A Structural Model of Informality with Constrained Entrepreneurship*

Pierre Nguimkeu†

Georgia State University

July 2020

Abstract

This paper presents and estimates a structural partial equilibrium model of informality and entrepreneurship using data from Cameroon. Institutional factors include registration costs, probability of detection of informal activity, credit constraints, and taxes in the formal sector. I show that the decision to formalize is U-shaped in skills with the turning point corresponding to secondary school completion. Initial wealth and average education are important drivers of informal entrepreneurship whereas higher education and parent's entrepreneurial status are significant determinants of formal entrepreneurship. The estimated model is then used to simulate the counterfactual impact of changes in registration costs, taxation, and enforcement, which are found to substantially affect formalization, aggregate income and government revenues. However, none of these policies is able to reduce the size of informality to less than 20-30%. This suggests that a different perspective than the traditional formalization objective should be considered when dealing with informality in Africa.

Keywords: Entrepreneurship, Informality, Regulation, GMM, Counterfactual analysis.

JEL Classification: O12, O17, H21, C51, C54.

*I thank Roland Pongou, Emmanuelle Auriol, Alex Karaivanov, Andrew Feltenstein, Cesar Calderon, Albert Zeufack, Omari Swinton, and conference participants of the 2014 World Economic Congress, the 2014 Southern Economic Association meeting, the 2015 Canadian Development Economics Study Group, and the 10th IZA/World Bank Conference on Employment and Development for useful comments and feedbacks. I thank two anonymous referees for their useful feedback that have greatly improved this paper. Finally, I thank the Cameroon National Institute of Statistics for making the data available to me for this study as well as Augustine Denteh for Excellent research assistance.

†Department of Economics, Andrew Young School of Policy Studies, Georgia State University, PO Box 3992, Atlanta, GA 30302-3992, USA; Email: nnguimkeu@gsu.edu.

1 Introduction

The importance of entrepreneurship for both developed and developing economies cannot be overemphasized. It is a crucial ingredient for promoting and sustaining economic growth, given its potential for creating jobs, delivering innovation and raising productivity. In the context of developing economies, however, the existence of a large informal sector (60-90% of the overall workforce) that coexists with the formal sector brings additional challenges to policy-makers' objective to understand and promote entrepreneurship. On the one hand, formal firms have higher growth prospects, higher productivity and income potential (Schneider and Enste 2000), and generate tax revenues that can be used by the state to improve the provision of public goods and services. On the other hand, informal businesses are generally low-scale, largely untaxed and create shortfalls in government revenues. At the same time they may represent an important avenue for job creation, an incubator for business potential or a stepping stone for accessibility to the formal economy, especially for less-educated workers (ILO 2002, Cano-Urbina 2015).

This paper extends Nguimkeu (2014)'s model of occupational choice in which workers with different productivities and capital choose between non-entrepreneurial work and entrepreneurship. The extension consists in distinguishing between informal and formal entrepreneurship by accounting for the role of institutional factors such as formal registration costs, probability of detection of informal activity, credit constraints, and taxes in the formal sector. Both forms of entrepreneurship use the same production technology but unlike formal entrepreneurs, informal entrepreneurs pay no taxes and no registration costs, can only borrow a fraction of their initial wealth and may have their profits confiscated with a certain probability. Formal entrepreneurs and workers are thus added to the sample to jointly estimate by GMM the parameters governing the ability distribution, the production function and the associated constraints. The main implications of the model are that less productive entrepreneurs choose informality whereas the most productive ones choose to formalize. Formality is costly because of both taxation and high entry regulations whose affordability is limited to firms promising sufficiently high returns and prospects to grow. This induces fewer entry decision from entrepreneurs with low ability so that the share of informal firms remains high, even when the tax rate is reasonably small. These predictions are tested employing reduced form techniques and specification analysis, and the estimated model is used to quantify the counterfactual impact of simulated changes in registration costs, tax rates, and probability of enforcement.

The existing research on formal/informal occupational choice and the impacts of policy efforts to increase formality comes from a few developing countries, concentrated in Latin America and Southeast Asia (see Bruhn and McKenzie 2014 for a recent review). Very little attention about these issues has been given to Sub-Saharan African countries, which differ from these regions by several relevant features. For instance, the size of informal employment and the entry costs to the formal sector are substantially higher in Africa compared to Latin America and Asia, whereas the levels of education and

financial development are relatively lower (see Figure 12 and Table 9 in the Appendix). Also, the structure of the informal sector (e.g., type of firms, performance, proportion of subsisters) or other type of regulatory barriers (e.g. titling, licensing, etc.) differ substantially across regions. To that extent, findings from other regions do not necessarily generalize to Africa, and this paper endeavours to provide some evidence from a typical African country in an effort to close this gap. The model discussed and estimated is an amended version of several models in the literature (e.g. Fortin et al. 1997, Amaral and Quentin 2006, Antunes and Cavalcanti 2007, Dabla-Norris et al. 2008, de Paula and Scheinkman 2011, Prado 2011, Ordoñez 2014). I present an extension that fits the context of many African countries as suggested by recent studies (e.g. World Bank 2005, Djankov et al. 2002, Auriol and Walters 2005, McKenzie 2011) and discussed in section 2.

The structural GMM estimation is performed by constructing moment restrictions that match occupational choice probabilities and earnings to their sample analogs. Evidence is provided using data from the 2005 Cameroon National Survey on Employment and Informal Sector (EESI).[1] The results show that there is a U-shaped relationship between entrepreneurs' education and their decision to formalize. Less educated entrepreneurs find it more profitable to remain informal. It is only above a certain educational threshold, estimated at secondary school completion, that formality becomes increasingly attractive to informal entrepreneurs. I also found that parents occupations play a key role in entrepreneurial choice and explain why more than 40% of formal entrepreneurs in the data are children of entrepreneurs. Specification analysis shows that failing to incorporate entry registration costs substantially deteriorates the model performance in terms of selection into the formal sector. To quantitatively assess the role of institutional factors as the main sources of the occupational inefficiencies observed in Cameroon (i.e. too many informal workers and too few entrepreneurs, whether formal or informal), counterfactual policy simulations are performed using the estimated model. I found that a 50% decrease in registration costs doubles the proportion of formal enterprises (through both formalization of informal firms and new formal enterprise creation) and increases aggregate income by 15%. Interestingly, total tax revenues net of the foregone registration fees increase by more than twice. Counterfactual results on tax policy uncover a Laffer curve where the efficient tax rate, estimated at 24% (i.e. about half of the current tax rate) generates 35% tax revenue gains above the current level, twice as much formal enterprises and 25% increase in aggregate income. By contrast, while an increase in the enforcement of formality status slightly increases the fraction of formal firms and the associated tax revenues, it has an overall perverse effect on the economy in terms of enterprise destruction. In all these scenarios, simulations reveal that none of these traditional formalization mechanisms is able to reduce the size of the informal sector to a range lower than 20-30%. This calls for a different perspective than the usual formalization objective to be considered when dealing with informality in sub-Sahara African countries.

[1] Available via the Cameroon National Institute of Statistics website at www.statistics-cameroon.org.

Section 2 presents a theoretical model of occupational choice and derives the main implications. Section 3 describes the data and provides reduced form results. Structural estimation and specification analysis are performed in Section 4. Counterfactual policy simulations are presented in Section 5, and Section 6 concludes. The Appendix gathers other technical material including the mathematical proofs and additional tables and figures.

2 Model Description

The economy is populated with individuals who differ in their entrepreneurial skills θ and their initial wealth endowment z. They choose their occupation at the beginning of the period, based on their expected end-of-period payoff. A non-entrepreneurial worker (e.g. wageworker or subsister), receives a fixed earning w at the end of the period. This income is compared to the profits the agent could receive if they start a firm. Regardless of whether they are formal or informal, an entrepreneur with skill θ uses capital k and hires l units of labor to produce goods according to the technology

$$y = \theta k^{\alpha} l^{\beta}, \tag{1}$$

where $\alpha, \beta \in (0,1)$ are the elasticities of output with respect to capital and labor, respectively. We also have $\gamma = \alpha + \beta < 1$, implying diminishing returns to scale in variable factors at the establishment level (see Lucas 1978).

2.1 Informal Entrepreneurship

Informal entrepreneurship means that the entrepreneur does not pay taxes to the government. Tax avoidance comes with a risk of being caught, in which case the informal entrepreneur's profit is forfeited.[2] Moreover, when operating in the informal sector entrepreneurs have limited access to credit. In order to get a loan from financial institutions, they need to provide collateral as a guarantee. They can therefore only borrow up to a fixed multiple, $\lambda \geq 0$, of their initial wealth, $z \geq 0$, that they use as collateral. Denote by p the probability of getting caught. The informal entrepreneur's optimal investment capital and labor then solves the profit maximization problem

$$\max_{k,l} \left\{ (1-p) \left[\theta k^{\alpha} l^{\beta} - wl - rk \right] : \quad 0 \leq k \leq \lambda z, \quad l \geq 0 \right\}.$$

For $\lambda = \infty$ the credit market is perfect and $\lambda = 1$ corresponds to financial autarky, where all capital is self-financed by the informal entrepreneurs, especially if z is a liquid wealth that can be readily used to finance a business. On the one hand, this specification captures the common prediction from models of limited contract enforcement typical of

[2] From anecdotal evidence, when a firm gets caught, the property is sealed until the owner pays a given amount of fine based on the amount of capital and goods found in the property, after which they usually can recuperate their belongings.

the informal economy where the amount of credit is limited by individual wealth.[3] On the other hand, the specification follows previous work such as Dabla-Noris et al. (2008), Mendicino and Prado (2014), Loyaza (1996) in assuming constant detection probability of informal firms. This simplifying assumption, however, substantially differs from other authors (e.g. Ordonez 2014, De Paula and Sheinkman 2011) who assumed that detection probability is a binary function of capital and takes the value of 1 (perfect detection) when k is larger than some exogenously fixed amount, and zero (no detection) otherwise, such that only small firms subsist in the informal sector. A particular feature of informality in sub-Sahara African countries, however, is that it is highly heterogeneous and comprises both small and large firms that operate informally. Hence, a size restriction by sector, as assumed by the latter authors, would be inconsistent with Cameroon data (See Choi et al. 2019, Chap 3). In addition, detection is operated by sending tax inspectors in markets and neighborhoods who prosecute all categories of informal business owners regardless of their size.[4]

The optimization constraint on capital then gives rise to two types of informal entrepreneurs. Those who are financially unconstrained, i.e., their optimal investment capital is an interior solution of the above optimization problem, and those who are financially constrained, i.e., their capital constraint is binding. The interior solutions of the entrepreneurs maximization problem are

$$k^* = \theta^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{r}\right)^{\frac{1-\beta}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} \quad \text{and} \quad l^* = \theta^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{1-\alpha}{1-\gamma}} \tag{2}$$

This solution is feasible only if k^* is lower than λz, or equivalently

$$\theta \leq (\lambda z)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{1-\beta} \left(\frac{w}{\beta}\right)^{\beta} \equiv \theta_c(z) \tag{3}$$

When the constraint is binding the investment capital and optimized labor are given by

$$k^{**} = \lambda z \qquad \text{and} \qquad l^{**} = \theta^{\frac{1}{1-\beta}} \left(\frac{\beta}{w}\right)^{\frac{1}{1-\beta}} (\lambda z)^{\frac{\alpha}{1-\beta}}.$$

The optimal informal entrepreneur's profits can therefore be expressed as follows:

$$\pi^I(z,\theta) = \begin{cases} \pi_u^I = (1-p)(1-\gamma)\theta^{\frac{1}{1-\gamma}} \left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} & \text{if } \theta \leq \theta_c(z) \\ \pi_c^I = (1-p)\left[(1-\beta)\theta^{\frac{1}{1-\beta}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta}} (\lambda z)^{\frac{\alpha}{1-\beta}} - \lambda r z\right] & \text{otherwise.} \end{cases} \tag{4}$$

[3] This collateral constraint can usually be derived from a limited liability problem. As suggested by an anonymous referee, a more practical specification should adopt a constraint of the form $k^f + \lambda z$, where k^f is the amount of social capital (e.g. from friends and family). This information is however not available in my data.

[4] As we assume in the empirical estimation, detection probability is more likely to depend on the size of the tax department which we can reasonably think of as constant within a period, for simplicity.

The informal entrepreneurs profit takes two possible values according to whether he is unconstrained, that is, $\theta < \theta_c$ or he is constrained, that is, $\theta > \theta_c$. The main difference between this profit function and the one derived by Nguimkeu (2014) is the enforcement probability account and the inclusion of paid labor in the entrepreneur's objective. While this might be of relatively low importance for informal entrepreneurs of whom very few hire paid labor outside of their household, hired labor turns out to be a key characteristic of formal firms in the data.

2.2 Formal Entrepreneurship

In order to formalize, entrepreneurs need to pay a registration cost, c, that represents both the license fees and the amount of time and bribes spent to get it. Moreover, at the end of the period, the government levies a tax τ on the profit of formal enterprises. Their formal status however provides them with a better access to credit from financial institutions (see Bruhn 2013, Laporta and Shleifer 2014). I therefore follow the literature and assume as in de Paula and Scheinkman (2011) (see also Ordoñez 2014) that the formal entrepreneurs' optimal investment capital and labor solves for the optimization problem[5]

$$\max_{k\geq 0, l\geq 0} \left\{(1-\tau)\left[\theta k^{\alpha} l^{\beta} - wl - rk\right] - rc\right\}$$

The optimal capital and labor are given by Equations (2) above and yield the following expression for the optimal formal entrepreneur's profit:

$$\pi^F(z,\theta) = (1-\tau)(1-\gamma)\theta^{\frac{1}{1-\gamma}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} - rc. \tag{5}$$

The cost c is fixed, implying that wealthy people may formalize more easily.[6]

The closest to this model specification is a general equilibrium calibration model of occupational choice recently presented by Ordoñez (2014), but both differ in several distinctive ways. First, unlike Ordoñez (2014) who assumed away entry registration costs into the formal sector, I formally include it in my model consistently with recent evidence showing that these costs are relatively much higher in African countries than in other regions of the world and are a serious constraint to formal entrepreneurship (see Djankov et al. 2002, World Bank 2005, Klapper et al. 2006, and Table 5 below for

[5]Notice that while the agent pays the one-time entry cost c to formalize his business, his periodic formal profit accounts for an amount of rc instead. This is because c can be seen as the present value of a sum of periodic payments of rc across the lifetime of the firm, i.e. $c = \sum_{t=1}^{\infty} \frac{rc}{(1+r)^t}$, assuming payments begin at the end of the current period. Thus, expensing rc each period allows to compare the lifetime streams of income for formal business against informal businesses and non-entrepreneurial work.

[6]In some unreported structural estimations, I allowed the entry cost to depend on wealth in order to assess the relevance of a price-discrimination in favor of wealthy people (e.g., through their political influence, as explained in Ayittey 2006) or against them (e.g. through possibly larger bribes, as suggested by an anonymous referee). However, the wealth effect was insignificant and this specification did not improve the baseline results and was hence dropped.

the Cameroon data). I then empirically show through statistical testing that failing to incorporate these costs could seriously undermine the model performance in terms of selection into the formal sector of Cameroon. Second, Ordoñez (2014) assumed that there is no financial constraints in the informal sector and that capital is freely optimized by informal entrepreneurs. These entrepreneurs may however choose to operate with less than an exogenously fixed amount, but only as a strategy to avoid detection. In contrast, I assume financial constraints in the informal sector and allow for this constraint to vary across individuals to account for heterogeneity in their initial wealth which is reflected in the heterogeneity in their capital observed in the data. This assumption is consistent with empirical works that have examined micro-enterprise financing in developing countries (e.g., Paulson et al 2006), and has been empirically tested by Nguimkeu (2014) with Cameroon data. The model in this paper is also more flexible than the one discussed by Prado (2011). Unless the probability of detection is higher than the tax rate, the Prado's model would deliver a corner solution equilibrium where all entrepreneurs prefer the informal sector even with a zero entry cost in the formal sector. This seems inconsistent with data from African countries where enforcement is weak and the probability of detection is usually lower than the tax rate, yet both formal and informal sectors coexist.

The assumption that formal entrepreneurs are not financially constrained in the credit market is clearly a theoretical simplification. In reality, some degree of financial frictions exists in the formal sector as well, although at a much lower extent than the informal sector. However, access to finance is the only true benefit of formalization within the model (and often also in actual economies) and allows for some important theoretical results in the paper to concur with the data, in particular the dominance of formal entrepreneurship by sufficiently high entrepreneurial ability.[7]

2.3 Occupational Choice

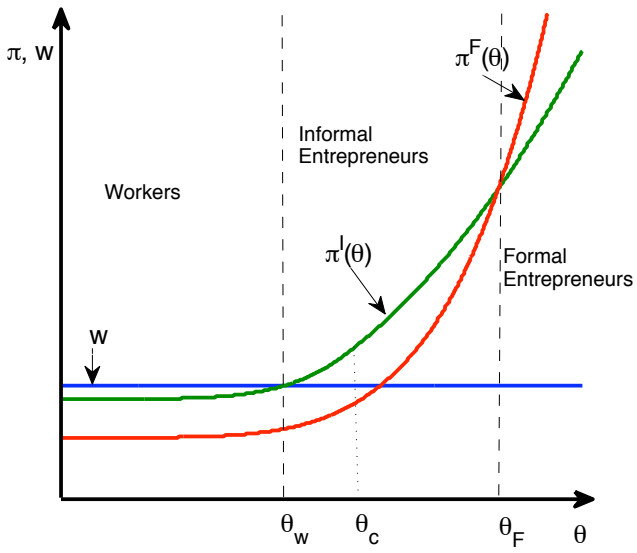
Each individual knows his personal attributes θ and z, market characteristics r and w and institutional factors τ, p and c, where it is assumed that $\tau > p$, as commonly observed in African countries. Given these factors, the agent chooses the occupation that would give him the maximum earning. In other words, the expected profit function of an agent with characteristics (z, θ) from the three categories of occupations can be written as

$$\pi(z, \theta) = \max\left\{w, \pi^I(z, \theta), \pi^F(z, \theta)\right\}$$

Earnings functions and related occupational choice are illustrated, as an example, in Figure 1. The agent's decision is characterized by three thresholds, θ_W, θ_F and θ_c, which summarize the occupational decisions of the agents and whether the capital choices of informal entrepreneurs are constrained or unconstrained. The threshold θ_c is the one given in Equation (3). As for θ_W, notice that entrepreneurs profits are increasing

[7] Trying to estimate a leverage parameter in the formal sector turned out to be infeasible due to identification issues.

Figure 1: Earning Functions and Occupations



with θ while workers' earnings do not vary with θ. It follows that there exists an ability threshold θ_W such that $w = \max\left\{\pi^I(z,\theta_W), \pi^F(z,\theta_W)\right\}$. Hence, all agents with $\theta < \theta_W$ remain workers and the rest become entrepreneurs. Finally, note that informal entrepreneurs with initial wealth z cannot operate with capital above λz. This constraint is more costly for higher skilled entrepreneurs as they would prefer larger scale firms given their high productivity. Hence, there exists an ability threshold, θ_F, determined by $\pi^I(z,\theta_F) = \pi^F(z,\theta_F)$, above which all entrepreneurs prefer the formal sector and the rest the informal sector. These results are summarized as follows.

Proposition 1. *Consider an agent with characteristics θ and z. There exist three critical entrepreneurial ability thresholds $\theta_W(z)$, $\theta_c(z)$ and $\theta_F(z)$, with $\theta_W(z), \theta_c(z) < \theta_F(z)$, such that*

(i) If $\theta < \theta_W(z)$ the agent is a wageworker or subsister

(ii) If $\theta_W(z) \leq \theta < \theta_F(z)$ the agent is an informal entrepreneur. In particular, when $\theta_W(z) \leq \theta < \theta_c(z)$ he is an unconstrained informal entrepreneur, and when $\theta_c(z) \leq \theta < \theta_F(z)$ he is a constrained entrepreneur.

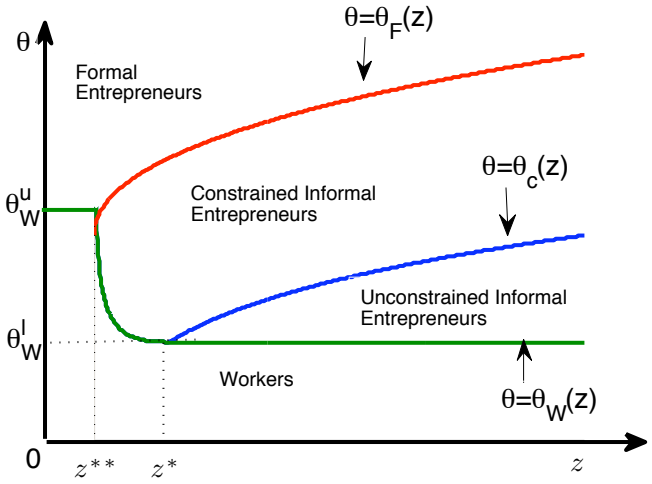
(iii) If $\theta \geq \theta_F(z)$ the agent is a formal entrepreneur.

Proof. See Appendix. □

Whether or not an informal entrepreneur is constrained is a key determinant for his prospects to eventually formalize. Clearly, as long as the tax rate τ is higher than the enforcement p, unconstrained informal entrepreneurs will not formalize, regardless of the entry cost c. Hence, inducing formality from registration reforms would mainly target

only constrained entrepreneurs and high skilled workers. In contrast, business training might induce both types of informal entrepreneurs to formalize as well as low skilled workers. The nature of the selection to the different types of occupations is illustrated, as an example, in Figure 2. This figure depicts a theoretical partition of the wealth-ability

Figure 2: Charaterization of Occupations



space that forms the basis of the structural estimation discussed in Section 4. The initial wealth cutoffs z^{**} and z^{*} illustrate the presence of credit constraints in the informal economy; z^{**} is the minimum wealth that average talents need to start an informal firm, while z^{*} is the minimum wealth they need to start an unconstrained informal enterprise.[8]

Finally, the model also provides a ground to examine how earning differences between occupations may be related to skills and wealth. Three comparative payoff functions can be considered to examine the agent's decision-making. The earning difference between non-entrepreneurial work and informal entrepreneurship is characterized by the comparative payoff function $V^{W,I}(z,\theta) = \pi^{I}(z,\theta) - w$; the difference between non-entrepreneurs and formal entrepreneurs earnings characterized by $V^{W,F}(z,\theta) = \pi^{F}(z,\theta) - w$; and the difference between informal entrepreneurs and formal entrepreneurs earnings characterized by $V^{I,F}(z,\theta) = \pi^{F}(z,\theta) - \pi^{I}(z,\theta)$. The properties of the comparative functions are illustrated in Figure 3 and summarized in the following proposition.

Proposition 2. *Consider agents with characteristics θ and z. Then*

(i) The functions $V^{W,I}(\theta,z)$ and $V^{W,F}(\theta,z)$ are increasing in θ.

(ii) The function $V^{I,F}(\theta,z)$ is U-shaped in θ. That is, there exists $\theta_m > 0$ such that

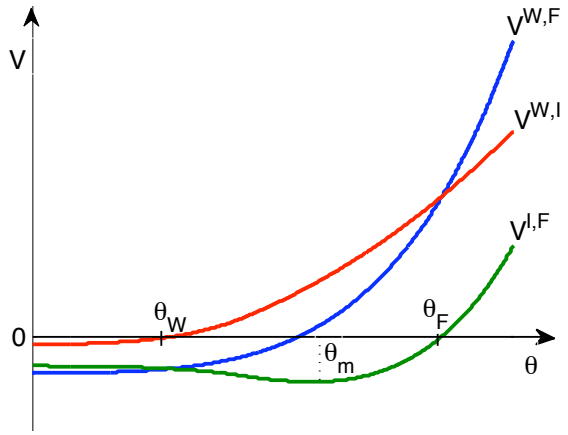
$$\frac{\partial V^{I,F}(\theta,z)}{\partial \theta} < 0 \quad \forall \theta \in [0,\theta_m), \qquad \text{and} \qquad \frac{\partial V^{I,F}(\theta,z)}{\partial \theta} \geq 0 \quad \forall \theta \in [\theta_m,\infty)$$

[8] Below the cutoff z^{**} one can only either remain in subsistence work or have such a great entrepreneurial idea that they get an uncollateralized loan to start a formal firm. In the data, however, z^{**} is nearly zero, making it more factual.

Proof. See Appendix. □

This result states that while the decision of a worker to start a firm is increasing in talent, the decision of an informal entrepreneur to formalize is non-monotonic in talent. In fact, this decision function is U-shaped in entrepreneurial skills (see Figure 3). It implies that low ability entrepreneurs find it more profitable to remain informal as their level of skills increases. This is especially true for unconstrained informal entrepreneurs who have no incentive to formalize. However, above a certain ability threshold formality becomes increasingly attractive for increasing levels of skills, because entrepreneurs with such skills would want to operate at a larger scale but would be constrained in the informal sector. While Figure 3 depicts a latent relationship that is never observed, it is possible to test it empirically using reduced form regressions as I do in Section 3.

Figure 3: Illustration of the Comparative Payoff Functions



Consistently with a large strand of the labor market literature (e.g. Magnac 1991, Maloney 1999, Pratap and Quentin 2006), this framework assumes that labor is free to flow between sectors and that individual's expected wage, w, is the same, regardless of the sector, although this wage is riskier in the informal sector. While evidence shows that on average formal wages are higher than informal wages, direct empirical tests of the premise that similar workers would expect higher wages in the formal sector compared to informal sector yield inconclusive results. For example, Magnac (1991), Maloney (1999), Pratap and Quentin (2006), and El Badaoui et al. (2008) do not find compelling evidence of labor market segmentation between formal and informal sector using data from Colombia, Mexico, Argentina, and South Africa, respectively. Lastly, implicit to this framework is the assumption that more entrepreneurship is broadly better. But it may be that fewer small firms and more large firms is better for the aggregate economy. To answer this question however would require a general equilibrium framework which allows wages and the interest rate to clear input markets.[9]

[9] I thank an anonymous referee for pointing this out.

3 Data and Reduced Form Evidence

This section describes some important features of the data used for this study and assesses the empirical relevance of some of the model predictions for Cameroon. While the choice of Cameroon for the empirical analysis is largely driven by the availability of the data to the author, there is an ongoing rethink of the Cameroon strategy of promoting entrepreneurship and implementing formal business law to the informal sector, for which this empirical work could be an important addition to policy-making. I give the background and provide descriptive statistics of these data and then test some implications of the model through reduced-form regressions.

3.1 Data and Background

The data used for the empirical analysis is a representative cross-sectional sample of households of Cameroon stemming from the Survey on Employment and the Informal Sector (EESI), conducted in 2005 by the National Institute of Statistics of Cameroon in partnership with the World Bank. This is a nationwide operation with two phases. The first phase collects sociodemographic and employment data while the second phase interviews a representative subsample of informal production units identified during the first phase. The methodology of the EESI is therefore similar to that of Phases 1 and 2 of the well-known "1-2-3 surveys" in West Africa (details can be found at www.afristat.org). For the analysis, I restrict the sample to households whose heads are active and are aged 15 or above, representing a total of 6112 observations. The definition of informality used for the EESI survey is on the basis of administrative records and on whether or not the business keeps formal accounts. Accordingly, informal enterprises are defined as "production units that do not have written formal accounts or are not registered with the tax authorities." Informal sector workers are therefore persons exercising their main economic activities in informal establishments. The informal sector accounts for the vast majority of activities and employs 89.5% of the Cameroon workforce aged 15 and above (INS 2005). The sample used in this study consists of 4337 households from the Informal sector and 1775 households from the formal sector, based on the above definition. Among them, very few are entrepreneurs of which very few are formal. We examine to what extent credit constraints, high registration costs and taxation drive these inefficiencies.

3.2 Descriptive Statistics

The average age of household heads is 36.2 years, 48.1% of which have a primary education, 36.5% a secondary education and 15.4% a tertiary education. The empirical analysis requires distinguishing entrepreneurs from non-entrepreneurs according to their activity. While this distinction is clearer in the formal sector, it is not so obvious in the informal sector. For the formal sector, the classic literature of entrepreneurial choice, which I follow, considers self-employment or business ownership as formal entrepreneurship (e.g., Evans and Jovanovic, 1989; Holtz-Eakin et al. 1994, Blanchflower and Oswald

1998). However, in the informal sector, where the majority of people are self-employed, using the same definition would be seriously misleading in this context. In fact, self-employment in the informal sector includes both the actual informal entrepreneurs as well as a wide category of subsisters. To distinguish between these activities, I follow Nguimkeu (2014) and define as informal entrepreneur a household that owns a business and employs others (see also Mondragón-Vélez and Peña 2010, for a similar definition). This definition particularly excludes purely self-employed (i.e. those who work just by themselves) most of which are subsisters.[10] Examples of informal entrepreneurs include taxi-drivers, grocers, tailors, carpenters, car mechanics who own shops, etc. Non-entrepreneurial workers are all other types of households including wage-earners employed by formal firms, informal firms or households as well as all the subsisters.

Table 1: Household Characteristics by Occupations

Characteristics	Occupations		
	Formal Entrepreneurs	Informal Entrepreneurs	Wageworkers /Subsisters
Num. of obs.	65	424	5 623
% of sample	1.1%	6.9%	92.0%
Female	12.3%	37.3%	41.7%
Av. household size	6.0	6.1	5.9
Av. age of head	42.4	37.3	36.1
Years of schooling			
0-6 years	11.1%	41.3%	48.4%
7-12 years	31.5%	48.6%	36.2%
13+ years	57.4%	10.1%	15.4%
Parent Entrep.	41.5%	13.6%	3.5%
Av. monthly income*	353.3	71.9	70.3
Av. wealth*	21 792.9	4 569.7	3 007.4

*In thousands of local currency (CFA); $1,000$ CFA $\sim$ \$2 US (in 2005)

Notes. These data are at the Household level and focus on the household head.

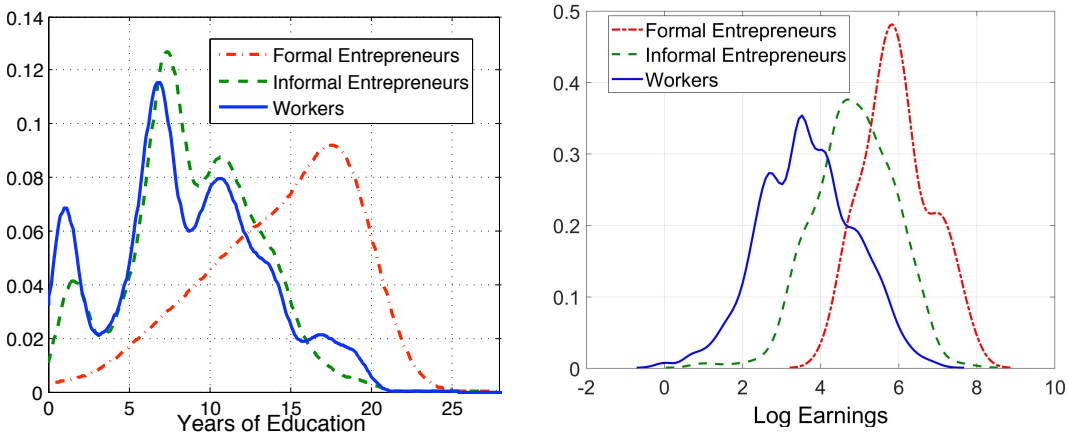
Table 1 summarizes the main characteristics of households in the sample according to their occupation (formal entrepreneurs, informal entrepreneurs, wageworkers/subsisters).[11]

[10]Exceptions are some activities that use high capital but are restricted to a single operator, such as taxi-drivers who own their taxi. This definition is obviously still not perfect. On the one hand, it would be inadequate to consider as entrepreneurs some employers who operate at a very low scale. On the other hand, even a purely self-employed with low physical capital could produce innovative products or services such that considering it as subsistence is misleading. Such cases are however infrequent in these data as evidenced by the sensitivity analyses performed in Nguimkeu (2014).

[11]The summary statistics are at the household level given that in the empirical analysis the agent is the household head

The sample consists of 92.0% workers, 6.9% of informal entrepreneurs and 1.1% of formal entrepreneurs. While it is possible that some households engage in two or more forms of activity at the same time, I take the primary activity as their main employment. In particular, a household is considered as entrepreneurial if at least one member is entrepreneur in the above sense. The average number of paid employees (i.e. hired outside of family) per informal enterprise is about 1.3 against more than 50 for formal enterprises. While entrepreneurs are on average older than workers, formal entrepreneurs are on average much older than informal entrepreneurs. The differences in the education composition across occupations is sizeable. The highest proportion of non-entrepreneurs (48.4%) has a primary education whereas the highest proportion of informal entrepreneurs (48.6%) has a secondary education, and the highest proportion of formal entrepreneurs (57.4%) has a post-secondary education. This basic description lines up with the theoretical prediction that in general these occupations are associated with low, medium and high skills, respectively. The left pattern of Figure 4 shows the density of years of education by occupation and confirms the superior educational level of entrepreneurs over workers'.[12] The data also show that a high proportion of entrepreneurs are sons and daughters of entrepreneurs. In particular, for 41.5% of formal entrepreneurs at least one parent was an entrepreneur. This is also true for 13.7% of informal entrepreneurs. By contrast, only 3.5% of workers are children of entrepreneurs.

Figure 4: Distribution of Education and Earnings by Occupation

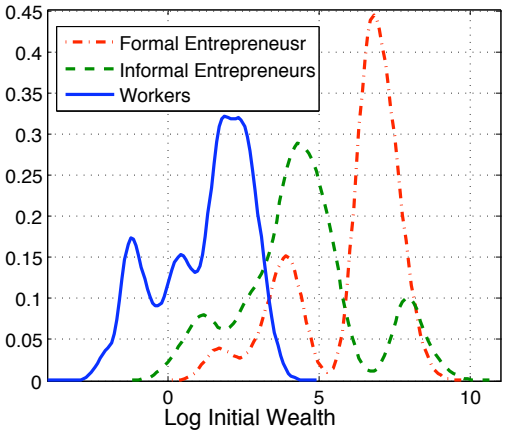


These patterns suggest, as I assess in the structural estimation presented in Section 4, that both education and parents entrepreneurial status are strongly associated with unobserved entrepreneurial ability, and could thus also be used as proxies for skills in a reduced form setting as I do below. On average, entrepreneurs also earn more than

[12]These are kernel-smoothed densities, so some kinks are smoothed out, e.g. a dip that only slightly appears at 13, representing the completion of secondary education. The modal formal entrepreneur has 17 years of education, corresponding to a "Licence" or some post-secondary vocational training.

workers. Consistent with the distribution of years of education the right panel of Figure 4 shows that while the earning distribution of informal entrepreneurs fairly dominates the workers earning distribution, formal entrepreneurs' earnings are largely above both.

Figure 5: Distribution of Log Initial Wealth by Occupation



As explained in the theory, initial wealth plays a key role in determining the amount of capital that a household is able to borrow from financial institutions. While the survey does not report households ex-ante total wealth, there are retrospective questions about household belongings acquired prior to starting their activity. I use this information to build a measure of household initial wealth by computing the net market value of their total initial belongings.[13] This measure of wealth is positively correlated with various measures of the quality of habitat, showing that it is reasonable.[14] Figure 5 shows the distribution of log initial wealth by occupation. Clearly, entrepreneurs are initially wealthier than workers, as the bulk of the distribution is more concentrated on higher values of the wealth range.

3.3 Regression Analysis

In this section, I use the Cameroon data to test some of the model predictions. I examine how entrepreneurship and informality are affected by financial constraints and observable entrepreneurial skills. In particular, I consider to what extent the likelihood of starting a formal or an informal business is related to education, parent occupation (both of which are associated with entrepreneurial skills as suggested by the descriptive statistics), and household initial wealth. I also added Experience (measured as the total time, in years, spent in the labor market) and an indicator variable for Female.

[13]These belongings are household durable goods including vehicles, TVs, radios, DVD/CD players, fridges, freezers, gas cookers, fans, sewing machines, mobile phones, computers, electric irons, number of houses owned by the household. A better proxy for initial wealth would have been the amount of inherited wealth as in Blanchflower and Oswald (1998). But this information is not available.

[14]In the survey, the quality of the habitat is assessed by reporting the type of housing, the type of walls, roof and floor material of the house, and access to clean water.

Table 2: Linear probability estimates – Choice between occupations

Variable	Formal (vs. Informal)		Formal (vs. Worker)		Informal (vs. Worker)	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	-0.2868***	0.1929**	-0.0106**	0.0225**	0.0713**	-0.0312
	(0.0473)	(0.0968)	(0.0051)	(0.0101)	(0.0124)	(0.0249)
Education	0.0201***	-0.0629***	0.0008***	0.0045***	0.0008*	0.0189***
	(0.0028)	(0.0131)	(0.0003)	(0.0015)	(0.0005)	(0.0036)
Education^2/10	–	0.0251***	–	0.0021***	–	-0.0071
		(0.0045)		(0.0005)		(0.0113)
Parent	0.4467***	0.4240***	0.0162***	0.0159***	0.0667***	0.0639***
	(0.0409)	(0.0400)	(0.0041)	(0.00403)	(0.0101)	(0.0101)
Wealth	0.0077***	0.0084***	0.0032	0.0010	0.0015***	0.0019**
	(0.0008)	(0.0015)	(0.0022)	(0.0013)	(0.0004)	(0.0009)
Experience/10	0.0058***	0.0049***	0.0000	0.0000	0.0001	0.0004
	(0.0017)	(0.0016)	(0.0002)	(0.0001)	(0.0004)	(0.0004)
Female	-0.0201	-0.0116	-0.0041	-0.0036	-0.0211***	-0.0229***
	(0.0248)	(0.0241)	(0.0029)	(0.0029)	(0.0070)	(0.0070)
Obs.	489	489	5688	5688	6047	6047
Adj-R^2	0.467	0.501	0.088	0.099	0.009	0.014

Notes. In the first panel (Formal vs Informal) the dependent variable is 1 if an individual is a formal entrepreneur and 0 if he is an informal entrepreneur; in the second panel (Formal vs Worker) the dependent variable is 1 if an individual is a formal entrepreneur and 0 if he is a worker; and in the third panel (Informal vs Worker) the dependent variable is 1 if an individual is an informal entrepreneur and 0 if he is a worker. Standard errors are in parenthesis. * $p < 0.001$, ** $p < 0.05$, *** $p< 0.01$.

Table 2 summarizes results from linear probability regressions of occupational choice. The first panel (Formal vs. Informal), reports the probability of formalizing, the second panel (Formal vs. Workers) reports the probability of creating a new formal firm, and the third panel (Informal vs. Workers) reports the probability of creating a new informal firm. For each set of regressions, Column (1) presents the baseline estimation while Column (2) includes a quadratic term for education to assess a possible non-monotonic effect of this skills proxy as implied by the theory (see Proposition 2(ii)). Each regression also controls for several other variables such as marital status, urban/rural dummies, physical disability (not reported). Consistent with the theory, the results show that education (which acts as one of the proxies for talent) is critical for both formalization and entrepreneurship. The baseline results show that education is positively and significantly associated with the probability of formal entrepreneurship. Interestingly, when a quadratic term for education is added in the regressions, the associated coefficient is significantly positive in the first panel while the initial coefficient becomes negative (see Column (2) of the first panel), and this result remains robust even when controlling

for various other regressors. This confirms that the probability that an entrepreneur formalizes is non-monotonic and U-shaped in education as suggested by the theory. The U-shaped probability of formalizing obtained in Column (2) of the first panel implies a turning-point estimated at 12.6 years of schooling above which informal entrepreneurs are increasingly likely to formalize. This result is consistent with the descriptive statistics presented in Table 1 showing that about 60% of formal entrepreneurs have 13 years of schooling or more, corresponding to secondary school completion in Cameroon. Likewise, education is also positively associated with entrepreneurship, whether it is informal or formal, as suggested by the theory. The coefficient on Parent is positive and significant in all regressions, and shows that individuals whose parents were entrepreneurs are more likely to become entrepreneurs compared to others. This is mostly due to learning factors as children of entrepreneurs may have received informal business skills from their parents. This evidence has been supported by a number of empirical studies such as Lentz and Laband (1990) and Parker and Van Praag (2012). The results show that the coefficient of initial wealth is significantly positive in all the regressions involving informal entrepreneurs. In particular, the probability of becoming an informal entrepreneur or to formalize increases with initial wealth. This result suggests that initial wealth determines the amount of capital required to start a business, thus reinforcing the evidence of important borrowing constraints.

Finally, I also run an empirical specification that accounts for all choices at the same time. Table 3 presents both ordinary least squares (OLS) and ordered probit regressions of the trichotomous choice between non-entrepreneurship, informal entrepreneurship and formal entrepreneurship. It should be noted that non-entrepreneurship, which comprises subsistence or wage work is a default occupation to which households resort to until they are able to create an informal or formal firm which are more lucrative options, as found in the descriptive statistics (see Figure 4). The ordinal aspect of these earnings across these three occupations suggest that an ordered multinomial choice model can also be used to analyze the reduced-form relationship.[15] I only present a specification that includes a quadratic term in education, as it provides the best fit. Both the OLS estimates and the ordinal probit estimates of the simultaneous choice model reported in Table 3 have similar signs and significance for all regressors. Consistently with previous reduced form results, they confirm that education, parent occupation and initial wealth are positively associated with the probabilities of entrepreneurship and formalization. However, although the construction of the initial wealth variable includes only assets acquired several years before the current occupational choice, there may still be some endogeneity in this measure, given that some unobserved factors may have driven both the initial saving rate and the current occupation. Likewise, education is also likely to be endogenous to occupational choice. These reduce-form regression results should therefore be understood as suggestive evidence provided by conditional correlations rather than pure causal effects.

[15] A nominal multinomial choice model would be equivalent to a combination of pairwise binary choices which are already captured in the regressions given in Table 2

Table 3: OLS and Ordinal Probit Estimates –Trichotomous Choice between Occupations

Variables	OLS Estimates	Ordinal Probit
Education	0.0091**	0.0830***
	(0.0044)	(0.0270)
Education^2/10	-0.0029	-0.0269
	(0.0015)	(0.0930)
Parent	0.0310**	0.265***
	(0.0123)	(0.0778)
Wealth	0.0070***	0.0226***
	(0.0004)	(0.0021)
Experience/10	0.0002	0.0027
	(0.0005)	(0.0030)
Female	-0.0290***	-0.1759***
	(0.0087)	(0.0516)
cut1	–	2.0077***
		(0.1939)
cut2	–	2.9694***
		(0.1991)
Obs	6112	6112
Pseudo R^2	0.0437	0.0411

Notes. The dependent variable is Occupation which is 0 for non-entrepreneurs, 1 for informal entrepreneurs and 2 for formal entrepreneurs. The first panel (OLS Estimates) presents ordinary least squares estimates from a linear regression specification. The second panel (Ordinal Probit) presents maximum likelihood estimates from an ordinal probit specification. Standard errors are in parenthesis. * p $<$ 0.001, ** p $<$ 0.05, *** p$<$ 0.01.

4 Structural Estimation

In this section, I estimate the theoretical model by the Generalized Method of Moments (GMM). The goal is to produce structural estimates that allow to examine the content of the mechanisms implied by the theoretical model and to use the estimated model in counterfactual simulations to quantify the impact of related policies.

4.1 Estimation Technique and Identification

While the initial wealth, z, is observable, entrepreneurial ability, θ, is not observable by the econometrician. Following previous studies, I make the structural assumption that ability depends on personal characteristics such as education and parent occupation

$$\ln \theta = \delta_0 + \delta_1 S + \delta_2 P + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

where S are the years of schooling, P is a dummy indicating whether at least one parent was an entrepreneur. Denote by F, I, and W dummy variables for formal en-

trepreneur, informal entrepreneur, and non-entrepreneur statuses, respectively. These are the three categories of agents observed in the data. Let y be the observed earning and $X = [1\ S\ P\ z]$ the vector of observable covariates. Denote by $\mathbf{1}[\cdot]$ an indicator function that takes the value one when its argument is true and zero otherwise. The net interest rate r is exogenously fixed at its observed annual average of 0.3. If we fix the wage rate at its sample average of 70.3, the remaining structural parameters over which the estimation is performed is defined by $\psi = [\delta_0,\ \delta_1,\ \delta_2,\ \alpha,\ \beta,\ \sigma,\ \lambda]$. These seven parameters are estimated by matching eight moments that are described in Table 4, and their analytical expressions from the model are derived in Appendix A.3. The first five moments match the probabilities of formal and informal entrepreneurship as well as their variance. The last three moments match their corresponding mean occupational earnings and covariance.[16] Matching these occupational probabilities as well as their associated income, variances and covariance ensures that in addition to matching choices at the extensive margin, the estimated parameters are also based on selection at the intensive margin driven by earnings.[17] Specifically, variations in education, parent entrepreneurial status and income are what allow to identify the parameters δ_0, δ_1, δ_2, α, β, σ, whereas variation in initial wealth is what allows to pin down the parameter λ.

Specifically, denote by $M_l(X, \psi)$ the l^{th} model-predicted moment, and by m_l its sample analog, $l = 1, \ldots, L$. The percentage discrepancy of a model-predicted moment from its sample analog is defined by

$$h_l(X, \psi) = \frac{M_l(X, \psi) - m_l}{m_l}, \quad l = 1, \ldots, L$$

The GMM estimator $\widehat{\psi}$ is then obtained by minimizing the criterion[18]

$$Q_n(\psi) = \left[\sum_{i=1}^{n} h(X_i, \psi)/n\right]' \Omega \left[\sum_{i=1}^{n} h(X_i, \psi)/n\right], \tag{7}$$

where $h(X_i, \psi) = [h_1(X_i, \psi),\ h_2(X_i, \psi), \ldots, h_L(X_i, \psi)]'$ is the $L \times 1$ vector of percentage moment discrepancies for household i, $i = 1, \ldots, n$, and Ω is a positive definite weighting matrix that we take as identity to give equal weight to all moments conditions. To evaluate model fit, we use the Hansen's J-test of overidentifying restrictions defined by

$$J_n = n \left[\sum_{i=1}^{n} h(X_i, \widehat{\psi})/n\right]' \left(\text{Var}\left[h(X, \widehat{\psi})\right]\right)^{-1} \left[\sum_{i=1}^{n} h(X_i, \widehat{\psi})/n\right], \tag{8}$$

[16] Matching the probability of non-entrepreneurship is redundant as it can be deducted from the probabilities of formal and informal entrepreneurship.

[17] Other moments such as the expected income of informal entrepreneurs whose parents where non-entrepreneurs and the expected income of formal entrepreneurs whose parents were entrepreneurs are not used to estimate the parameters, but are used as a goodness-of-fit check to evaluate the ability of the estimated model to match untargeted moments.

[18] For the GMM estimation, I use initial values resulting from an extensive grid search over the parameter space and an optimization routine called *particleswarm* which is robust to local extremes.

Table 4: Model Moments and Sample Analogs

	Description	Model $M_l(X_i, \psi)$	Sample m_l
1.	Prob. of formal entrep	$\Pr[F_i = 1\|X_i]$	$\bar{F} = \frac{1}{n}\sum_{i=1}^{n} F_i$
2.	Variance of formal entrep	$\Pr[F_i = 1\|X_i] - \Pr[F_i = 1\|X_i]^2$	$\frac{\sum_{i=1}^{n}(F_i - \bar{F})^2}{n-1}$
3.	Prob. of informal entrep	$\Pr[I_i = 1\|X_i]$	$\bar{I} = \frac{1}{n}\sum_{i=1}^{n} I_i$
4.	Variance of informal entrep	$\Pr[I_i = 1\|X_i] - \Pr[I_i = 1\|X_i]^2$	$\frac{\sum_{i=1}^{n}(I_i - \bar{I})^2}{n-1}$
5.	Variance of non-entrep	$\Pr[W_i = 1\|X_i] - \Pr[W_i = 1\|X_i]^2$	$\frac{\sum_{i=1}^{n}(W_i - \bar{W})^2}{n-1}$
6.	Income of formal entrep	$\mathbf{E}[y_i\|F_i = 1, X_i]$	$\frac{\sum_{i=1}^{n} y_i F_i}{\sum_{i=1}^{n} F_i}$
7.	Income of informal entrep	$\mathbf{E}[y_i\|I_i = 1, X_i]$	$\frac{\sum_{i=1}^{n} y_i I_i}{\sum_{i=1}^{n} I_i}$
8.	Covar. income vs formal	$\mathbf{E}[y_i\|F_i = 1, X_i]\Pr[F_i = 1\|X_i]$	$\frac{\sum_{i=1}^{n} y_i F_i}{n}$

Notes. The table reports averages of the described quantities.

where

$$\text{Var}\left[h(X, \widehat{\psi})\right] = \frac{1}{n}\sum_{i=1}^{n} h(X_i, \widehat{\psi}) h'(X_i, \widehat{\psi}).$$

The statistic J_n has an asymptotic χ^2 distribution with degrees of freedom equal to the number of moments minus the number of parameters (i.e., 1 in this case). This test provides a statistical assessment of whether the vector of moment conditions, i.e., the vector of model's moments discrepancies from the data, are jointly statistically different from the vector of zeros. As a supplement to the J-test, I also compare each estimated moment to its sample counterpart by computing their relative deviations. Matlab minimization routines are used to numerically search for the minimum. The standard errors of the estimated parameters are computed by the bootstrap method using random draws of the original sample with replacement.

4.2 Institutional Parameters

There are three exogenous institutional parameters entering the model: the tax rate τ, the entry cost c, and the probability of detection, p. Table 5 provides useful information that may help to set the values for some of these parameters. In particular, the total

tax rate as a percentage of enterprise profit is readily available and fixed at $\tau = 48.9\%$ as per the measurements of the World Bank's *Doing Business* Survey (2005).[19]

Table 5: Characteristics of the Institutional Environment

Indicator	Starting a Business	Indicator	Paying Taxes
Number of procedures	12	Number of payments/year	44
Number of days	37	Number of days	90
Registration fees (% GNI/capita)	182.5	Total tax rate (% profit)	48.9
Min. capital (% GNI/capita)	232.0		
GNI per capita = \$640 ≈ CFA 320,000		*Source: Doing Business in 2005*	

For the fixed entry cost to formality, I take the registration fees estimated by Doing Business (2005) which I top up by the foregone income incurred during the days spent in the registration office for the procedures. That is,

$$c = \text{Registration Fees} + \text{Number of days} \times \text{Average daily Earnings}$$

From Table 5 the registration fee can be calculated at CFA 582,400, the number of days for the registration procedure is 37, and the average daily earnings in our data is measured at CFA 3,200. Hence, the parameter c is exogenously fixed at $c =$ CFA 700,800 $\approx \$1,402$. Note that this is still an underestimation of the actual fixed cost since it does not account for the amount of bribes, which De Soto (1989) has shown to be significant. Nevertheless, it is already quite high as it represents about 75% of the average yearly profit of informal entrepreneurs in Cameroon (see Table 1).

However, the probability of detection p is not directly available from the data. In order to approximate it, I use information on the number of tax inspections and the level of corruption in the country. Since inspections are on-site, the probability of being in trouble with tax authorities can be proxied by the ratio of the total number of tax inspections over the total number of firms. This should however be discounted by the degree of integrity of tax inspectors. Thus, I calculate the probability p of getting caught and forfeiting the firm's profit by

$$p = \frac{\text{Total number of tax inspections}}{\text{Total number of firms}} \times \text{Degree of Integrity of tax inspectors.}$$

The total number of tax inspections is obtained by multiplying the number of tax inspectors (proxied by the size of the tax department of the Ministry of Finance) by the number of per-period inspections (all available in the Cameroon Statistical Yearbook,

[19]Note that this is for all taxes as a share of profits, not just taxes on profits. So it also includes, for example, social security contributions, national employment fund, etc., and I use it to approximate the total burden on the formal firm. I am also implicitly assuming that this is the effective tax assuming complicance, and not just a statutory tax.

at www.statistics-cameroon.org). The total number of businesses with fixed locations was measured during the 2009 General Enterprise Census. Finally, I use the Corruption Perception Index (CPI) produced by *Transparency International* as a measure of the integrity of tax authorities in Cameroon. The CPI is a score that indicates the perceived level of public sector corruption on a scale of 0 (highly corrupt) to 100% (very clean) (see www.transparency.org for details). The ratio of tax inspections to the number of firms is calculated at 0.0356 whereas the CPI for Cameroon is 0.22. These measures imply a detection probability of $p = 0.78\%$.[20]

4.3 Structural Results

The GMM estimation results of the baseline model are presented in Table 6. The association between entrepreneurial ability and years of education, δ_1 is estimated at 0.035, implying that any additional year of schooling increases entrepreneurial ability by 3.5 percent. This suggests that education may be a reasonable driver of entrepreneurial talent in Cameroon. Having a parent who was an entrepreneur is also an important factor, since the estimated coefficient that relates parent occupation and entrepreneurial ability, δ_2, is positive and significantly estimated at 0.157. It shows that individuals whose parents were entrepreneurs are 15.7% more likely to become entrepreneurs than others. An important finding is that there are binding capital constraints. The degree of

Table 6: Structural GMM Estimates of the Model

Parameter	Name	Estimate	Std. Error
Log Ability Parameters			
Constant	δ_0	-3.7217	0.0205
Education	δ_1	0.0347	0.0026
Parents	δ_2	0.1570	0.0126
Std. Deviation	σ	2.8041	0.0352
Technology and Constraints			
Capital share	α	0.2252	0.074
Labor share	β	0.4601	0.096
Borrowing constraint	λ	7.413	3.523
J-statistic		3.0002	
Number of Obs.		6112	

Standard errors are calculated using bootstrap samples.

financial friction, λ, is estimated at 7.4 meaning that total initial investment can be up to 7.4 times the value of initial wealth. The implication for borrowing constraints should,

[20] While these values of institutional parameters may be imperfect, the simulations performed in Section 5 allow to examine the sensitivity of the results for a wide range of possible values.

however, be understood with caution. It does not necessarily mean that households can borrow up to 7.4 times the value of their wealth in a financial institution. In the data, informal entrepreneurs claimed that about 90% of the total initial business investment came from personal savings, gifts and transfers from family, relatives and ROSCAs (rotating saving and credit associations). Loans from commercial banks and other financial institutions represented the remaining 10%. Thus the above multiplier should be discounted by about 10% to get a better sense of the degree of borrowing constraints.[21] The estimates of α and β mean that a 10 percent increase in the capital devoted to a business leads to a 2.3 percent increase in output, while a 10 percent increase in hired labor increases output by 4.6 percent, respectively. Since returns to capital are usually high in the informal sector (e.g. Udry and Anagol 2006, De Mel et al. 2008) it must be that the estimated value of α obtained here is driven by firms of the formal sector where returns to capital turn out to be very low in most studies (e.g., Alby, Auriol and Nguimkeu 2020).

Before taking the estimated model seriously to a counterfactual policy simulation exercise, it is useful to first examine how well it fits the data. I first calculate a goodness of fit test statistic for the empirical model, as well as percentage discrepancies between the model predicted moments and their sample analogs. I also check how sensitive the results are to various specifications of the log entrepreneurial talent, entry cost and more importantly to the omission of registration costs, and I calculate a goodness of fit test statistic in all these alternative cases.

4.4 Specification Analysis and Robustness Checks

To assess how well the model fits the data, I first compare the model predicted moments and their sample analogs. Table 7 reports the model fit for both the first nine chosen moments that are targeted in the GMM estimation as defined in Table 4 as well as the two last moments that were untargeted. Apart from the proportion of formal entrepreneurs and the non-central covariance between income and formal entrepreneurship which are matched within a 10% discrepancy, all the remaining targeted moments are matched within 5% or lower of their sample counterparts. The observed good fit in the last two untargeted moments can be seen as a validation of the model with data that have not been used directly in the estimation. The results show that the model also fits well in these additional dimensions, within a 5% discrepancy. Second, a more formal standard statistical test is the Hansen's J-statistic that computes the distance between the vector of moments discrepancies and the vector of zeros, as given by Equation (8) above. The J-statistic is computed at 3.0002, which is less than 3.84, the critical value of the χ^2_1 at the 5% statistical level, suggesting that the empirical model is not at odds with the data (see bottom panel of Table 6).

To assess the robustness of the results, I look at how sensitive these results are to

[21] The main reasons evoked for this low involvement in borrowing through financial institutions are high transaction costs, high interest rates and excessive collateral requirements.

Table 7: Matched and Umatched Moments at the GMM estimates

	Description	Model	Data	Discrepancy (%)
1.	Prob. of formal entrep (%)	1.19	1.10	8.18
2.	Variance of formal entrep (%)	0.92	0.88	4.08
3.	Prob. of informal entrep (%)	6.75	6.90	-2.17
4.	Variance of informal entrep (%)	6.78	6.46	4.95
5.	Variance of non-entrep (%)	7.51	7.21	4.16
6.	Income of formal entrep	350.1	353.3	-0.91
7.	Income of informal entrep	75.3	71.9	4.73
8.	Covar. income vs formal	4.166	3.886	7.20
9.	Income of formal entrep, $P = 1$	371.2	374.7	-0.93
10.	Income of informal entrep, $P = 0$	73.5	70.1	4.85

Notes. The table reports averages. Income is in thousands of CFA ($1∼ 500 CFA)

the specification of log ability, to the definition of entrepreneurship, and to the inclusion of registration costs in the model. Panel Mod 1 in Table 8 reports structural estimates when ability is allowed to also vary with initial wealth. The idea is to check whether the baseline specification given in Equation (6) where the unobserved ability distribution is independent of initial wealth may be potentially problematic. Similar to Nguimkeu (2014), the results show that the correlation between log ability and initial wealth is insignificant. Moreover, including wealth does not significantly change the values of

Table 8: Robustness checks

Parameter	Name	Baseline	Mod 1	Mod 2	Mod 3	Mod 4
Constant	δ_0	-3.7217	-4.0431	-2.0432	1.958	2.381
Education	δ_1	0.0347	0.0383	0.0422	0.0201	0.0262
Parents	δ_2	0.1570	0.1752	0.2052	-0.3340	-0.2931
Std. Deviation	σ	2.8041	2.2931	1.9054	5.9789	4.7040
Capital share	α	0.2252	0.2503	0.2637	0.3410	0.4305
Labor share	β	0.4601	0.4714	0.4502	0.4916	0.3856
Capital constraint	λ	7.4130	9.6480	10.841	5.0133	-
J-statistic		3.0002	3.8820	5.0310	30.540	32.906

Notes. Mod 1: Alternative specification of log entrepreneurial ability featuring initial wealth.
Mod 2: Alternative definition of informal entrep based on whether business location is fixed.
Mod 3: Alternative model specification where entry registration costs are completely omitted.
Mod 4: Alternative model with no entry costs and capital-dependent detection probability.

baseline coefficients while the overall model fit deteriorates. This confirms that wealth is not acting as a proxy for entrepreneurial talent. The same is true for the cost function. Alternative specifications that excluded the schooling or parent covariates from log talent

yielded worse model fits. Panel Mod 2 reports estimates when informal entrepreneurship is defined based on whether or not an employer or pure self-employed operates in a fixed location (excluding informal car or motorcycle taxi drivers). With this definition, the fraction of informal entrepreneurs in the sample increases to 7.4% and the predicted proportions of formal entrepreneurs, informal entrepreneurs and non-entrepreneurs are now 1.2%, 7.3% and 91.8% respectively, but the model fit slightly deteriorates compared to the baseline, presumably because some subsistence activities may enter in this definition.

Finally, I assess how the model would perform should the entry registration costs be omitted as in some previous studies, and the probability being a step function of capital as in Ordonez (2014). Panel Mod 3 reports the structural estimates of the model where the registration costs channel is entirely shut down. The results show that the estimated coefficients are substantially different from those previously obtained in Baseline, Mod 1 and Mod 2, and the J-statistic is much larger, indicating a substantial deterioration of fit. The estimated probabilities obtained in this scenario predict a proportion of 10% for formal entrepreneurs, 3% for informal entrepreneurs and 87% for workers which are far from the actual sizes. The test statistic for this alternative specification is computed at 30.54, thus rejected by the data at any conventional level. Panel Mod 4 reports the structural estimates of a model where there is no entry cost and the detection probability is 0 for values of capital below a threshold b and is 1 otherwise.[22] Likewise, this alternative model is also rejected by the data, given the specification test statistic computed at 32.906 against a χ^2_1-distribution benchmark.

5 Counterfactual Policy Analysis

The main inefficiencies in Cameroon that motivate this analysis are that there are too few entrepreneurs (formal or informal) and too many informal workers (mostly in subsistence activities), suggesting important misallocations of skills and resources (see, e.g. Figure 12). Indicators of business environment such as taxation, entry regulations, and access to credit show that Cameroon is doing relatively worse than many neighboring countries (World Bank 2005). The results obtained in the preceding sections further confirm that these factors are significant drivers of the above-mentioned inefficiencies. In this section, I perform a set of counterfactual experiments to evaluate the impact of several policies on entrepreneurship and informality with the Cameroon data. The focus is therefore on entry costs, taxation and enforcement.[23] The estimated model is considered at the current equilibrium and departures from this initial state are assessed by evaluating discrepancies that may occur from changes in policy parameters due to various possible reforms. In particular, for each policy change, new occupational choices and earnings,

[22] Without the variability from initial wealth, z, the capital threshold parameter b is difficult to identify in the structural model estimation and was hence set a the 90th percentile of the capital of informal firms in the data.

[23] The simulated impacts of micro-lending and business training on informal entrepreneurship can be found in Nguimkeu (2014). So I do not examine these policies further and refer the reader to this work.

including the associated tax revenues, are computed assuming the same distribution of wealth and the same exogenously fixed interest rate.

5.1 Registration Reforms

One policy that is becoming very popular in developing countries is the firm registration reform. This policy consists in substantially reducing the cost of registration and the number of procedures required to start a firm (see, e.g., Bruhn 2013 for details on the Mexico case). Since the fixed entry cost to formal entrepreneurship is captured by the institutional parameter, c, in the model, the counterfactual experiment consists in examining how the equilibrium would change should there be a reduction of c from the current level to a smaller amount. The policy is thus implemented as

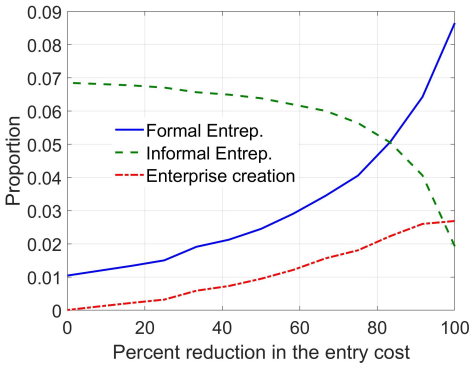
$$c' = c - b, \qquad 0 \leq b < c,$$

where b represents the decrement in the entry cost implied by the reform, at the expense of government revenues.

The aggregate impact of this registration reform with the Cameroon data within the context of the model is depicted in Figure 6. The effect is quantified for a range of percent cost reduction, b/c, starting from 0%, the current state as produced by the structural estimates, to 100%, the idealistic state where the entry cost is zero or insignificant. The left panel of Figure 6 shows the variation in the fraction of formal entreprises, informal entreprises and new enterprise creation. As for increased reductions, the fractions of formal entrepreneurs and new enterprises increase while the fraction of informal entrepreneurs decreases. In particular, a 50% decrease in registration costs (i.e b/c goes from 0 to 0.5) doubles the proportion of formal enterprises, through both formalization of informal firms and new formal enterprise creation. The results depicted in Figure 7 show how these effects would vary by levels of education and initial wealth. The left panel shows that while individuals with 17 years of schooling or higher are the most likely to formalize or create new formal firms as a result of a reduction in the entry cost, individuals with 6 years of schooling or less (i.e., with primary school education) would not be sensitive to this policy and would prefer to stay informal or in non-entrepreneurial activities, especially for entry costs reductions that are below 50 percent of the current cost. The right panel of Figure 7 shows that individuals at the first quartile of the initial wealth distribution are more likely to formalize as a result of a reduction of the entry cost. Individuals at the second and third quartiles are almost insensitive to this policy. These results show that changes in the entry cost appeals more to individuals with low levels of education or high initial wealth.

The right panel of Figure 6 depicts the variation in aggregate income gains, computed as the total income gain from all sectors and the tax revenue gains, computed as the total tax revenues net from the foregone registration fees due to the reform (right vertical axis). Results show that both quantities eventually increase with increased reductions in the entry costs. In particular, a 50% reduction in registration costs increases aggregate

Figure 6: Aggregate Impact of a Registration Reform



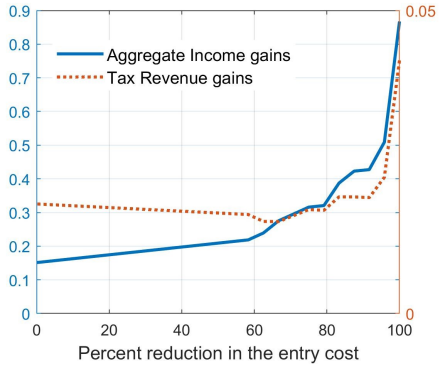
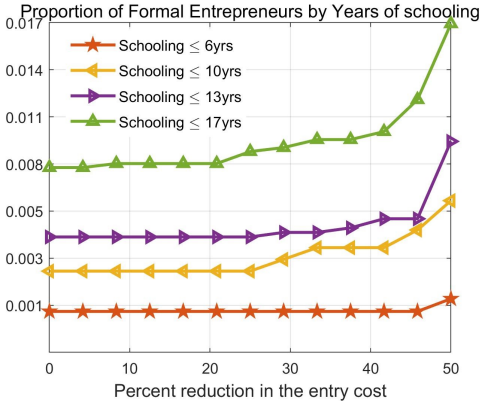
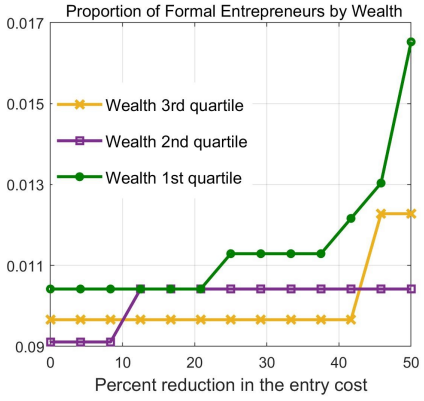


Figure 7: Heterogenous Impact of a Registration Reform on Formal Entrepreneurship





income by 15%. The net tax revenues slightly decrease for small reductions in entry cost, but eventually increase substantially when the reduction is above 60%. These findings are consistent with empirical results obtained by Bruhn (2013) and Kaplan et al. (2011) about the effect of the "System of Fast Opening of Firms" (SARE) on entrepreneurship in Mexico. The effects found here are however significantly higher than those generally found in Latin America (Bruhn and McKenzie 2014) presumably because, as explained earlier, entry costs are much higher and the initial size of the formal sector is much smaller in African countries. Moreover, the counterfactual results for Cameroon show that about 30% of informal entrepreneurs would not formalize even if the entry cost was reduced to zero.

5.2 Tax Reforms

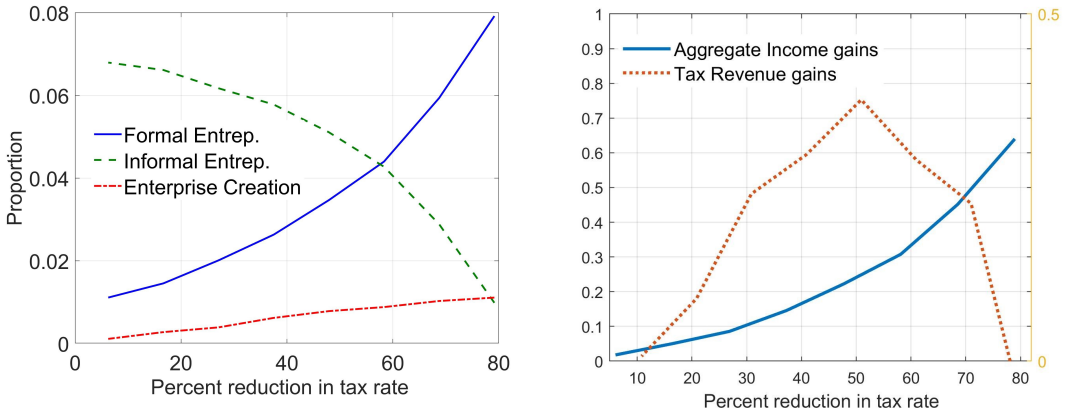
Given the impressive size of the informal sector, a natural question to ask is whether the government is choosing the tax rate in the most efficient way. While high taxation

may increase tax revenues and provide the state with the resources necessary to build law enforcement capacity and the capacity to offer some of the benefits of being formal, excessive taxation may as well drive economic activity out of the formal economy. Here, I investigate the impact of tax reforms on entrepreneurship and informality. Formally, I assume reductions in tax rates of magnitude d, such that

$$\tau' = \tau - d, \qquad 0 \leq d < \tau.$$

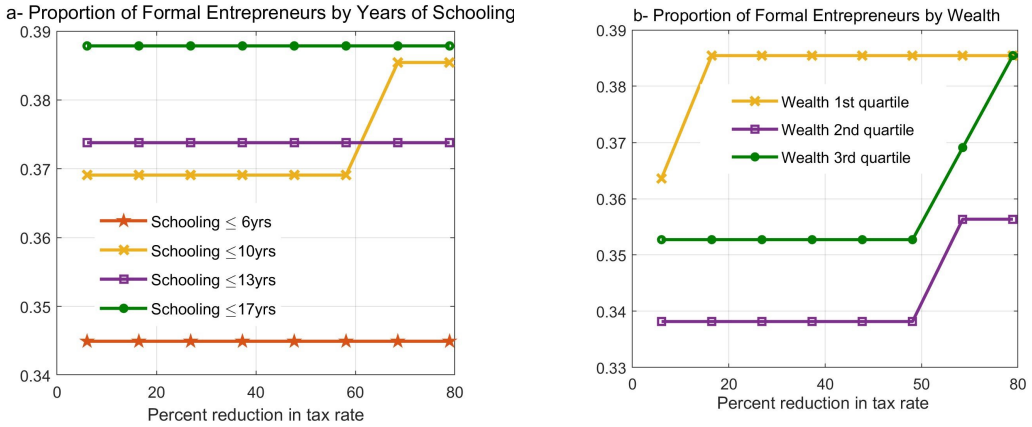
Figure 8 depicts the aggregate effects of variations in tax rates on the economy. The percent reduction in tax rate, d/τ, ranges from 0%, the current state, to about 80% tax reduction. As one would expect, a reduction in taxes increases the fraction of formal entrepreneurs, both in terms of formalization of informal firms as well as new enterprise creation (see left panel of Figure 8). However, only constrained entrepreneurs are affected by this policy whereas unconstrained entrepreneurs remained less sensitive to tax changes. This can be seen more clearly from the heterogeneous impact given in the second panel of Figure 9 where the sensitivity to initial wealth is evaluated. Only individuals at the first quartile of the wealth distribution quickly formalize or create new formal firms for moderate reductions in tax rates. This is because formalization is more beneficial to them (e.g., in terms of access to credit) than to those that are already relatively wealthy. It is only for sufficiently large tax reductions, e.g. 50% reduction or above, that the formal sector becomes attractive to relatively wealthy individuals, especially because at this point the tax rate becomes lower than the enforcement rate. Levels of education are almost insensitive to this policy.

Figure 8: Aggregate Impact of a Tax Reduction Policy



The most interesting pattern is the evidence of the inefficiency of the current tax rate in the Cameroon data as depicted in the right panel of Figure 8. The counterfactual results show that taxation rates have the well-known Laffer's inverted-U impact on government revenues. There exists an efficient tax rate, estimated at 24% (i.e. about

Figure 9: Heterogenous Impact of a Tax Reduction Policy on Formal Entrepreneurship



half of the current tax rate) that generates a maximum tax revenue gains at 35% above the current revenues. At the same time this efficient tax rate induces twice as much formal enterprises and a 25% increase in aggregate income. These results are consistent with those instilled by the SIMPLES tax reduction program implemented in 1996 in Brazil (see Monteiro and Assunção 2012, Fajnzylber et al. 2011). However, just like for the registration policy, the counterfactual simulations for Cameroon show that about 20% of informal entrepreneurs would not formalize even if the tax rate was reduced to the implausible rate of 10% of profits.

5.3 Law Enforcement

While the theoretical literature emphasizes the role of low enforcement as a possible cause of large informal sector (e.g. Dabla-Norris et al. 2008, de Paula and Scheinkman 2011, Ordoñez 2014), there is little empirical evidence on the impacts of enforcement on entrepreneurship and firm informality in African countries. In this study, I use the structural framework to quantify the effect of increased enforcement in Cameroon as follows.

$$p' = p + e, \qquad 0 \leq e < 1 - p,$$

where e represents increments in the probability of detection. Unlike in the preceding experiments, I am unable to compute the costs incurred by enforcing the formality status in this framework. The enforcement probability increments, e, range from 0%, the current state, to 40% corresponding to the state where the probability of detection is close to the tax rate. The left panel of Figure 10 shows that contrary to the previous policies, increased enforcement has a net negative impact on enterprise creation. While some informal firms are formalizing, new formal enterprises are not being created and many informal firms are shutting down. In particular, increased enforcement is more detrimental to entrepreneurs with low initial wealth and lower educational levels. Figure

Figure 10: Aggregate Impact of Law Enforcement

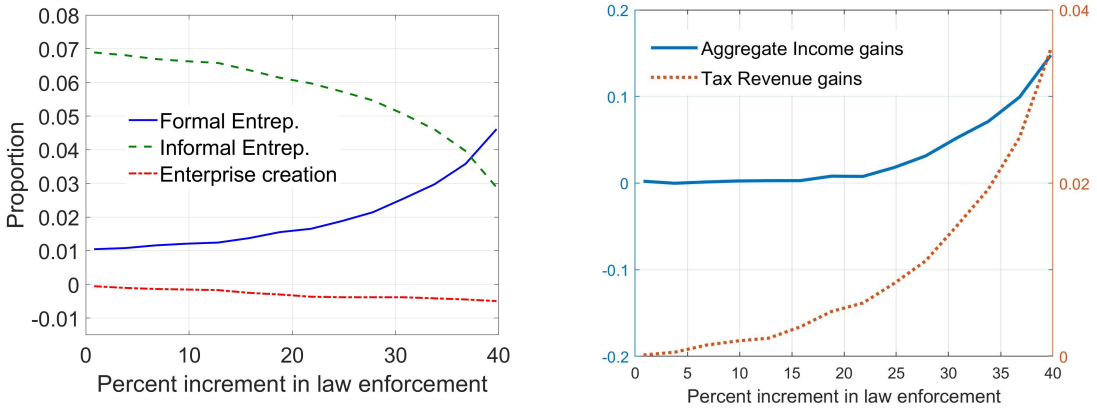
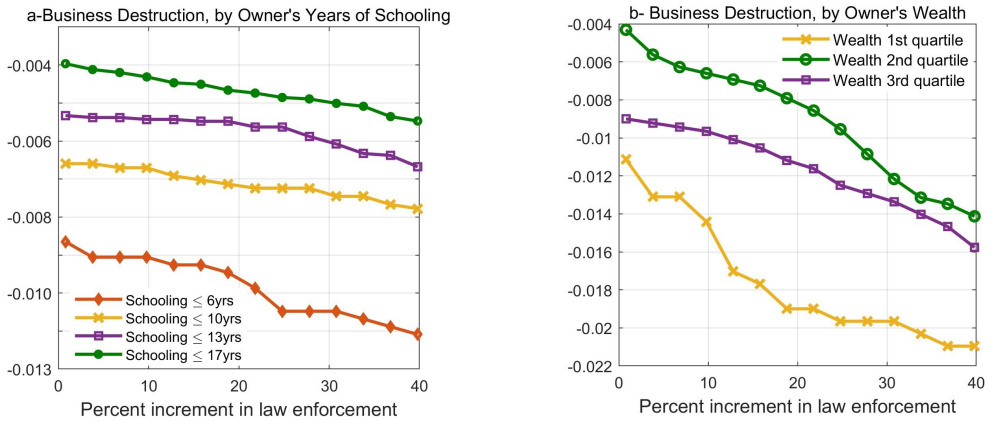


Figure 11: Heterogenous Impact of Law Enforcement on Business Destruction



11 shows that both individuals with primary education or less and those within the first or second quartiles of the initial wealth distribution are the most affected (their rates of business destruction are higher for each percent increment in enforcement). These results show that with increased enforcement, high productivity informal enterprises formalize more compared to informal firms with low productivity. The latter shut down instead and return to non-entrepreneurial work. The counterfactual analysis performed here can also serve as a robustness check for the calibrated probability of detection p. In fact, approximating the detection probability using the size of the tax department (as a proxy for the number of tax inspectors) is quite optimistic. It could be that the number of inspectors is actually much lower than the assumed quantity. Figure 10 shows that for a wide range of reasonable values of p, i.e. between 0 and 10%, the aggregate outcomes

do not change much. This suggests that the results should not change much relative to the benchmark estimates in case the calibration for p turned out to be different from the assumed value of 0.78%, but still falls within a 10% probability.

Finally, it is not obvious how the above policies compare in terms of relative efficiency. The first two exercises (registration reforms and tax reforms) account for both benefits and costs, whereas the last one (enforcement) merely assume improvement in the existing institutional setup. The former may therefore give a better sense of the plausible net gains from these policies compared to the latter where the associated costs are not captured. However, none of these reforms, taken individually, is able to induce, for reasonable values of the underlying policy, more than 30 percent of informal firms to formalize. Even when the tax on profits is set at an unlikely rate of 10% or the entry cost is unrealistically set to zero, there are still about $20 - 30$ percent of informal firms that would not formalize. These are typically small-scale, less productive firms or non-classical occupations (e.g., traditional African spiritualists) that seem to view no net benefits from formalization. As the simulations in Figure 11 show, these small enterprises are also the ones most likely to shut down if formality status is strongly enforced. In the absence of wage-work opportunities for these individuals, the government may prefer to leave them operate rather than have them shut down, given the possible high social costs if they end up unemployed otherwise. In any case, none of these policies are enough on their own and combining them together would possibly produce better effects, especially if they are accompanied by other government measures that would make the formal sector more attractive.

6 Conclusion

The overwhelming importance of informality in African countries poses considerable policy challenges in understanding and promoting entrepreneurship. I present a simple structural model of occupational choice to analyze the role of skills and entry registration costs to the formal entrepreneurial sector, while accounting for other factors such as financial frictions, taxation and enforcement. The main implications of the model are first assessed using reduced form estimates of occupational choice between non-entrepreneurial work, informal and formal entrepreneurship with data from Cameroon, an economy where 90% of the labor force operates in the informal sector. The results show that while initial wealth and average education are important factors associated with the probability of becoming an informal entrepreneur, higher education and parent's entrepreneurial status are the main determinants of formal entrepreneurship. Moreover, it is shown, both theoretically and empirically, that the probability of formalizing is U-shaped in skills, with the turning point corresponding to secondary school completion.

To evaluate the contribution of institutional characteristics on the observed occupational patterns, I structurally estimate the theoretical model using GMM and I check its validity using specification and goodness-of-fit tests. I empirically show that ignoring

the critical role of registration costs substantially undermines the model performance in terms of selection to formal entrepreneurship. Counterfactual simulations are then performed to quantify the impact of various policies. In particular, I found that if the state reduces the registration costs by half, it can induce twice as much formal enterprises and levy more than twice the amount of tax revenues that is currently collected. Similarly, the efficient tax rate that would generate twice as much formal enterprises and produce four-thirds of the current tax revenues is found to be half of the current rate. In contrast, a law enforcement policy whose objective is to increase the probability of detection would have an overall perverse effect in terms of firms shut downs. These counterfactual results are consistent with empirical findings recently obtained in both developed and developing countries. However, my simulations in this African context as shown with Cameroon data also reveals that even in the unrealistic case where entry costs would be zero and the unlikely scenario where tax rates would be as low as 10% of formal profits, there would still be about 20-30 percent of informal entrepreneurs in the economy. This suggests that a different perspective than the traditional formalization objective should be considered when dealing with informality in Africa.

While existing research on informality and the impacts of related policies has mostly focused on countries in Latin America and Southeast Asia, this is, to the best of my knowledge the first work to provide a structural test of the effect of skills, registration costs and other institutional attributes on the formal/informal entrepreneurial dilemma in the context of a sub-Sahara African country. While the use of a static model is an obvious limitation of the study, the choice to resort to it was imposed by data availability. For the same reasons, the role of risk aversion could not be explicitly incorporated in the model. Future work to address these limitations would require that more data be collected in Africa, particularly individual/household/firm level panel data, with detailed information about wealth, socioeconomic activities and program participation.[24]

A Appendix

A.1 Proof of Proposition 1

Proof. (i) The critical threshold $\theta_W(z)$ is solution to the equation $w = \max\{\pi^I(\theta, z), \pi^F(\theta, z)\}$. Denote by θ_W^u and θ_{IW} the unique solutions of the equations $w = \pi^F(\theta, z)$ and $w = \pi^I(\theta, z)$ respectively. Then solving for these equations using the expressions given in (4) and (5) yields $\theta_W^u(z) = \left(\frac{w + rc}{(1-\gamma)(1-\tau)}\right)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{\beta}\right)^{\beta}$ and

$$\theta_{IW}(z) = \begin{cases} \left(\frac{w}{(1-\beta)(1-p)} + \frac{\lambda r z}{1-\beta}\right)^{1-\beta} (\lambda z)^{-\alpha} \left(\frac{w}{\beta}\right)^{\beta} & \text{if} \quad z \leq z^* \\ \left(\frac{w}{(1-\gamma)(1-p)}\right)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{\beta}\right)^{\beta} & \text{otherwise} \end{cases}$$

[24] According to Mckenzie (2011) there is not a single african country for which such data is currently readily available.

where

$$z^* = \frac{\alpha w}{(1-\gamma)(1-p)\lambda r} \tag{9}$$

The desired solution is then defined by $\theta_W(z) = \min\{\theta^u_W(z), \theta_{IW}(z)\}$, that is,

$$\theta_W(z) = \begin{cases} \theta^u_W = \left(\dfrac{w+rc}{(1-\gamma)(1-\tau)}\right)^{1-\gamma}\left(\dfrac{r}{\alpha}\right)^{\alpha}\left(\dfrac{w}{\beta}\right)^{\beta} & \text{if} \quad z \le z^{**} \\ \theta^m_W = \left(\dfrac{w}{(1-\beta)(1-p)} + \dfrac{\lambda r z}{1-\beta}\right)^{1-\beta}(\lambda z)^{-\alpha}\left(\dfrac{w}{\beta}\right)^{\beta} & z^{**} < z \le z^* \\ \theta^l_W = \left(\dfrac{w}{(1-\gamma)(1-p)}\right)^{1-\gamma}\left(\dfrac{r}{\alpha}\right)^{\alpha}\left(\dfrac{w}{\beta}\right)^{\beta} & \text{otherwise,} \end{cases} \tag{10}$$

where the cut-off z^{**} is given by

$$\left(\frac{w}{(1-\beta)(1-p)} + \frac{\lambda r z^{**}}{1-\beta}\right)^{1-\beta}\left(\frac{\lambda r z^{**}}{\alpha}\right)^{-\alpha} = \left(\frac{w+rc)}{(1-\gamma)(1-\tau)}\right)^{1-\gamma} \tag{11}$$

(ii) Consider the function $V^{I,F}(z,\theta) = \pi^F(z,\theta) - \pi^I(z,\theta)$. For $\theta \le \theta_c(z)$, we clearly have $V^{I,F}(z,\theta) < 0$. Suppose $\theta > \theta_c(z)$; then, since $\tau > p$ and $\dfrac{1}{1-\gamma} > \dfrac{1}{1-\beta}$, $V^{I,F}(z,\theta)$ is strictly increasing in θ, and moreover, $\lim_{\theta\to\infty} V^{I,F}(z,\theta) = +\infty$ for any given z. It follows by the intermediate value theorem that there exists a unique threshold $\theta_F = \theta_F(z) \in (\theta_c(z), \ \infty)$ such that $V^{I,F}(z,\theta_F) = 0$, that is,

$$\pi^F(z,\theta_F) = \pi^I(z,\theta_F) \tag{12}$$

□

A.2 Proof of Proposition 2

Proof. (i) Since w is independent of θ and z and both $\pi^I(\theta,z)$ and $\pi^F(\theta,z)$ are increasing in θ, then $V^{W,I}(\theta,z)$ and $V^{W,F}(\theta,z)$ are clearly also increasing in θ.

(ii) For $\theta \le \theta_c(z)$,

$$\frac{\partial V^{I,F}(z,\theta)}{\partial\theta} = -(\tau-p)\theta^{\frac{\gamma}{1-\gamma}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} < 0.$$

For $\theta > \theta_c(z)$,

$$\begin{aligned} \frac{\partial V^{I,F}(z,\theta)}{\partial\theta} &= (1-\tau)\theta^{\frac{\gamma}{1-\gamma}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} - (1-p)\theta^{\frac{\beta}{1-\beta}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta}}(\lambda z)^{\frac{\alpha}{1-\beta}} \\ &= (1-\tau)\theta^{\frac{\beta}{1-\beta}}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}}\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}}\left[\theta^{\frac{\alpha}{(1-\gamma)(1-\beta)}} - \theta_m^{\frac{\alpha}{(1-\gamma)(1-\beta)}}\right], \end{aligned}$$

where

$$\theta_m = \theta_m(z) = \left(\frac{1-p}{1-\tau}\right)^{\frac{(1-\gamma)(1-\beta)}{\alpha}}\left(\frac{r}{\alpha}\right)^{1-\beta}\left(\frac{w}{\beta}\right)^{\beta}(\lambda z)^{1-\gamma}.$$

Hence, when $\theta < \theta_m$, $\dfrac{\partial V^{I,F}(z,\theta)}{\partial\theta} < 0$, and when $\theta \ge \theta_m$, $\dfrac{\partial V^{I,F}(z,\theta)}{\partial\theta} \ge 0$ □

A.3 Model-Predicted Moments

Here, I derive the model quantities that are used to compute the model-predicted moments in the GMM estimation, as described in Table 4. Denote by $\bar{\theta} = \delta_0 + \delta_1 S + \delta_2 P$ the mean function of the log-ability distribution. Given the model predictions:

- The probability of formal entrepreneurship is

$$\Pr[F = 1|X, \psi] = \Pr[\theta > \theta_F(z)] = \Phi\left\{-\frac{\ln \theta_F(z) - \bar{\theta}}{\sigma}\right\} \tag{13}$$

  where $\theta_F(z)$ is given by Equation (12) above.

- The probability of non-entrepreneurial activity is

$$\begin{aligned}\Pr[W = 1|X, \psi] = &\Pr\left[\theta \leq \theta_W^u(z)\right] \mathbf{1}\left[z \leq z^{**}\right] + \Pr\left[\theta \leq \theta_W^l\right] \mathbf{1}\left[z \geq z^*\right] \\ &+ \Pr\left[\theta \leq \theta_W^m(z)\right] \mathbf{1}\left[z^{**} \leq z \leq z^*\right] \\ = &\Phi\left\{\frac{\ln \theta_W^u(z) - \bar{\theta}}{\sigma}\right\} \mathbf{1}\left[z \leq z^{**}\right] + \Phi\left\{\frac{\ln \theta_W^l - \bar{\theta}}{\sigma}\right\} \mathbf{1}\left[z \geq z^*\right] \\ &+ \Phi\left\{\frac{\ln \theta_W^m(z) - \bar{\theta}}{\sigma}\right\} \mathbf{1}\left[z^{**} \leq z \leq z^*\right]\end{aligned} \tag{14}$$

  where the thresholds ability levels $\theta_W^u(z)$, $\theta_W^m(z)$ and θ_W^l are given by Equation (10) above, whereas z^* and z^{**} are given by Equations (9) and (11), respectively.

- The probability of informal entrepreneurship is then obtained by

$$\Pr[I = 1|X, \psi] = 1 - \Pr[F = 1|X, \psi] - \Pr[W = 1|X, \psi] \tag{15}$$

- The mean income of formal entrepreneurs is

$$\begin{aligned}\mathbf{E}[y|F = 1, \psi] = &\mathbf{E}[\pi^F(z, \theta)|X, \psi] \\ = &(1 - \tau)(1 - \gamma)\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} \mathbf{E}\left[\theta^{\frac{1}{1-\gamma}}\right] - rc. \\ = &(1 - \tau)(1 - \gamma)\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} \left[\frac{\bar{\theta}}{1-\gamma} + \frac{\sigma^2}{2(1-\gamma)}\right] - rc.\end{aligned} \tag{16}$$

- The mean income of informal entrepreneurs is

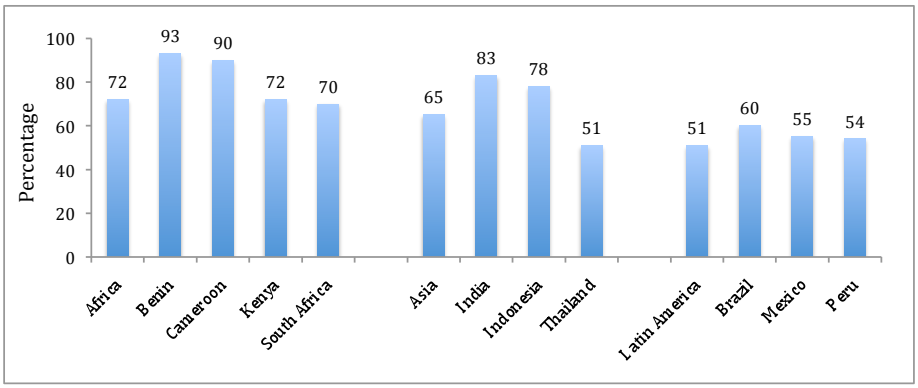
$$\begin{aligned}\mathbf{E}[y|I = 1, \psi] = &\mathbf{E}[\pi^I(z, \theta)|X, \psi] \\ = &\mathbf{E}[\pi_u^I(z, \theta)] \Pr[\theta \leq \theta_c(z)] + \mathbf{E}[\pi_c^I(z, \theta)] \Pr[\theta > \theta_c(z)] \\ = &(1 - p)(1 - \gamma)\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\gamma}} \left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\gamma}} \left[\frac{\bar{\theta}}{1-\gamma} + \frac{\sigma^2}{2(1-\gamma)}\right] \Phi\left\{\frac{\ln \theta_c(z) - \bar{\theta}}{\sigma}\right\} \\ &+ (1 - p)\left((1 - \beta)\left(\frac{\beta}{w}\right)^{\frac{\beta}{1-\beta}} (\lambda z)^{\frac{\alpha}{1-\beta}} \left[\frac{\bar{\theta}}{1-\beta} + \frac{\sigma^2}{2(1-\beta)}\right] - \lambda r z\right) \Phi\left\{-\frac{\ln \theta_c(z) - \bar{\theta}}{\sigma}\right\}\end{aligned} \tag{17}$$

Using the above quantities, the moments and conditional moments required for the GMM estimation can then be readily obtained by taking weighted averages with the appropriate indicator functions, as described in Table 4.

A.4 Additional tables and figures

Figure 12: Importance of the Informal Sector in Developing Countries

Informal employment as % of non-agricultural employment in selected regions and countries, various years (1995-2000)



Source: UN-Habitat (2006)

Table 9: Cost of registration, Education Attainment and Financial Development

Region	Requirement to Start a Formal Business			Avg. Years	Private Credit
	# of Steps	# of Days	Fees (% GNIPC)	of Schooling	to GDP (%)
OECD	6	25	8.0	10.6	124.0
South Asia	9	46	45.4	4.22	35.3
East Asia & the Pacific	8	51	47.1	6.82	46.8
Middle East & North Africa	10	39	51.2	5.90	34.5
Latin America & the Caribbean	11	70	60.4	7.13	41.5
Subsahara Africa	11	63	225.2	4.62	20.1

Source: Various sources. Cost of registration (UN-Habitat 2006), Schooling (Barro and Lee 2001), Private Credit to GDP ratio (Beck et al. 2010)

References

[1] Alby, P., Auriol, E., and Nguimkeu, P. (2020) "Does Social Pressure Hinder Entrepreneurship in Africa? The Forced Mutual Help Hypothesis", Economica 87, 299–327.

[2] Amaral, P. and E. Quentin (2006) "A Competitive Model Of The Informal Sector", Journal of Monetary Economics, 53, 1541-1553.

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

[4] Auriol, E. and Walters. M. (2005) "Taxation base in developing countries", Journal of Public Economics 89, 625-646

[5] Ayittey, George (2006), *Indigenous African Institutions*, 2nd edition, Transnational Publishers.

[6] Barro, R., Lee, J.W., (2001), "International data on educational attainment: updates and implications", Oxford Economic Papers 53 (3).

[7] Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine (2010), "Financial Institutions and Markets across Countries and over Time", World Bank Economic Review 24(1), pp. 77-92.

[8] Blanchflower, D. and Oswald, A. (1998) "What makes an Entrepreneur?" Journal of Labor Economics 16(1): 26-60.

[9] Bruhn, M., (2013) "A tale of two species: Revisiting the effect of registration reform on informal business owners in Mexico", Journal of Development Economics, 103, 275-283.

[10] Bruhn, M. and D. McKenzie (2014) "Entry regulation and the formalization of microenterprises in developing countries", The World Bank Research Observer 29 (2), 186-201.

[11] Cano-Urbina, J. (2015) "The role of the informal sector in the early careers of less-educated workers", Journal of Development Economics 112, 33-55.

[12] Choi, Jieun, Mark Dutz, and Zainab Usman (2019). *The Future of Work In Africa: Harnessing the Potential of Digital Technologies for All*, World Bank Group.

[13] 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, 2008.

[14] De Mel, S., McKenzie, D., and , C. (2008) "Returns to Capital in Microenterprises: Evidence from a Field Experiment", Quarterly Journal of Economics, 123(4), 1329-1372.

[15] de Paula, A. and J. Scheinkman (2011) "The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil", Review of Income and Wealth, Series 57, Special Issue, May 2011.

[16] De Soto, H. (1989). *The Other Path: The Invisible Revolution in the Third World*, I B Tauris & Co Ltd.

[17] Djankov, S., R. LaPorta, F. Lopez-De Silanes, and A. Shleifer (2002) "The Regulation of Entry", Quarterly Journal of Economics, 117, 1-37.

[18] El Badaoui, E., E. Strobl and F. Walsh (2008), "Is There an Informal Employment Wage Penalty? Evidence from South Africa," Economic Development and Cultural Change, 56, 683-710.

[19] Evans, David S. and Boyan Jovanovic (1989) "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." Journal of Political Economy 97, 808- 827.

[20] Fajnzylber, P., W.F. Maloney, and G.V. Montes-Rojas (2011) "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program", Journal of Development Economics, 94(2), 262-276.

[21] Fortin, B., N. Marceau, and L. Savard (1997) "Taxation, Wage Controls and the Informal Sector", Journal of Public Economics, 66, 293-312.

[22] Holtz-Eakin D., D. Joulfaian and H. Rosen (1994) "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints", Journal of Political Economy, 102(1), 53-75.

[23] International Labour Organisation (2002) "Decent Work and the Informal Economy", International Labor Conference, Geneva.

[24] Institut National de la Statistique (2005) *Enquête sur l'emploi et le secteur informel au Cameroun en 2005*, Rapport principal, Yaounde, Cameroon.

[25] Kaplan, D., Piedra, E., Seira, E., (2011) "Entry regulation and business start-ups: evidence from Mexico", Journal of Public Economics 95 1501-1515

[26] Klapper, L., Laeven, L., Rajan, R. (2006) "Entry regulation as a barrier to entrepreneurship", Journal of Financial Economics 82 (3),591-629.

[27] Laporta, R. and A. Schleifer (2014) "Informality and Development", Journal of Economic Perspectives, 28(3), 109-126.

[28] Lentz, B. F., and Laband, D.N. (1990) "Entrepreneurial Success and Occupational Inheritance among Proprietors", Canadian Journal of Economics, 23(3) 563-579

[29] Lucas Jr., R.E. (1978) "On the size distribution of business firms." Bell Journal of Economics 9, 508-523.

[30] Magnac, T. (1991) "Segmented or competitive labor markets", Econometrica 59, 165-187.

[31] Maloney, W.F. (1999) "Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico", World Bank Economic Review 13, 275-302.

[32] McKenzie, D. (2011), "How can we learn whether firm policies are working in Africa? Challenges (and solutions?) for experiments and structural models", Journal of African Economics, 20(4): 600-25.

[33] Mondragón-Vélez, C., Peña, X. (2010) "Business ownership and self-employment in developing economies: the Colombian case", In Lerner, J. and A. Schoar (Eds.), International Differences in Entrepreneurship. University of Chicago Press, 89-127.

[34] Monteiro, J.C.M. and J.J. Assunção (2012) "Coming out of the Shadows? Estimating the Impact of Bureaucracy Simplification and Tax cut on Formality in Brazilian Microenterprises", Journal of Development Economics, 99(1), 105-115.

[35] Nguimkeu, P. (2014) "A structural econometric analysis of the informal sector heterogeneity", Journal of Development Economics 107, 175-191.

[36] Ordoñez, J.C., (2014) "Tax collection, the informal sector, and productivity", Review of Economic Dynamics 17 (2), 262-286.

[37] Parker S. and Van Praag, M. (2010) "The Entrepreneur's Mode of Entry: Business Takeover or New Venture Start", Journal of Business Venturing, 23-82

[38] Paulson, A., R. Townsend and A. Karaivanov (2006) "Distinguishing Limited Commitment from Moral Hazard in a Model of Entrepreneurship", Journal of Political Economy, 144(1), 100-144.

[39] Pratap, S. and E. Quentin (2006) "Are labor markets segmented in developing countries? A semiparametric approach", European Economic Review 50, 1817-41.

[40] Rossi, M.A. (2014), "The impact of individual wealth on posterior political power", Journal of Economic Behavior & Organization, 106, 469-480.

[41] Prado, M., (2011) "Government policy in the formal and informal sectors", European Economic Review 55 (8), 1120-1136.

[42] Schneider, F., Enste, D.H. (2000) "Shadow economies: size, causes and consequences", Journal of Economic Literature 38, 77-114.

[43] Udry, C. and Anagol S. (2006) "The return to capital in Ghana" American Economic Review 96 (2) 388-393.

[44] World Bank (2005) *Doing Business in 2005: Removing Obstacles to Growth*, Washington, DC.