UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Development Conference

Public economics for development

5-6 July 2017 | Maputo, Mozambique

An Estimated Model of Informality and Entrepreneurship

Pierre Nguimkeu

Department of Economics, Andrew Young School of Policy Studies;
Georgia State University
PO Box 3992, Atlanta, GA 30303, USA;
Email: nnguimkeu@gsu.edu.

January 2017

Abstract

What are the main constraints to formal entrepreneurship in Africa, and what kinds of reforms can successfully stimulate business creation and integrate informal businesses into the formal economy? We examine these questions from both a structural econometric approach and a reduced-form multivariate analysis using data from Cameroon. We present and estimate a structural model of occupational choice where agents choose between formal entrepreneurship, informal entrepreneurship and non-entrepreneurial work, and we explicitly incorporate the presence of registration costs, imperfect credit markets and taxation with low enforcement. We show both theoretically and empirically that the decision to formalize is U-shaped in skills with the turning point corresponding to secondary school completion, and that failing to account for registration costs significantly undermines the model performance in terms of selection to the formal sector. Initial wealth and average education are found to be important drivers of informal entrepreneurship while higher education and parent's entrepreneurial status are the main determinants of formal entrepreneurship. Counterfactual policy simulations show that substantial enterprise creation, increased formalization and aggregate income gains can be obtained from registration reforms, reduced taxation and business training. In contrast, a pure enforcement policy against informality has an overall perverse effect on entrepreneurship.

Keywords: Entrepreneurship, Informality, Regulation, Maximum likelihood, Counterfactual analysis.

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

1 Introduction

The importance of entrepreneurship for both developed and developing economies cannot be overemphasized. Entrepreneurship is universally considered a crucial ingredient in promoting and sustaining economic growth because of its potential for creating jobs, delivering innovation and raising productivity. In the context of developing countries, 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 enterprises 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 entrepreneurship is generally low-scale, is largely untaxed and creates important shortfalls in government revenues. At the same time it represents an important avenue for job creation, an incubator for business potential and a stepping stone for accessibility and graduation to the formal economy, especially for less-educated workers (ILO 2002, Cano-Urbina 2015).

This paper investigates how individual and institutional characteristics drive entrepreneurial choice and the formation of formal and informal firms in developing countries. While accounting for traditional factors such as taxation, enforcement and financial constraints, I emphasize the role of entry registration costs and reexamine the effects of skills. These elements not only enable a better fit of the model with the data as evidenced by the likelihood analysis that I perform, but also play critical roles in quantitatively assessing the relevance of policies and programs aiming to promote entrepreneurial growth and increase output while harnessing informality. The underlying structural model that I present assumes heterogeneous agents who choose to become non-entrepreneurs, informal entrepreneurs or formal entrepreneurs based on both their personal characteristics (i.e. skills and wealth endowment) and institutional factors (i.e. entry costs, taxation, enforcement and degree of financial frictions). The implications of the model are tested using reduced form techniques and specification analysis, and structural estimates are provided using data from Cameroon, an economy where 90% of the labor force operates in the informal sector, which accounts for 31% of the gross national product. The estimated model is then used in counterfactual policy simulations to quantify the impact of registration reforms, taxation, enforcement, and business training programs on the Cameroon economy.

The existing research on formal/informal occupational choice and the impacts of policy efforts to increase formality comes from only a few 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 finan-

cial development are relatively lower (see Figure 10 and Table 6 in the Appendix). Also, the heterogeneity 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, using a structural approach. 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). Specifically, the model assumes an institutional environment that accounts for the presence of important registration costs, imperfect credit markets and a tax collection policy with low enforcement. A worker receives a fixed income while an entrepreneur establishes a firm with capital investment and hired labor and realizes profits from a decreasing-returns-to-scale technology. To become formal, an entrepreneur needs to pay a registration cost and once formal, also pays taxes but enjoys better access to credit. In contrast, informal entrepreneurs do not pay taxes but are more likely to face borrowing constraints while also facing a probability of getting caught and forfeiting their profit. The main implications of the model are as follows. Less productive entrepreneurs choose informality whereas the most productive ones choose the formal sector. Formality is costly not only because of high taxation but also because of high registration costs whose affordability is limited to firms promising sufficiently high returns and prospects to grow. This induces fewer entry decision from entrepreneurs with low ability, which in turn increases the number of unproductive enterprises in the informal sector, even when the tax rate is reasonably small.

To structurally estimate the model, I build a likelihood function by matching the probability of each occupation generated by the theoretical model with the corresponding occupational status observed in the data. Evidence is provided using data from the 2005 Cameroon National Survey on Employment and Informal Sector (EESI), which gathered a comprehensive set of information on a cross-sectional representative sample of households.[1] The results confirm that observable talent such as education is a key determinant of entrepreneurial choice. There is however a non-monotonic, U-shaped, relationship between entrepreneurs' education and their decision to formalize. In other words, less educated entrepreneurs find it more profitable to remain informal as their level of education increases. It is only above a certain educational threshold that formality becomes increasingly attractive to informal entrepreneurs with increasing levels of education. 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. This means that informal business training received at home may have an effective impact on entrepreneurial success and suggests that business training may be useful to foster entrepreneurship. Finally, I empirically show through statistical

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

testing and likelihood analysis that failing to incorporate the critical role of entry registration costs substantially deteriorates the model performance in terms of selection into the formal sector.

To quantitatively assess the role of institutional factors such as registration costs, tax rates, and law enforcement, I perform a set of counterfactual policy simulations using the estimated model. I found that a 50% decrease in registration costs (e.g. through government subsidies and/or substantial reduction of administrative steps) 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 optimal tax rate, estimated at 24% (i.e. about half of the current tax rate) generates 30% tax revenue gains above the current level, twice as much formal enterprises and 20% 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 a perverse effect on the economy in terms of enterprise destruction. I also examine the impact of business training programs as recently reviewed by McKenzie and Woodruff (2014). The results indicate that a moderate increase in entrepreneurial skills resulting from business training has a fair impact on entrepreneurship, tax revenues and aggregate income. Put together, these results show that important income gains and efficient allocation of skills can obtain from substantially reducing the cost of registration and optimally choosing the tax rate while fostering entrepreneurial skills and enterprise creation through business training and better access to credit. They also show, however, that these policies alone cannot induce many informal firms to formalize, especially small-scale and less productive ones, suggesting that other government measures should be undertaken to make the formal sector more attractive.

The rest of the paper is organized as follows. Section 2 presents a theoretical model of occupational choice and derives the main implications. Section 3 presents the data, descriptive statistics and reduced form results. Section 4 presents structural estimates, specification tests and robustness checks. Policy simulations are presented in Section 5 and concluding remarks are given in Section 6. The Appendix gathers other technical material.

2 Model Description

The economy is populated with individuals who differ in their skills θ - distributed according to the the CDF $G(\theta)$ - 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. 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, $\lambda \geq 1$, 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. This specification captures the common prediction from models of limited contract enforcement typical of the informal economy where the amount of credit is limited by individual wealth.[2]

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

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

[2]The collateral constraint can be derived from a limited liability problem where at the beginning of the period an individual deposits his wealth $z \geq 0$ in the financial intermediary to rent k units of capital. If this individual can abscond with this capital with probability $1/\lambda$ without any other form of punishment than loosing his collateral z, then in the equilibrium the financial intermediary will rent capital only to the extent that no individual will renege on the rental contract, implying $k \leq \lambda z$.

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:

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

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 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, $c(z)$, 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[3]

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

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

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

The cost $c(z)$ is allowed to be decreasing in the wealth z, to reflect the fact that in developing countries wealthy people have political influence which they could exercise

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

to partially reduce their entry costs to the formal sector (see, e.g., Ayittey 2006 for an extensive discussion and Rossi 2014 for micro-level evidence).

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 the descriptive Table 4 below for 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. 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 developing 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.[4]

2.3 Occupational Choice

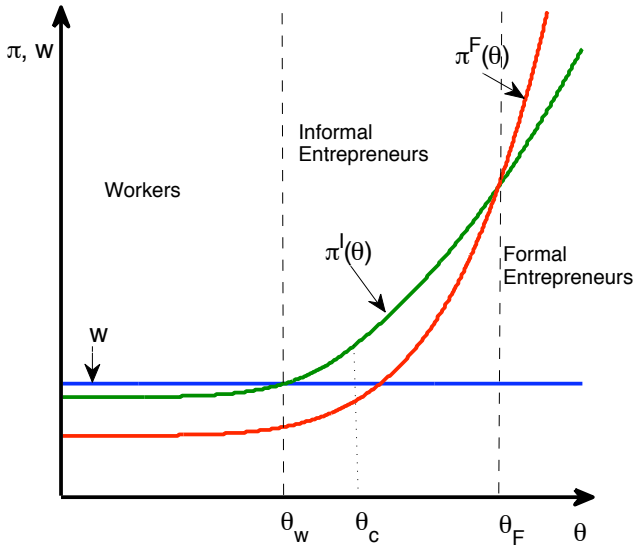
Each individual knows his personal attributes θ and z, market characteristics r and w and institutional factors τ, p and $c(z)$, where it is assumed that $\tau > p$, as commonly observed in developing countries. Given these factors, the agent chooses the occupation

[4]Trying to estimate a leverage parameter in the formal sector turned out to be infeasible, given the already large dimensionality of the problem.

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

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

Figure 1: Earning Functions and Occupations



Earnings functions and related occupational choice are illustrated 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 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 $\theta < \theta_W$ become 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, such that $\pi^I(z,\theta_F) = \pi^F(z,\theta_F)$, 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 chooses to be a wageworker

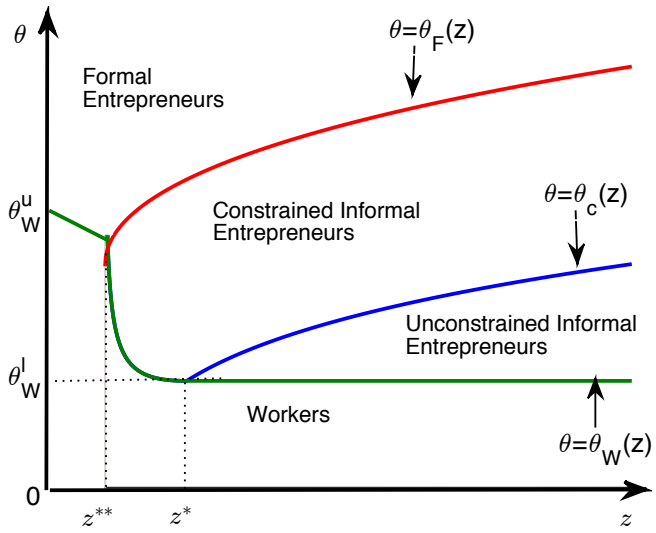
(ii) *If $\theta_W(z) \leq \theta < \theta_F(z)$ the agent is an informal entrepreneur. In particular, when $\theta_W(z) \leq \theta < \theta_c(z)$ he is an unconstrained informal entrepreneur, and when $\theta_c(z) \leq \theta < \theta_F(z)$ he is a constrained entrepreneur.*

(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 $c(z)$. 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 in Figure 2. This figure depicts a partition of the wealth-ability space that

Figure 2: Charaterization of Occupations



forms the basis of the structural estimation discussed in Section 4. Note again how the implications derived here significantly differ from Ordoñez (2014) and other authors. In particular, consistent with empirical evidence both constrained and unconstrained informal entrepreneurs always coexist in my model (see Figure 2). In contrast, Ordoñez (2014) model implies the existence of a corner solution equilibrium where all informal entrepreneurs are constrained.

Total demand for labor from both formal and informal entrepreneurs must equal the supply of labor from workers to generate the labor market equilibrium that the wage w

is required to satisfy:

$$\int_{\theta_W}^{\theta_F} l_I(\theta, w) dG(\theta) + \int_{\theta_F}^{\infty} l_F(\theta, w) dG(\theta) = \int_0^{\theta_W} dG(\theta), \tag{6}$$

where $l_I = l^* \mathbf{1}[\theta < \theta_c] + l^{**} \mathbf{1}[\theta > \theta_c]$ and $l_F = l^*$. The left hand side of Equation (6) is decreasing in w and the right hand side is increasing in w. Moreover, when w tends to 0, the left hand side exceeds the right hand size, and when w tends to ∞, the opposite holds. The existence and unicity of an equilibrium wage then follows from the continuity of these functions. This equilibrium wage is sensitive to the various policies discussed in Section 5, as these policies indirectly affect the labor supply and labor demand curves.

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 $V^{I,F}(z, \theta) = \pi^F(z, \theta) - \pi^I(z, \theta)$. Note that since $\partial c(z)/\partial z \leq 0$ and $\partial \pi^I(z, \theta)/\partial z \geq 0$, the sign of $\partial V^{I,F}(\theta, z)/\partial z$ is ambiguous and is therefore left for empirical examination in the next section. The unambiguous 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 and z.

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

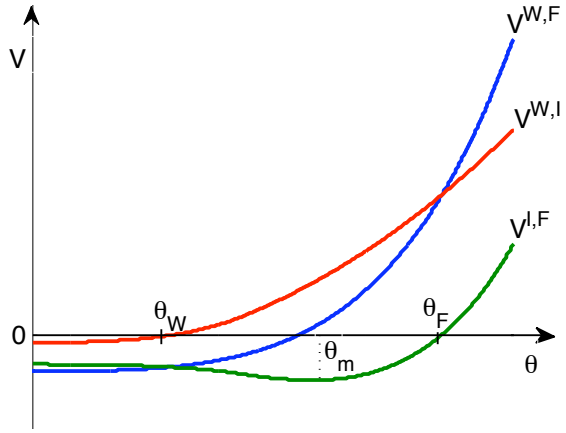
$$\frac{\partial V^{I,F}(\theta, z)}{\partial \theta} < 0 \quad \forall \theta \in [0, \theta_m), \qquad \textit{and} \qquad \frac{\partial V^{I,F}(\theta, z)}{\partial \theta} \geq 0 \quad \forall \theta \in [\theta_m, \infty)$$

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 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.

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

Figure 3: Illustration of the Comparative Payoff Functions



flow between sectors and that individual's expected wage, w, is the same, regardless of the sector. This means that workers have no intrinsic preferences for the sector they work for and only derive their utility from the wages they receive at work. 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) do not find compelling evidence of labor market segmentation between formal and informal sector using data from Colombia, Mexico and Argentina, respectively.

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 and 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 and/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.

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.[5] 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 summarizes the main characteristics of households in the sample according to their occupation (formal entrepreneurs, informal entrepreneurs, workers). The

[5]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).

Table 1: Household Characteristics by Occupations

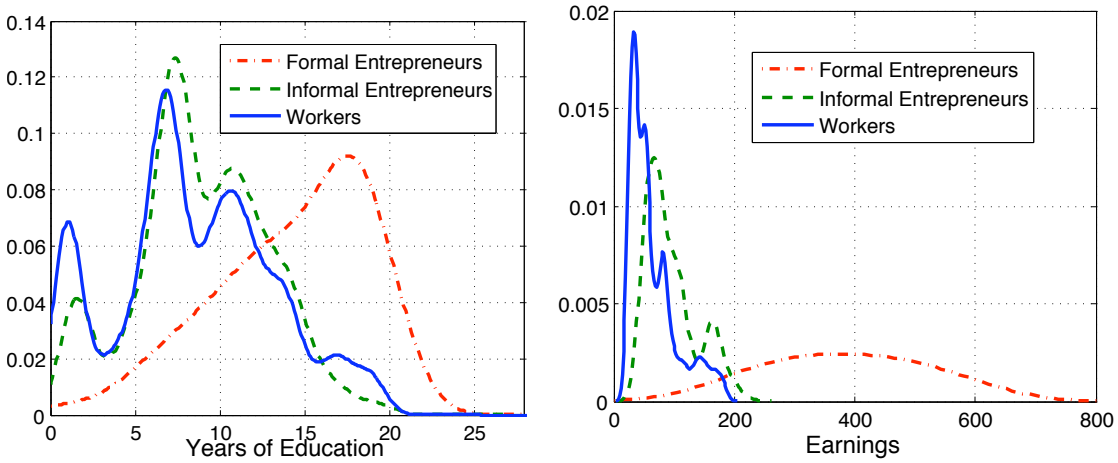
Characteristics	Occupations		
	Formal Entrepreneurs	Informal Entrepreneurs	Wageworkers /Subsisters
Num. of obs.	65	424	5 623
% of sample	1.1%	6.9%	92.0%
% of women	12.3%	37.3%	41.7%
Av. household size	6.0	6.1	5.9
Av. age of head	42.4	37.0	36.1
Years of schooling			
0-6 years	11.1%	41.3%	48.4%
7-12 years	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	77.2	75.3
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)

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 require 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'. The data also shows 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.

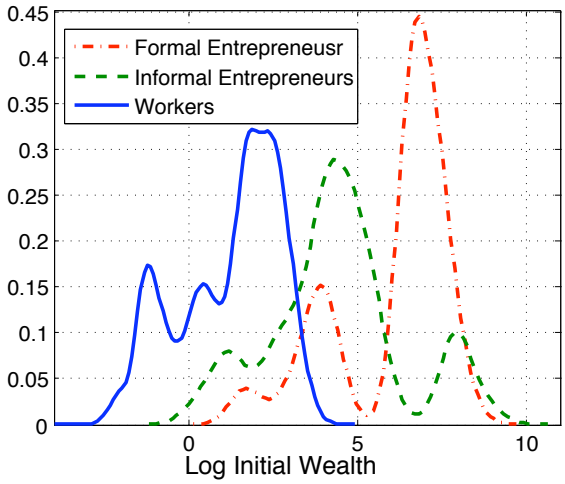
These patterns suggest, as we assess in the structural estimation presented in Section 4, that both education and parents occupation are strongly correlated with unobserved

Figure 4: Distribution of Education and Earnings by Occupation



entrepreneurial ability. On average, entrepreneurs also earn more than workers. Consistent with the distribution of years of education the right panel of Figure 4 shows that while the earning distribution of informal entrepreneurs slightly dominates the workers earning distribution, formal entrepreneurs' earnings are largely above these two.

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.[6] This measure of wealth is positively correlated with various measures of the quality of habitat, confirming its validity.[7] 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 is 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 likely to be correlated with business skills as suggested by the descriptive statistics), and household initial wealth. Table 2 summarizes results from probit regressions of occupational choice. The first panel (Informal Entr. vs. Formal Entr.) presents the choice between informal entrepreneurship and formal entrepreneurship, the second panel (Workers vs. Formal Entr) between non-entrepreneurial work and formal entrepreneurship, and the third panel (Worker vs. Informal Entr) between non-entrepreneurial work and informal entrepreneurship. 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 observable skills 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 confirm that observable skills such as education are critical for formal 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 choice between informal and formal entrepreneurship is non-monotonic and U-shaped in skills as suggested by the theory. However, the quadratic term is insignificant in the remaining panels and its inclusion deteriorates the fit of the choice probability between wagework and formal entrepreneurship (according to the Akaike Information Criterion - AIC). The U-shaped probability of formalizing obtained in Column (2) of the first panel implies a turning-point estimated at 13.4 years of schooling above which informal entrepreneurs are increasingly likely to formalize. This result is consistent with the descriptive statistics presented in Table 1 and the distribution of years of education depicted in Figure 4.

[6]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.

[7]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: Probit Estimates - Choice between Occupations

Variable	Informal Entr. vs. Formal Entr .		Worker vs. Formal Entr.		Worker vs. Informal Entr.	
	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	-3.853***	-3.129***	-3.555***	-3.571***	-1.529***	1.789***
	(0.483)	(0.564)	(0.249)	(0.354)	(0.078)	(0.094)
Education	0.149***	-0.231*	0.046***	0.050	-0.006	0.087
	(0.029)	(0.016)	(0.015)	(0.056)	(0.005)	(0.079)
Education^2/100	-	0.863**	-	-0.017	-	-0.616
	-	(0.426)		(0.256)	-	(0.412)
Parent	1.770***	1.765***	0.435***	0.434***	0.713*	0.693*
	(0.271)	(0.269)	(0.128)	(0.128)	(0.411)	(0.411)
Wealth	0.302***	0.288***	0.310***	0.003***	0.090***	0.113***
	(0.060)	(0.061)	(0.033)	(0.000)	(0.026)	(0.027)
Wealth^2	-	0.288***	-	0.003***	-	0.113***
	-	(0.061)	-	(0.000)	-	(0.027)
Experience	0.040***	0.039***	-0.003	-0.003	0.000	0.002
	(0.015)	(0.015)	(0.008)	(0.007)	(0.003)	(0.003)
Female	-0.256	-0.213	-0.308*	-0.308*	-0.153***	-0.169***
	(0.275)	(0.273)	(0.169)	(0.169)	(0.053)	(0.053)
Number of Obs.	489	489	5688	5688	6047	6047
AIC	194.43	192.54	506.37	508.37	3005.3	2974.5

Notes. In the first panel (Informal Entr. vs. Formal Entr.) the dependent variable is 1 if an individual is a formal entrepreneur and 0 if he is an informal entrepreneur; in the second panel (Workers vs. Formal Entr) the dependent variable is 1 if an individual is a formal entrepreneur and 0 if he is a worker; and in the third panel (Worker vs. Informal Entr) the dependent variable is 1 if an individual is an informal entrepreneur and 0 if he is a worker. Standard errors are in parenthesis.

In contrast, the association between education and the choice probability between non-entrepreneurial work and informal entrepreneurship is insignificant. This result might arise from two conflicting effects due to the coexistence of both extremely low educated individuals and extremely highly educated individuals among workers, compared to a more homogenous average educated group of Informal Entrepreneurs. These heterogenous workers may then be driving the regression coefficient in opposite directions, resulting in an insignificant effect. This intuition is confirmed by the results in Table 3 showing the output of the probit regression of the choice between wagework and informal entrepreneurship performed on two subgroups of workers: those with education below the sampling median of 7 years of schooling (denoted Low Skills) and those with education above this sampling median (denoted High Skills). The coefficient on Parent is positive and significant in all regressions, implying that individuals whose parents were entrepreneurs are more likely to become entrepreneurs compared to others. This

Table 3: Probit Estimates - Choice between Wagework and Informal Entrepreneurship

Variable	High Skills		Low Skills	
	(1)	(2)	(1)	(2)
Intercept	1.301***	5.402***	-2.503***	-1.757***
	(0.151)	(0.406)	(0.111)	(0.112)
Education	-0.260***	-0.270***	0.223***	0.346***
	(0.013)	(0.070)	(0.012)	(0.039)
Education^2/100	-	3.812	-	-2.645
	-	(2.981)	-	(5.221)
Parent	0.739*	0.766*	0.687*	0.656*
	(0.432)	(0.439)	(0.405)	(0.406)
Initial Wealth	0.011***	0.014***	0.018***	0.018***
	(0.003)	(0.003)	(0.004)	(0.004)
Experience	0.002	-0.001	0.007**	0.006**
	(0.005)	(0.005)	(0.004)	(0.003)
Female	-0.230***	-0.227***	-0.202***	-0.154**
	(0.071)	(0.076)	(0.064)	(0.068)
Number of Obs.	3114	3114	3357	3357
AIC	659.46	834.85	569.65	848.54

Notes. The dependent variable is 1 if an individual is an informal entrepreneur and 0 if he is a worker. Standard errors are in parenthesis.

suggests that 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). Finally, the results show that the coefficient of initial wealth is significantly positive in all the regressions. In particular, the probability of becoming entrepreneur 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 financial constraints. 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. These regression results should therefore be understood as suggestive evidence provided by conditional correlations rather than pure causal effects.

4 Structural Estimation

In this section, I estimate the model by maximum likelihood. What is fitted is the probability of being formal entrepreneur, informal entrepreneur or non-entrepreneurial

worker as a function of wealth and ability generated by the model with the actual household occupational status from the data. 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 relevant policies on entrepreneurship and informality.

4.1 Estimation Technique

While the initial wealth, z, is observable, the individual 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

$$\ln \theta = \delta_0 + \delta_1 s + \delta_2 P + \varepsilon \tag{7}$$

where $s = \ln(1 + S)$ is the log of years of schooling S, P is a dummy indicating whether at least one parent was an entrepreneur. In this baseline specification current entrepreneurial ability is assumed independent from initial wealth. But I relax this assumption in the robustness checks to allow for entrepreneurial skills to be correlated with wealth. I also assume a log linear specification for the registration cost function defined by

$$c(z) = c_0 \exp(-c_1 z), \tag{8}$$

where $c_0 > 0$ is an exogenous fixed cost (determined in Section 4.2) and $c_1 \geq 0$ is an elasticity to be estimated. There are three categories of agents observed in the data: Formal entrepreneurs, $(F = 1, I = 0, W = 0)$, Informal entrepreneurs $(F = 0, I = 1, W = 0)$ and non-entrepreneurial workers $(F = 0, I = 0, W = 1)$. Denote by $X = [1 \;\; s \;\; P \;\; z]$ the vector of observable covariates and by $\mathbf{1}[\cdot]$ an indicator function that takes the value one when its argument is true and zero otherwise.

Given the model predictions, the probability of formal entrepreneurship is

$$\Pr[F = 1|X] = \Pr[\theta > \theta_F(z)] = 1 - \Phi\left\{\frac{\ln \theta_F(z) - \delta_0 - \delta_1 s - \delta_2 P}{\sigma}\right\} = H_F(\psi, X) \tag{9}$$

where ψ denotes the vector of all structural parameters of the model and $\theta_F(z)$ is given by Equation (16) in the appendix.[8]

[8]Note that unlike in a standard probit model, the standard error σ is identifiable at least because of the nonlinearity of the model in the parameters.

The probability of non-entrepreneurial activity is given by

$$\begin{aligned}\Pr[W=1|X] = \Pr\left[\theta \leq \theta_W(z)\right] &= \Pr\left[\theta \leq \theta_W^u(z)\right]\mathbf{1}\left[z \leq z^{**}\right] + \Pr\left[\theta \leq \theta_W^l\right]\mathbf{1}\left[z \geq z^*\right] \\ &\quad + \Pr\left[\theta \leq \theta_W^m(z)\right]\mathbf{1}\left[z^{**} \leq z \leq z^*\right] \\ &= \Phi\left\{\frac{\ln\theta_W^u(z) - \delta_0 - \delta_1 s - \delta_2 P}{\sigma}\right\} d_i^u(z) + \Phi\left\{\frac{\ln\theta_W^l - \delta_0 - \delta_1 s - \delta_2 P}{\sigma}\right\} d_i^l(z) \\ &\quad + \Phi\left\{\frac{\ln\theta_W^m(z) - \delta_0 - \delta_1 s - \delta_2 P}{\sigma}\right\}\Big(1 - d_i^u(z) - d_i^l(z)\Big) = H_W(\psi, X)\end{aligned} \tag{10}$$

where $d_i^u(z) = \mathbf{1}\left[z \leq z^{**}\right]$ and $d_i^l(z) = \mathbf{1}\left[z \geq z^*\right]$. The thresholds ability levels $\theta_W^u(z)$, $\theta_W^m(z)$ and θ_W^l are given by Equation (14) in the appendix, while z^* and z^{**} are respectively given by Equations (13) and (15) in the appendix.

The probability of informal entrepreneurship is then obtained by

$$\Pr[I=1|X] = 1 - \Pr[F=1|X] - \Pr[W=1|X] = 1 - H_F(\psi, X) - H_W(\psi, X) = H_I(\psi, X). \tag{11}$$

Given a sample of independent observations of size n, $\{(F_i, I_i, W_i, X_i), i = 1, \ldots, n\}$, the log-likelihood function of the econometric model can therefore be written as:

$$L_n(\psi) = \sum_{i=1}^{n} \left[F_i \ln H_F(\psi, X_i) + I_i \ln H_I(\psi, X_i) + W_i \ln H_W\right] \tag{12}$$

where F_i, I_i and W_i are zero-one indicator variables for the observed occupational choice of household i, and X_i is the vector of their observable characteristics as defined above. Both the interest rate r and the wage rate w are exogenously fixed at their observed averages. The maximum likelihood estimation is therefore performed over the set of parameters $\psi = [\delta_0, \delta_1, \delta_2, \alpha, \beta, \sigma, \lambda, c_1]$. These parameters correspond respectively to the constant term of the ability distribution, δ_0; the interaction between education and ability, δ_1; the interaction between parents occupation and ability, δ_2; the productivity of capital in the production technology, α; the productivity of labor in the production technology, β; the standard deviation of the ability distribution, σ; the degree of financial friction, λ; and the elasticity of entry costs to initial wealth, c_1. With the given observations, maximization routines can be used to search for the maximum numerically. The standard errors of the estimated parameters can be computed by bootstrap methods using draws of the original sample with replacement. Details about the maximization algorithm are presented in the Appendix.

4.2 Institutional Parameters

There are three exogenous institutional parameters entering the model: the tax rate τ, the entry cost c_0, and the probability of detection, p. Table 4 provides useful information that may help to set the values for some of these parameters. In particular, the total

Table 4: 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 $\approx$ CFA 320,000		*Source: Doing Business in 2005*	

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).

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,

$$c_0 = \text{Registration Fees} + \text{Number of days} \times \text{Average daily Earnings}$$

From Table 4 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_0 is exogenously fixed at c_0 = 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 deflated 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, 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 $p = 0.78\%$.[9]

4.3 Structural Results

The results of the maximum likelihood estimation of the baseline model are presented in Table 5, Column (1). The estimated correlation between entrepreneurial ability and education, δ_1 is estimated at 0.4, implying that a 10 percent increase in the amount of schooling increases entrepreneurial ability by 4 percent. This suggests that education may be a reasonable indicator 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 significant. In the second column, the estimated correlation between entrepreneurial ability and assets is statistically insignificant. Thus we can reject the hypothesis that our measure of initial wealth is a positive proxy for entrepreneurial ability. Moreover, the inclusion of initial wealth in the log talent specification deteriorates the model fit (see the discussion in Section 4.4 below). An important finding is that there are binding capital constraints. The degree of financial friction, λ, is estimated at 11.4 meaning that total initial investment can be up to 11.4 times the value of initial wealth. The implication for borrowing constraints should, however, be understood with caution. It does not necessarily mean that agents can borrow up to 10.4 times the value of their wealth. In the data, about 90% of the total initial business investment of informal entrepreneurs 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.[10]

The estimates of α and β mean that a 10 percent increase in the capital devoted to a business leads to a 2.2 percent increase in output, while a 10 percent increase in hired labor increases output by 4.5 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 pinned down 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 2014). For the labor returns coefficient β, on the one hand most informal firms produce output without hiring labor out of their household (corresponding to $\beta = 0$); and on the other hand, formal firms output are more elastic to labor (corresponding to β attending values as high as 0.85 as in Olley and Pakes 1996). The average therefore reasonably falls within the estimated range of 0.45. The elasticity of entry costs to initial

[9] 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.

[10] 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 5: Structural Maximum Likelihood Estimates of the Model

Parameter	Name	Estimate		
		(1)	(2)	(3)
Log Ability Parameters				
Constant	δ_0	-2.8372	-3.0314	1.755
		(0.0118)	(0.0980)	(0.1801)
Education	δ_1	0.4013	0.4270	0.3207
		(0.0210)	(0.0118)	(0.0550)
Parents	δ_2	0.0241	0.0252	-0.2664
		(0.0079)	(0.0118)	(0.3061)
Wealth	δ_3	...	0.0186	0.1706
		...	(0.0214)	(0.0153)
Stand. Deviation	σ	2.4610	2.2692	3.2951
		(0.0380)	(0.0330)	(0.0151)
Technology and Constraints				
Capital share	α	0.2201	0.2252	0.232
		(0.059)	(0.028)	(0.031)
Labor share	β	0.4502	0.4702	0.4506
		(0.092)	(0.022)	(0.049)
Capital constraint	λ	11.417	10.624	7.041
		(3.410)	(3.512)	(3.620)
Cost parameter	c_1	0.0007	0.0006	-
		(0.0004)	(0.0005)	-
Log-likelihood		-1.9602	-1.9692	-2.873
Number of Obs.		6112	6112	6112

Bootstrap standard errors in parenthesis

wealth is significant but very low, estimated at 0.0007.

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 check how sensitive the results are to various specifications of the log entrepreneurial talent and more importantly to the omission of registration costs, and also calculate a goodness of fit test statistic for the empirical model in all these alternative cases.

4.4 Robustness and Goodness-of-fit

A common concern in the entrepreneurship literature is the possible correlation between initial wealth and entrepreneurial ability. For example, to study the relationship between

starting a firm and wealth, some authors (e.g., Holtz-Eakin et al. 1994, Blanchflower and Oswald 1998) use data on inheritances which are likely exogenous. Lacking such data I make an effort to control for endogeneity by using assets acquired by households prior to starting their activity as the initial wealth variable in the model (see also Paulson et al. 2006, Nguimkeu 2014). However, the baseline specification given in Equation (7) where the unobserved ability distribution is independent of wealth may still be potentially problematic. Here, I discuss the implications of relaxing this assumption by allowing interaction between ability θ and initial wealth z, that is, $\ln\theta = \delta_0 + \delta_1 s + \delta_2 P + \delta_3 z + \varepsilon$. The results from re-estimating the model using this specification are given in Column (2) of Table 5. The parameter δ_3 is estimated at 0.0186 with a standard error of 0.0214. This indicates a positive but insignificant correlation between ability and the measure of initial wealth. Moreover, with this new specification, the other coefficients do not significantly change while the overall likelihood slightly deteriorates. This confirms that wealth is not acting as a proxy for entrepreneurial talent. Alternative specifications that exclude the schooling or parent covariates yielded worse likelihoods.

Another concern is to assess how well the model fits the data. First, a comparison between the actual sizes of each occupation with their fitted counterparts shows a close proximity. While the data features 1.1% of formal entrepreneurs, 6.9% of informal entrepreneurs and 92.0% of workers, the estimated model delivers fitted probabilities of 1.3%, 6.8% and 91.3, respectively. Second, a more formal standard statistical maneuver for performing a specification test is to compare observed and expected values, since large departures between them would seemingly indicate lack of fit. I apply a simple Pearson test statistic based on the standardized residuals within each occupation. The test is defined by

$$\widehat{T} = \sum_J \frac{(n_J - n\bar{\hat{H}}_J)^2}{n\bar{\hat{H}}_J} = \sum_J \frac{\left(\sum_{i=1}^n \left[\mathbf{1}(J_i = 1) - H_J(\widehat{\psi}, X_i)\right]\right)^2}{\sum_{i=1}^n H_J(\widehat{\psi}, X_i)}, \quad J \in \{F, I, W\},$$

and has a limiting chi-squared distribution with 2 degrees of freedom under the null of correct model specification. Large values of the statistic would imply that the model is inappropriate. This statistic is computed at 1.906 for the estimated model, that is, a p-value of 0.3856, suggesting that the empirical model is not at odds with the data.

Finally, I assess how the model would perform should the entry registration costs be omitted as in previous studies. Column (3) of Table 5 presents 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 Columns (1) and (2), and the likelihood statistic is substantially lower, indicating an important deterioration of fit. More importantly, the estimated probabilities obtained in this scenario predict a proportion of 9% for formal entrepreneurs, 3% for informal entrepreneurs and 88% for workers which are far from the actual sizes. The test statistic for this alternative specification is computed at 57.340, thus rejected by the data at the 5% significance level.

5 Counterfactual Policy Analysis

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 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, an equilibrium wage rate is computed assuming the same distribution of wealth and the same exogenously fixed interest rate, and agents make their choices based on these factors. Since the theory suggests that entry costs, taxation, enforcement and entrepreneurial skills are the main drivers of entrepreneurship and formality, the focus of the exercise is on each of these attributes.[11]

5.1 Registration Reforms

One policy that is becoming very popular in developing countries is the firm registration reform. The policy consists in substantially reducing the cost of registration and the number of procedures required to start a firm (see Bruhn 2013 for details on the Mexico case). Since the fixed entry cost to formal entrepreneurship is captured by the institutional parameter, c_0, in the model, the counterfactual experiment consists in examining how the equilibrium would change should there be a reduction of c_0 from the current level to a smaller amount. The policy is thus implemented as

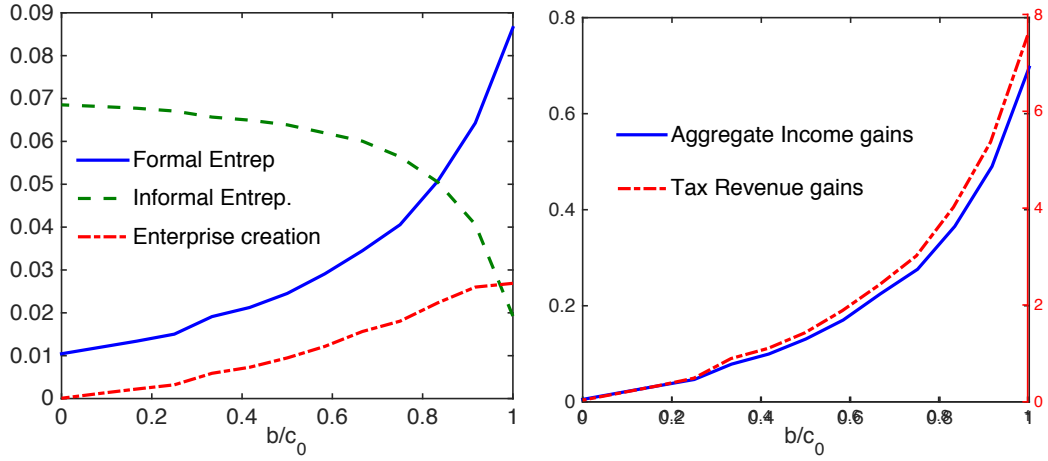
$$c_0' = c_0 - b, \qquad 0 \leq b < c_0,$$

where b represents the decrement in the entry cost implied by the reform.

The 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 relative cost decrements, b/c_0, starting from 0, the current state as produced by the structural estimates, to 1, the idealistic state where the entry cost is zero. The left panel of Figure 6 shows the variation in the fraction of formal entreprises, informal entreprises and new enterprise creation. As b/c_0 increases, 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_0 goes from 0 to 0.5) doubles the proportion of formal enterprises, through both formalization of informal firms and new formal enterprise creation. 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 aggregates increase with decreasing entry costs. In particular, a 50% decrease in registration costs increases aggregate income by 15% and total net tax revenues by more than twice the current amount. These findings

[11] As found in Nguimkeu (2014) the main impact of Microfinance would be a rise of informal entrepreneurship. So I do not examine this policy further and refer the reader to this work. A recent review of microfinance and other financial alternatives in seeding entrepreneurship can be found in Bruton, Khavul, Siegel and Wright (2015).

Figure 6: Impact of a Registration Reform/Subsidy



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.

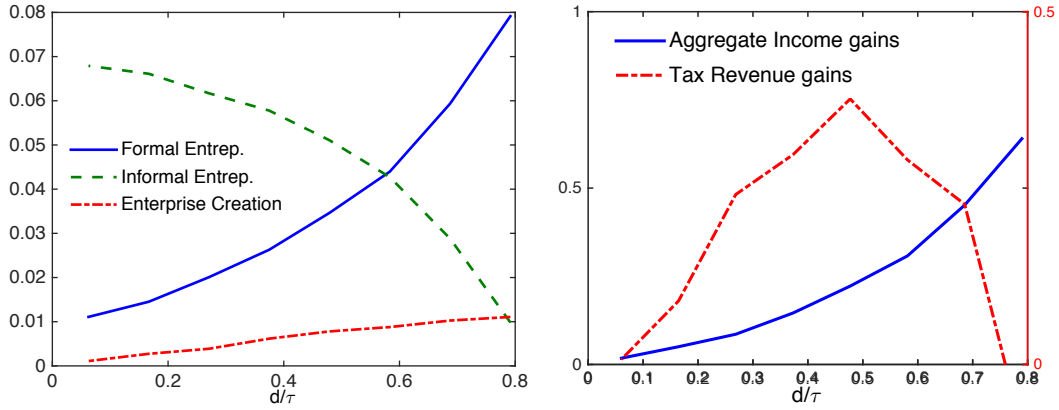
5.2 Tax Reforms

Given the impressive size of the informal sector, a natural question to ask is whether the government is choosing the tax rate in the best possible 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, too much taxation may as well push 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 7 depicts the effects of variations in tax rates on the economy. The relative reduction in tax rate, d/τ, ranges from 0, the current state, to about 0.8, representing an 80% tax reduction. As one would expect, a decrease 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 7). However, only constrained entrepreneurs are affected by this policy whereas unconstrained entrepreneurs remained insensitive to tax changes. The more interesting pattern, however, is the evidence of the sub-optimality of the current tax rate in the Cameroon data as depicted in the right panel of Figure 7. The counterfactual results show that taxation rates have the well-known Laffer's inverted-U impact on government revenues; there exists an optimal tax rate, estimated

Figure 7: Impact of a Tax Reduction Policy



at 24% (i.e. about half of the current tax rate) that generates a maximum tax revenue gains at 30% above the current revenues. At the same time this optimal tax rate induces twice as much formal enterprises and a 20% increase in aggregate income. These results are consistent with those instilled by the SIMPLES tax reduction program implemented in Brazil in 1996 (see Monteiro and Assunção 2012, Fajnzylber et al. 2011).

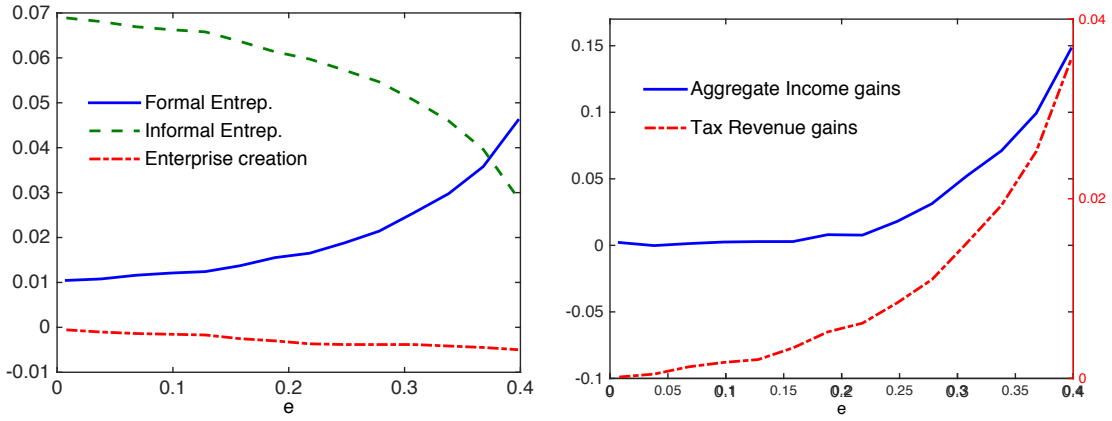
5.3 Law Enforcement

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 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 increases 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 0.41 corresponding to the state where the probability of detection is equal to the tax rate, i.e. 49%. The left panel of Figure 8 shows that contrary to the previous policies, increased enforcement has a net negative impact on enterprises creation. While some informal firms are formalizing, new formal enterprises are not being created and many informal firms are shutting down. The results show that with increased enforcement, high productivity informal enterprises formalize while informal firms with low productivity close down. These results are consistent with experimental evidence by Andrade et al. (2014) who found that increased inspections can provide up to a 27 percentage point increase in the likelihood of formalizing in Belo Horizonte, Brazil.

Figure 8: Impact of Law Enforcement



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 8 shows that for a wide range of reasonable values of p, i.e. between 0 and 0.1, the aggregate outcomes do not change much. This suggests that the results should not change much relative to the benchmark calibration in case the estimates for p turned out to be different from the assumed value of 0.78%.

5.4 Business Training

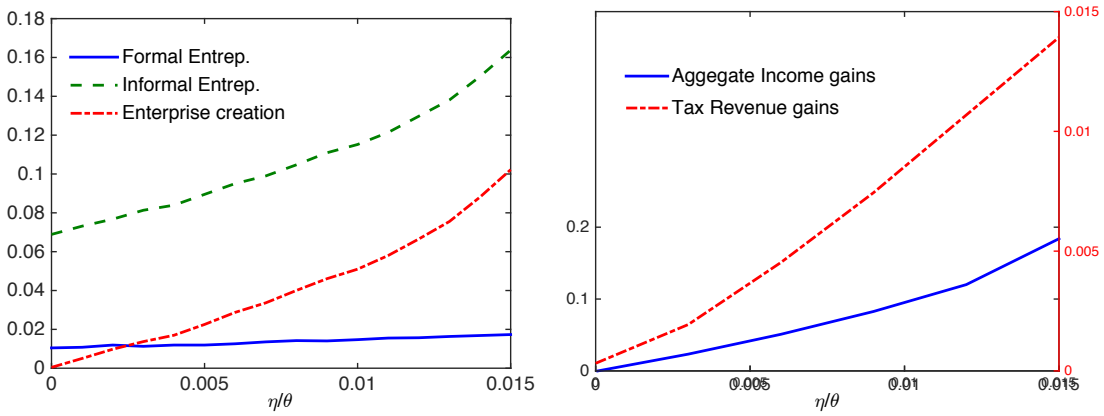
Several studies have found a positive effect of business training on business startups (see McKenzie and Woodruff 2014 for a recent review). However, how does such a policy affect informality? Moreover, what would be the size of the impact of a business training program on entrepreneurship in the present context? The impact of this policy in the model can be quantified by allowing individuals' talent to shift as a result of a business training, that is,

$$\theta' = \theta + \eta, \quad 0 \leq \eta < \infty,$$

where η is the increase in ability inculcated by the training.[12] There are several ways to support such policy. One could take the form of a government support program like the Small Business Innovative Research (SBIR) program (Audretsch, Link and Scott 2002), or a program supported by international organizations such as the Start and Improve Your Business (SIYB) program sponsored by the International Labor Organization. The impact is quantified for a range of increments η in ability starting from 0, the current state of no business training program then progressively increases in terms of fractions of the initial average ability level η/θ. The results presented in Figure 9 indicate that

[12] In this setting, neither can I account for the costs incurred by financing and operating training programs, nor for imperfect take-up.

Figure 9: Impact of Business Training Programs



a small increase (from 0 to 1.5% of average talent) in entrepreneurial skills has some impact on entrepreneurship - mostly through informal enterprise creation - and generate a rather small amount of tax revenues and aggregate income. This result is consistent with the experimental findings of Klinger and Schündeln (2011) in Central America (El Salvador, Guatemala, and Nicaragua).

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 two (enforcement and training) 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. In comparison to the other exercises, a policy of business training appears to be quite insignificant in terms of formalization of informal firms and with weaker results in terms of aggregate income, for reasonable talent shifts. Also, regardless of the type of formalization policy considered (registration, taxation, or enforcement), none of them, taken individually, is able to induce, for rasonable values of the underlying policy, more than 30 percent of informal firms to formalize. Even when the net income tax is set to 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 and less productive firms that seem to view no net benefits from formalizing. As the simulations in Figure 8 show, these small enterprises are also the ones most likely to close down if the law is strongly enforced. In the absence of wagework opportunities for these individuals, the government may prefer to leave them operate rather than have them close 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 should produce better effects, especially if they are accompagnied 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 maximum likelihood and check its validity using specification tests and likelihood analysis. I empirically show that ignoring the critical role of registration costs substantially undermines the model performance in terms of selection between formal and informal 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 optimal 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.

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 choice in the context of a Sub-Saharan 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.[13]

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

7 Appendix

7.1 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 $\theta_W^u(z) = \left(\frac{w + rc(z)}{(1-\gamma)(1-\tau)}\right)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{\beta}\right)^{\beta}$ and

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

where

$$z^* = \frac{\alpha w}{(1-\gamma)(1-p)\lambda r} \tag{13}$$

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

$$\theta_W(z) = \begin{cases} \theta_W^u = \left(\frac{w + rc(z)}{(1-\gamma)(1-\tau)}\right)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{\beta}\right)^{\beta} & \text{if} \quad z \leq z^{**} \\ \theta_W^m = \left(\frac{w}{(1-\beta)(1-p)} + \frac{\lambda r z}{1-\beta}\right)^{1-\beta} (\lambda z)^{-\alpha} \left(\frac{w}{\beta}\right)^{\beta} & z^{**} < z \leq z^* \\ \theta_W^l = \left(\frac{w}{(1-\gamma)(1-p)}\right)^{1-\gamma} \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{\beta}\right)^{\beta} & \text{otherwise,} \end{cases} \tag{14}$$

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

$$\left(\frac{w}{(1-\beta)(1-p)} + \frac{\lambda r z^{**}}{1-\beta}\right)^{1-\beta} \left(\frac{\lambda r z^{**}}{\alpha}\right)^{-\alpha} = \left(\frac{w + rc(z^{**})}{(1-\gamma)(1-\tau)}\right)^{1-\gamma} \tag{15}$$

(ii) Consider the function $V^{I,F}(z, \theta) = \pi^F(z, \theta) - \pi^I(z, \theta)$. For $\theta \leq \theta_c(z)$, we clearly have $V^{I,F}(z, \theta) < 0$. Suppose $\theta > \theta_c(z)$; then, since $\tau > p$ and $\frac{1}{1-\gamma} > \frac{1}{1-\beta}$, $V^{I,F}(z, \theta)$ 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,

$$\pi^F(z, \theta_F) = \pi^I(z, \theta_F) \tag{16}$$

I provide in Section 7.3 an algorithm that numerically solves for this equation and show how it is simultaneously used for the maximum likelihood estimation. □

7.2 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 and z, then $V^{W,I}(\theta, z)$ and $V^{W,F}(\theta, z)$ are clearly also increasing in θ and z.

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

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

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 $\theta \geq \theta_m$, $\dfrac{\partial V^{I,F}(z,\theta)}{\partial \theta} \geq 0$ □

7.3 Algorithm for the maximum likelihood estimation of the model

The following algorithm is used for the likelihood maximization.

1. Define a grid for parameters α and β within $[0.0001, 0.9999]$, say $\alpha(i) = \frac{i}{10000}$, and $\beta(j) = \frac{j}{10000}$, $i, j = 1, 2, \ldots, N$, where $N = 9999$. Define a grid for λ within $[1, 100]$, say $\lambda(k) = 1 + \frac{k-1}{100}$, $k = 1, 2, \ldots, N$, and a grid for c_1 within $[0, 1]$, say $c_1(l) = (l-1)/10000$, $l = 1, \ldots N$.

2. For each pair (i, j) such that $\alpha(i) + \beta(j) < 1$, k and l,

   (a) Solve for the equation $\pi^F(z, \theta_F) - \pi^I(z, \theta_F) = 0$. This yields a value $\theta_F^{ijkl} = \theta_F(\alpha(i), \beta(j), \lambda(k), c_1(l), z)$, for each z.

   (b) Solve the maximization problem

   $$\max_{\{\delta_0, \delta_1, \delta_2, \sigma\}} L_n(\delta_0, \delta_1, \delta_2, \alpha(i), \beta(j), \sigma, \lambda(k), c_1(l))$$

   This yields:
   $L_n^{ijkl} = L_n(\delta_0(i,j,k,l), \delta_1(i,j,k,l), \delta_2(i,j,k,l), \alpha(i), \beta(j), \sigma(i,j,k,l), \lambda(k), c_1(l))$

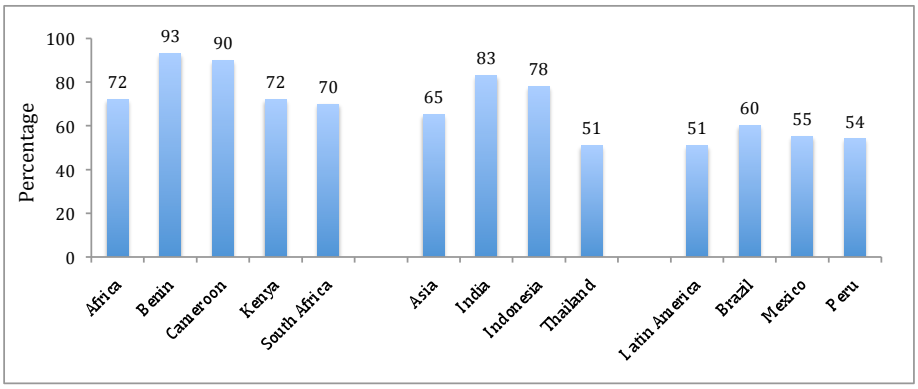
3. The maximum likelihood is then given by $L_n^{i^*,j^*,k^*,l^*} = \max_{1 \leq i,j,k,l \leq N} L_n^{i,j,k,l}$, and the optimal parameter values are $\delta_0(i^*, j^*, k^*, l^*)$, $\delta_1(i^*, j^*, k^*, l^*)$, $\delta_2(i^*, j^*, k^*, l^*)$, $\alpha(i^*), \beta(j^*)$, $\sigma(i^*, j^*, k^*, l^*)$, $\lambda(k^*))$, $c_1(l^*)$, where i^*, j^*, k^*, l^* are the optimal grids indices.

To implement the algorithm I used Matlab routines starting from a wide variety of predetermined guesses. The standard errors of the estimated parameters are computed by bootstrap method using 100 draws of the original sample with replacement.

7.4 Additional tables and figures

Figure 10: 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 6: 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. (2014) "Does Social Pressure Hinder Entrepreneurship in Africa? The Forced Mutual Help Hypothesis", Working Paper.

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

[3] Andrade, G. H., Bruhn, M., McKenzie, D. (2014), "A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms", World Bank Economic Review , 1-31.

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

[5] Audretsch, D. B., A. N. Link, and J.T. Scott (2002) "Public/Private Technology Partnerships: Evaluating SBIR-Supported Research", Research Policy, 31:145-58.

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

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

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

[9] 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.

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

[11] 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.

[12] 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.

[13] Bruton, G., S. Khavul, D. S. Siegel, and M. Wright (2015) "New Financial Alternatives in Seeding Entrepreneurship: Microfinance, Crowdfunding, and Peer-to-Peer Innovations", Entrepreneurship Theory and Practice, 39 (1), 9-26.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

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

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

[21] 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.

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

[23] 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.

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

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

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

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

[28] Klinger, B. and M. Schündeln (2011) "Can Entrepreneurial Activity be Taught? Quasi-Experimental Evidence from Central America", World Development, 39(9), 1592-1610.

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

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

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

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

[33] 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.

[34] 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.

[35] McKenzie, D. and C. Woodruff (2014) "What are we Learning from Business Training and Entrepreneurship Evaluations around the Developing World?", World Bank Research Observer, 29(1), 48-82.

[36] 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.

[37] 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.

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

[39] Olley, G.S. and A. Pakes (1996) "The dynamics of productivity in the telecommunications equipment industry", Econometrica, 64(6), 1263-1297.

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

[41] 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

[42] 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.

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

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

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

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

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

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