A.2) and the definition of the sets I_k imply that $\left\|\hat{P}_k\right\| < 1$. Hence, the argument presented in Hopenhayn (1992, Lemma 4) follows directly and the operator $\left(\mathrm{I}-\hat{P}_k\right)$ has an inverse. Hence, the invariant measures μ_i and μ_f are welll-defined and can be written as

$$\mu_i = \left(\mathrm{I}-\hat{P}_1\right)^{-1}\Lambda_i\tilde{G}_i \tag{23}$$

$$\mu_f = \left(\mathrm{I}-\hat{P}_3\right)^{-1}\left(\Lambda_f\tilde{G}_f+\hat{P}_2\mu_i\right) \tag{24}$$

Write the above measures as $\mu_i = m_i\left(\underline{\alpha}_i, \overline{\alpha}_i, \Lambda_i\right)$ and $\mu_f = m_f\left(\underline{\alpha}_f\overline{\alpha}_i, \Lambda_f\right)$, where I use the fact that the operators $\hat{P}_k$ are functions of the cut-offs that define the sets I_k. It is then useful to establish the following result:

LEMMA 3: *The functions* $m_i\left(\underline{\alpha}_i, \overline{\alpha}_i, \Lambda_i\right)$ and $m_f\left(\underline{\alpha}_f, \overline{\alpha}_i, \Lambda_f\right)$ *are continuous in all of their arguments. Moreover,* $m_i\left(\underline{\alpha}_i, \overline{\alpha}_i, \Lambda_i\right)$ is *strictly increasing in* Λ_i *and* $m_f\left(\underline{\alpha}_f, \overline{\alpha}_i, \Lambda_f\right)$ is *strictly increasing in* Λ_f.

Proof: From expressions 23 and 24, it is clear that the functions $m_s(\cdot)$ are continuous and strictly increasing in Λ_s. The continuity in the cut-offs is a consequence of the fact that the operator $\hat{P}_k$ is continuous in the cut-offs that define the corresponding set I_k [see Hopenhayn (1992, Lemma 5) for a proof].□

Step 5: Existence and uniqueness

The starting point of this final part of the proof is to use the following results established in Hopenhayn (1992): (i) there exists a stationary equilibrium with invariant measure μ, which is associated to an unique aggregate input-output pair, (L, Q), and unique prices (Theorem 2); (ii) if the entry cost is low enough, a stationary equilibrium with positive entry exists (Theorem 3); (iii) if the profit function is multiplicatively separable between productivity and prices, $\pi(\theta, p, w) = h(\theta)\, g(p, w)$, then if there exists an equilibrium with entry and exit, it is unique (Theorem 4).[7] Given these results, it remains to show that there exists an unique stationary equilibrium with an unique formal-informal partition and entry into both sectors.

First, write $\mu = m\left(\underline{\alpha}_i, \overline{\alpha}_i, \underline{\alpha}_f, \Lambda_f, \Lambda_i\right) = m_i\left(\underline{\alpha}_i, \overline{\alpha}_i, \Lambda_i\right) + m_f\left(\underline{\alpha}_f, \overline{\alpha}_i, \Lambda_f\right)$. Fix μ; from Step 3 (equations 17-20) we know that there are unique cut-off points for entry, exit and transition between sectors. In particular, because $c_f^e > c_i^e$ the following ordering between formal and informal entry thresholds holds: $\underline{\nu}_f > \underline{\nu}_i$. Thus, for any

[7]The functional form assumed for the production function belongs to this class and thus the theorem applies directly.

$\nu \in [\underline{\nu}_i, \underline{\nu}_f)$, entry occurs into the informal sector and for any $\nu \geq \underline{\nu}_f$, entry occurs into the formal sector. Hence, the unique thresholds $\left(\underline{\nu}_i, \underline{\nu}_f\right)$ pin down the mass of entrants into the informal sector, Λ_i. Additionally, fixing μ uniquely determines thresholds $(\underline{\alpha}_i, \overline{\alpha}_i)$, which therefore pins down an unique informal sector size, $\mu_i = m_i\left(\underline{\alpha}_i, \overline{\alpha}_i, \Lambda_i\right)$.

Finally, the industry size μ also uniquely determines the formal sector's threshold, $\underline{\alpha}_f$. But $\mu = \mu_i + \mu_f$ and $\mu_f = m_f\left(\underline{\alpha}_f, \overline{\alpha}_i, \Lambda_f\right)$ is strictly increasing in Λ_f. Thus, once the informal sector size is determined and the thresholds $\left(\underline{\alpha}^f \overline{\alpha}^i\right)$ are fixed, there is an unique value of Λ^f that satisfies the identity $\mu = \mu_i + \mu_f$.

Therefore, there is an unique stationary equilibrium with invariant sector measures, μ_s, entry and exit thresholds, $(\underline{\nu}_s, \underline{\alpha}_s, \overline{\alpha}_i)$, $s = i, f$, aggregate prices and quantities, (Q, L, w), and entry levels in both sectors, (Λ_f, Λ_i).

□

A.3 Proof to Proposition 2

Define the following operators:

$$T_k(B) = \begin{cases} \int_B dF\left(x|\theta\right), & \text{if } \theta \in X_k \\ 0 & \text{otherwise} \end{cases}$$

for $k = 1, 2, 3$ and all Borel sets $B \subseteq \Theta$. As before, $X_1 \equiv A_i^n \cap \left[\underline{\alpha}_i, \overline{\alpha}_i\right)$, $X_2 \equiv A_f^n \cap \left[\underline{\alpha}_f, \theta_h\right)$, and , $X_3 \equiv A_i^n \cap \left[\overline{\alpha}_i, \theta_h\right)$. One can then rewrite the expressions for the λ_s^n as follows:

$$\begin{aligned} \lambda_i^{n+1} &= T_1 \lambda_i^n \\ \lambda_f^{n+1} &= T_2 \lambda_f^n + T_3 \lambda_i^n \end{aligned}$$

The productivity distribution of newborn firms, however, is simply given by $\lambda_i^1([\theta_l, \theta)) = \int_{\nu > \underline{\nu}_i} F\left(\theta|\nu\right) dG(\nu)$ and $\lambda_f^1([\theta_l, \theta)) = \int_{\nu > \underline{\nu}_f} F\left(\theta|\nu\right) dG(\nu)$.

From the second period on, the productivity distribution in every subsequent period is obtained by first applaying the truncation implied by the conditions $\theta \in X_k$ in the operators T_k. That is, the operators T_k not only condition on the "stayers" region of the informal and formal sectors, but on the intersection of these regions and the set of active firms. This always implies a truncation on the lower tail of the productivity distribution and hence this truncation is increasing in the FOSD ordering (the truncated distribution FOSD the untruncated distribution). Finally, assumption (A.2) implies that after the truncation, the operator T_k is also monotone and hence the operator T_k as defined is increasing in the FOSD criterion. Thus, the following holds almost by definition: $\lambda_i^2 \equiv T_1\lambda_i^1 \succeq \lambda_i^1$. By induction, $\lambda_n^{n+1} \succeq \lambda_i^n$ for any n.

The analysis of λ_f^n is not so straightforward because of the presence of the term $T_3\lambda_i^n$, so a more careful argument is needed. First, note that because of the argument just made for λ_i^n, in the absense of the term $T_3\lambda_i^n$ one would observe $\lambda_f^{n+1} \succeq \lambda_f^n$ for any n. Second, the conditioning embeded in operator T_3 is stronger than the one in T_2, as $\overline{\alpha}_i > \underline{\alpha}_f$. Put differently, the productivity distribution of those that make the informal to formal transition is truncated on the left at a higher point than the distribution among the formal "stayer" firms. Thus, $\lambda_f^2 = T_2\lambda_f^1 + T_3\lambda_i^1 \succeq \lambda_f^1$. But given the result that λ_i^n is increasing in n, and that the T_k operators are increasing, then λ_f^n will be also increasing in n. □

A.4 Proof to Corollary 2

As discussed in Section 2.1, there is an unique threshold $\tilde{l}$ above which the formal firm only hires formal workers (on the margin). Thus, if the firm grows in size above $\tilde{l}$, all the workers in excess of $\tilde{l}$ will be hired formally. For a formal firm that has its optimal level of labor below the threshold, $l_f^*(\theta) < \tilde{l}$, $s_i(\theta) = 1$. In this case, the within informality rate will stay constant at one for some period while the firm is growing, but as soon as $l_f^*(\theta_t) = \tilde{l}$ the informal share will start declining monotonically with firm's size. Similarly, for any initial value of $s_i < 1$, as the firm grows the s_i will decline monotonically. Hence, the s_i is a constant function of θ if $l(\theta) < \tilde{l}(\theta)$ and it is strictly decreasing in θ over the range where $l(\theta) \geq \tilde{l}(\theta)$. Combined with the result in Proposition 2, this imples that the average within firm informality for a given cohort will be monotonically decreasing with cohort's age almost everywhere in the relevant range $\left[\underline{\alpha}_f, \theta_h\right]$. In fact, as long as there is always at least one firm within the cohort that has a $s_i(\theta) < 1$, the average within firm informality will be strictly decreasing in the cohort's age. □

A.5 Proof to Corollary 3

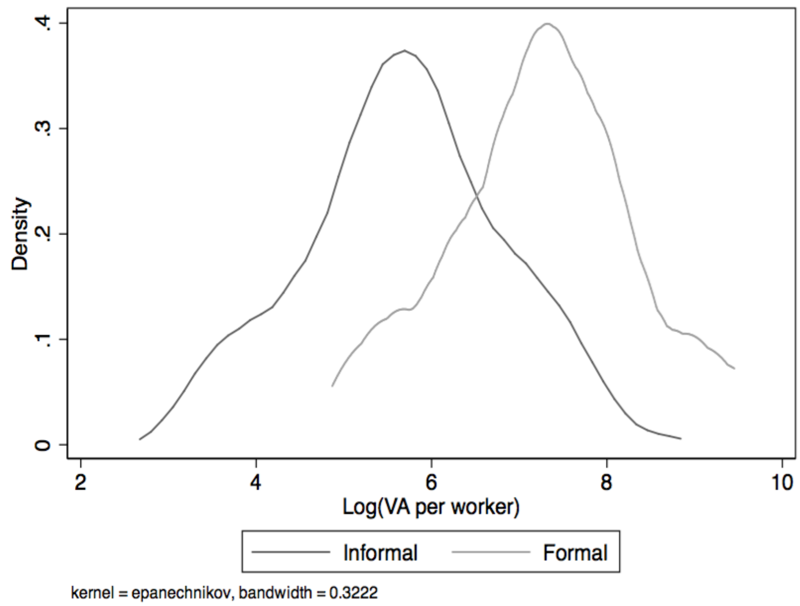
Proposition 2 established that the distribution of productivity is increasing with firms' age in both sectors. At the same time, the informal to formal transition threshold, $\overline{\alpha}_i$ remains constant. Hence, as the cohort gets older a smaller number of firms (among the survivors) will remain informal as the most productive ones keep making the transition into the formal sector. This implies that $\mu\left(A_i^{n+1}\right) \leq \mu\left(A_i^n\right)$. The same is not true for $\mu\left(A_f^n\right)$, as the upper limit of the stayers region is the upper bound of the set Θ itself. This means that formal firms do not face the additional exit margin that informal firms do, as there is no upper limit for their growth. Since the productivity shock in both sectors is the same, even if the set of active firms in the formal sector reduces in size as the cohort ages, it does so at a lower rate than informal firms in the same cohort. Hence, the ratio $\gamma_b^n = \frac{\mu\left(A_i^n\right)}{\mu\left(A_i^n\right)+\mu\left(A_f^n\right)}$ is weakly decreasing in the cohort's age. $\square$

Table 1: Descriptive Statistics

	Formal		Informal	
	Mean	Sd.Dev.	Mean	Sd.Dev.
Owner's schooling (%)				
0 to 8	0.287	–	0.614	–
9 to 11	0.391	–	0.292	–
12+	0.322	–	0.094	–
Has debts (%)	0.333	–	0.155	–
Expanded past year (%)	0.589	–	0.407	–
Invested past year (%)	0.299	–	0.222	–
Sector composition				
Services	0.394	–	0.402	–
Industry	0.078	–	0.110	–
Commerce	0.439	–	0.281	–
Construction	0.049	–	0.160	–
Wages†	0.777	1.232	0.594	0.925
Revenue†	9.119	16.679	1.363	3.323
Profit†	2.53	12.82	0.66	2.61
Firm's age (months)	110.01	98.53	106.04	105.68
# workers	2.72	1.73	1.28	0.72
Capital/Labor	2.39	5.77	0.58	2.35
Obs.	6,632		42,032	

Source: ECINF 2003.
† Normalized by the (country-wide) average wage of prime-age workers who are heads of their household and work in the formal sector.

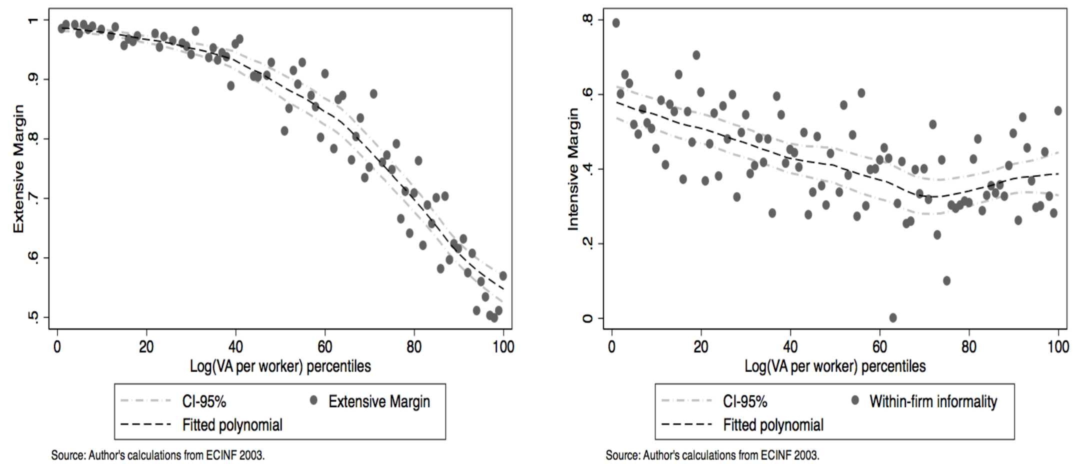
Figure 1: Productivity distribution among entrants in both sectors



Note: The figure shows the kernel density plots of the residuals from a regression of log (value-added per worker) on aggregate industry dummies (e.g. manufacturing, services, and so on) for the formal and informal sectors separately. The sample used contains only firms that are at most one year old.

	(1)	(2)
VARIABLES	x1	x1
age	-0.187**	-0.179**
	(0.088)	(0.088)
age2	0.040**	0.037**
	(0.019)	(0.019)
age3	-0.002**	-0.002**
	(0.001)	(0.001)
_Iindustry_2	-0.091	-0.089
	(0.063)	(0.064)
_Iindustry_3	-0.017	-0.018
	(0.042)	(0.043)
outside		-0.166***
		(0.061)
nocapital		-0.017
		(0.094)
ownk		-0.08127
		(0.059)
Constant	0.677***	0.872***
	(0.112)	(0.118)
Observations	1,851	1,851
R-squared	0.016	0.035

Figure 2: Firm's productivity and the margins of informality

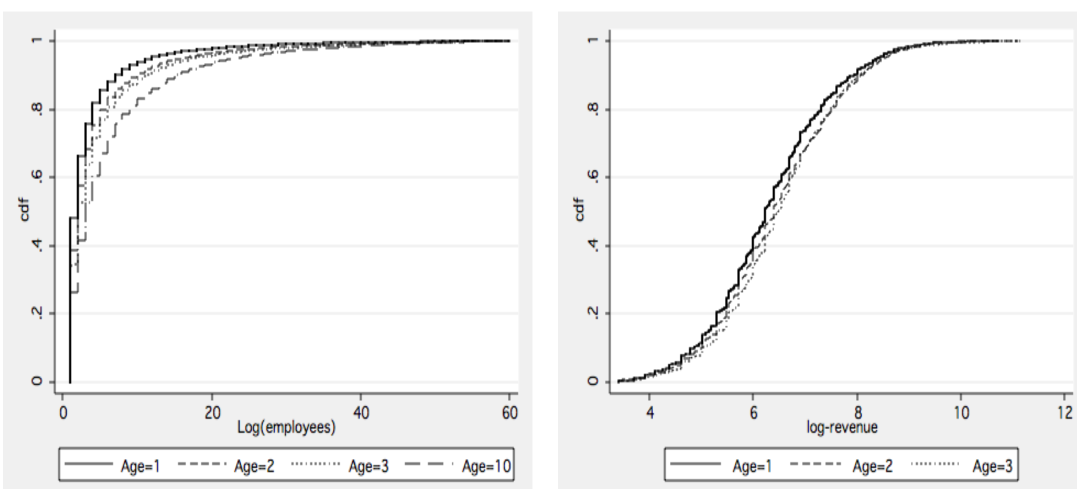


(a) Extensive margin and productivity

(b) Extensive margin and productivity

Note: The left panel displays the share of informal firms within each percentile of the log-productivity distribution. The right panel contains the average share of informal workers within formal firms for each percentile of the log-productivity distribution.

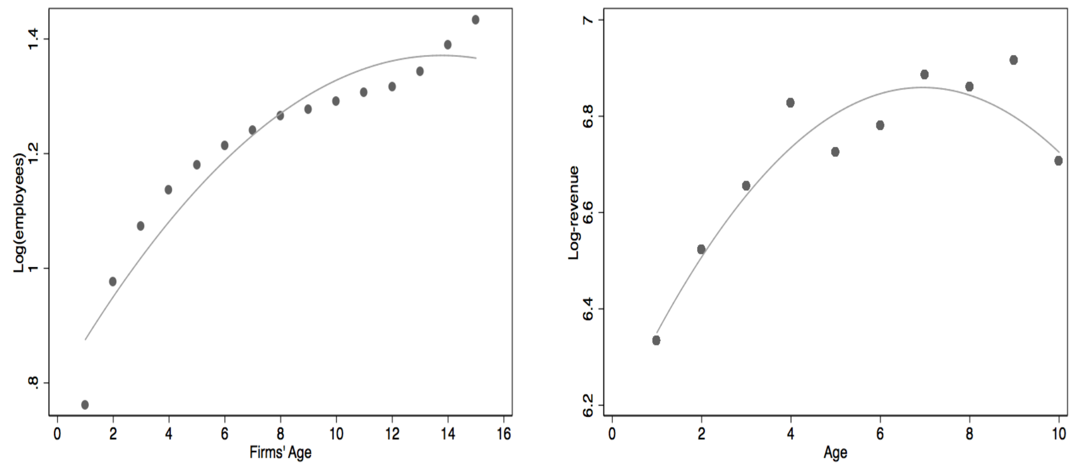
Figure 3: Size distribution and cohorts' age



(a) Formal firms

(b) Informal firms

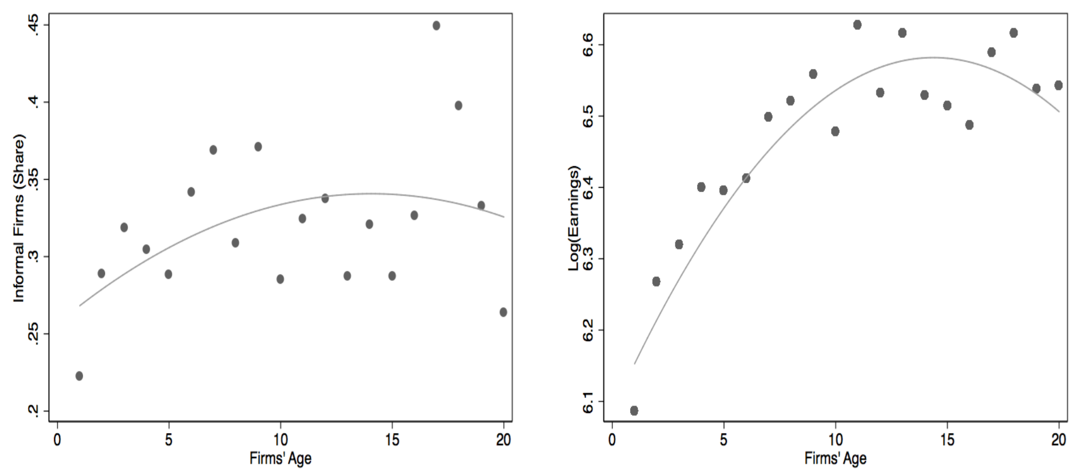
Figure 4: Age-size profiles



(a) Formal firms: size = log(employees)

(b) Informal firms: size = log(revenue)

Figure 5: Age-size profiles from PNAD



(a) Formal firms: size = log(revenue)

(b) Informal firms: size = log(revenue)

Figure 6: Age effects from a fixed effects model – formal firms

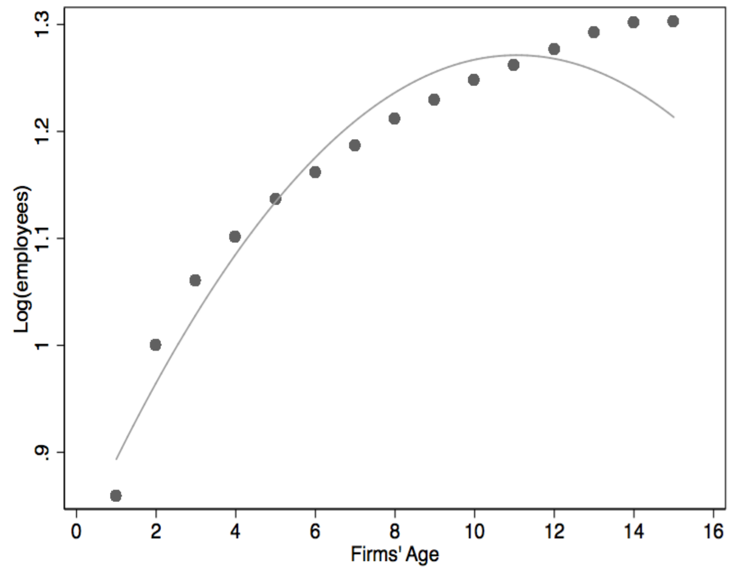


Figure 7: Avg. within-firm informality rate and cohort's age

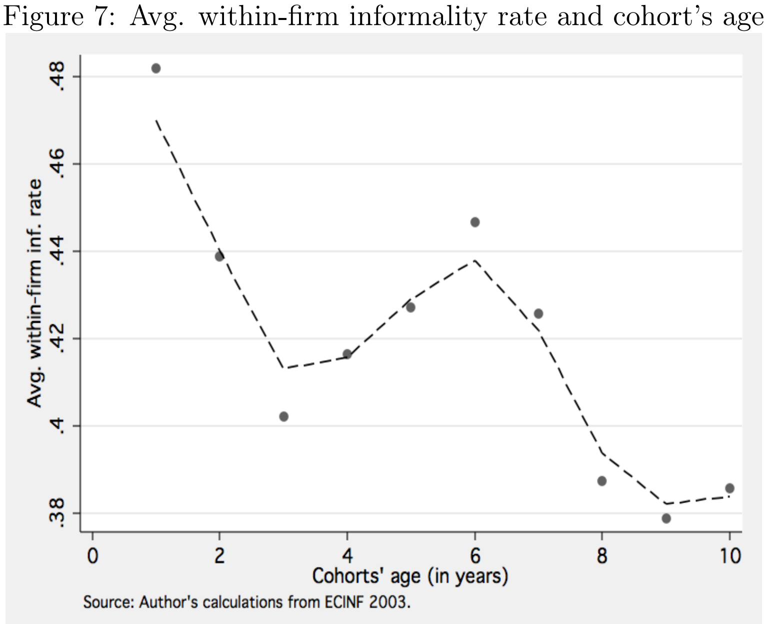


Figure 8: Firm informality rate and cohort's age

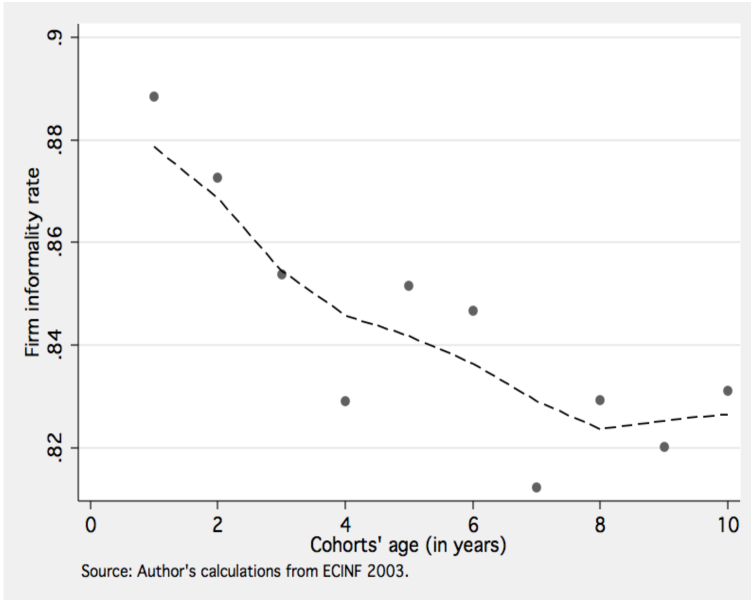
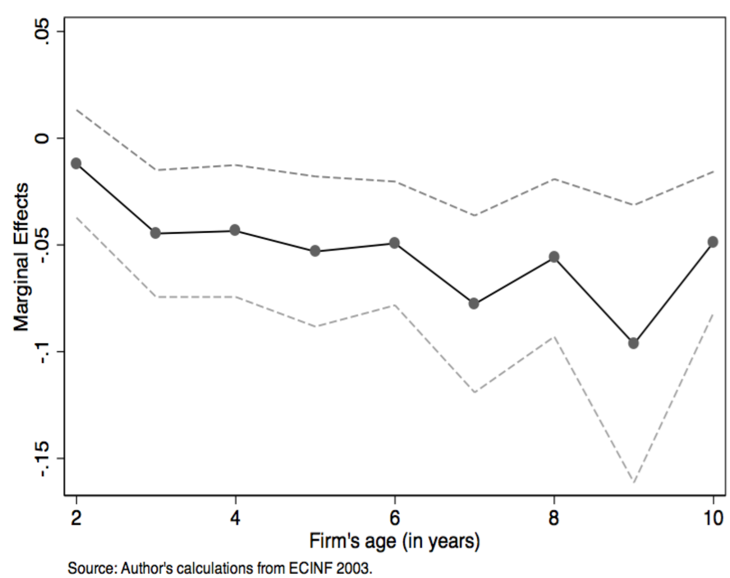


Figure 9: Probability of being informal: Age Marginal Effects



Note: Solid line denotes marginal effects from firms' age. Dashed line denote the 95% confidence interval. Marginal effects computed from a probit model where the dependent variable is a dummy of whether the firm is informal and the controls are firms' age, industry dummies and dummies for wether the entrepreneur conducts her business outside her home, if she needed capital to start her business and wether she used own capital to start her business.