A Structural Econometric Analysis of the Informal Sector Heterogeneity*

Pierre E. Nguimkeu†

Georgia State University

Abstract
Understanding the informal sector - that represents about 60 – 90% of urban employment in developing countries - has a significant importance for any strategy and policy interventions aiming to alleviate poverty and improve welfare. I formulate and estimate a model of entrepreneurial choice to address the heterogeneity in occupations and earnings observed within the informal sector. I test the implications of the model with reduced form and nonparametric techniques, and use a structural econometric approach to empirically identify occupational patterns and earnings using data from the Cameroon informal sector. The empirical validity of the structural estimates is tested and the estimated model is used in counterfactual policy simulations to show how microfinance and business training programs can strengthen the efficiency of the informal sector and substantially improve its earning potential.

Keywords: Occupational choice; Financial constraints; Entrepreneurial skills; Maximum likelihood estimation; Informal economy; nonparametric estimation; Specification testing.

JEL Classification: O12, O17, C51, C52, C54.

*I gratefully acknowledge financial support from the Social Science and Humanities Research Council of Canada (SSHRC). I thank the National Institute of Statistics of Cameroon for making the data available to me for this study. Comments from Emmanuelle Auriol, Jean-Paul Azam and participants of the CEA and WEAI conferences are gratefully acknowledged. I am much indebted to Pascal Lavergne and Alexander Karaivanov for intensive discussions and advices.

†Department of Economics, Andrew Young School of Policy Studies, Georgia State University, PO Box 3992 Atlanta, GA 30302-3992, USA, Tel: (1) 404.413.0162, Email: nnguimkeu@gsu.edu
1 Introduction

In developing countries, the informal sector has an overwhelming importance for at least three reasons. First, it is a response to poverty and unemployment; informal entrepreneurs drive job formation through small-business creation.\(^1\) Second, it is considered an incubator for business potential and a stepping stone for accessibility and graduation to the formal economy (ILO 2002). Third, it absorbs the majority of the workforce: about 60 – 90% of the overall employment is in the informal sector (UN-Habitat 2006, ILO 2009). However, in spite of involving so much workforce the informal sector produces only 10 – 40% of the gross national product, essentially because informal subsistence activities characterized by labor-intensive and low-income generating businesses predominate this economy. In fact, a typical feature of the informal sector which is increasingly referred to in the literature as “upper” and “lower” tiers segmentation, is that it includes both urban poor people, who depend on informal subsistence activities for their livelihood, as well as relatively higher-income people most of whom are micro-entrepreneurs running small or medium size enterprises that use capital and hire labor (Fields 1990). Understanding this heterogeneity that characterizes the informal sector therefore has significant implications for any strategy and policy interventions aiming to alleviate poverty and improve economic welfare.

This paper contributes to this objective by providing an econometric analysis of the heterogeneity observed in the informal sector. I formulate and estimate a structural model of occupational choice where informal sector agents choose between subsistence activity and entrepreneurship, and their decision-making process depends on their entrepreneurial skills as well as their access to credit. I test the implications of the model with reduced form and nonparametric techniques, and use a structural econometric approach to empirically identify the relationship among skills, initial wealth distribution, and occupational patterns and earnings using data from the Cameroon informal sector. The empirical validity of the structural estimates is tested using the Lavergne and Nguimkeu (2011) specification test, and the estimated model is used in a counterfactual policy simulation to show how microfinance and business training programs can strengthen the efficiency of the informal sector and substantially improve its earning potential.

The model used to approximate individual behavior is an extension of a widely used empirical specification that is featured in Evans & Jovanovic (1989). Each agent compares the expected gain he or she would obtain from subsistence activity to the expected profit accruing from running a firm. An individual in the subsistence activity receives a fixed income while an entrepreneur establishes a firm with capital investment and realizes profit from a decreasing-returns-to-scale production technology. Entrepreneurship promises higher expected earnings, but requires higher skills and higher starting capital.

\(^1\)More than 90% of new jobs created between 1990 and 1994 in Africa were in the informal sector (Kuchta-Helbling 2000).
that can be supplemented by borrowing. However, because of credit market imperfection, borrowing is exogenously limited to a fixed multiple of initial asset. Initial wealth therefore plays the role of collateral and is required by financial institutions as a strategy to reduce default. Thus low-wealth entrepreneurs may be constrained in their investment and some potential entrepreneurs may be unable to borrow to finance their projects in spite of having a good entrepreneurial idea. The main implication of the model is that entrepreneurial choice and entrepreneurial earnings are positively related not only to skills but also to initial wealth. In order words, if profitable entrepreneurial activities (entrepreneurship) require getting a certain level of initial wealth to use as collateral in the capital market, poorly endowed individuals will not engage in such activities. They will rather remain in subsistence activities that require no capital, so that, the higher the share of initially poor people in the economy, the bigger the size of the subsistence segment.

The main focus of this work is to use data from Cameroon to address the heterogeneity observed in the informal sector whose stylized facts are comparable to those produced by the theoretical model, and then use the estimated model to formulate and perform welfare analysis that allow to quantify the impact of relevant policies. The structural parameters of the model are estimated using maximum likelihood estimation technique. The likelihood function is constructed by matching the expected probability of becoming a micro-entrepreneur generated by the theoretical model (expressed as a function of initial wealth and other individual and market characteristics) with the corresponding household occupational status observed in the data. The econometric analysis is performed for the whole sample, as well as for various data stratifications including urban, rural and metropolitan areas. The model implications are also tested through reduced form estimation (including Probit and Quantile regressions) that make use of multivariate controls as well as non-parametric estimation that do not impose any a priori structure to the relationship among the variables. I use these techniques not only to check the robustness of the structural findings but also to examine to what extend other factors not featured in the theoretical model may influence entrepreneurial choice. A validity check to determine how close the specified model is to the data is also performed by applying the Lavergne & Nguimkeu (2011) model specification test.

The empirical analysis uses data from a cross-sectional sample of households of Cameroon stemming from the 2005 National Survey on Employment and Informal Sector (EESI). This survey is a two-phase nationwide operation conducted by the National Institute of Statistics in partnership with the World Bank. 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 Central and West Africa. Only data from Phase 1 are used here. The estimation results reveal that more than 80% of micro-entrepreneurs are credit constrained and that these constraints are the main source of the differences in informal enterprises
returns, which is consistent with previous results in the literature. More interestingly, they reveal that credit constraints also affect people’s occupational choices and explain the preponderance of massive subsistence activities characterized by low earnings that coexist with micro-enterprises. Observable skill such as education also appears to be a key determinant of occupational patterns and entrepreneurial earnings in the informal sector. This result has been already emphasized in Nguetse (2009), who estimated returns to education in the Cameroon informal sector using this same dataset. These findings then suggest that appropriate financing schemes as well as skills enhancement strategies can improve welfare by improving the shape of allocation of talents across occupations and raising the earning potential of the informal sector. To confirm this intuition, I use the structural estimates to perform two counterfactual policy experiments. The first experiment evaluates the impact of a micro-lending program using an approach similar to Buera et al. (2011) while the other quantifies the impact of a business training program as explained in McKenzie and Woodruff (2012). The results show that by allowing individuals to borrow up to three times the value of the average household wealth the proportion of micro-entrepreneurs can substantially increase by up to 9%, representing twice as much micro-entrepreneurs in the economy, while the average earnings of the overall informal sector can increase by more than 30%. Likewise, a training program that increases the average entrepreneurial ability by up to 15% produces similar results on entrepreneurship. Moreover, these experiments show that the heterogeneity of returns among informal enterprises stemming from differences in levels of starting capital and skills can be considerably lowered.

Several other studies have addressed the heterogeneity observed in the informal sector and the related differences in occupations and earnings. A growing literature argues that social norms and solidarity prevailing in Africa constitute a significant handicap to micro-entrepreneurial investment (see Duflo et al 2009; Nillesen, Beekman & Gato 2011), and a serious barrier to entrepreneurial choice (Alby, Auriol & Nguimkeu 2011; Grimm et al. 2010). However, the most widely documented constraints to informal entrepreneurship in Africa are lack of skills and initial capital requirements (Fields 1990; Cunningham & Maloney 2001). Following various theoretical models of inequality and poverty traps that emphasize the role of wealth distribution as an explanation of the coexistence of high and low returns in economic activities (see Banerjee & Newman 1993, Piketty 1997, Aghion & Bolton 1997, Lloyd-Ellis & Bernardt 2000, Ghatak & Jiang 2002, Gine & Townsend 2004), skills and starting capital requirements have been empirically posited by some authors as being the main sources of the segmentation observed in the informal sector. A noticeable empirical contribution in the Sub-Saharan Africa context is Grimm et al. (2010) who studied how entry costs and starting capital affect marginal returns to capital in micro-enterprises. Using micro-enterprises data from West-African countries they show that different levels of starting capital can explain the observed heterogeneity in capital returns. Their findings, however, fit only entrepreneurial businesses and fail to identify the massive subsistence segment of informal activities characterized by very low capital stocks and low returns that coexists with these micro-enterprises in
Moreover, their framework does not allow to test and quantitatively evaluate the implications and magnitude of policy strategies, as a structural approach would do. This paper is therefore a complementary work to Grimm et al. (2010) in these respects.

The rest of the paper is organized as follows. Section 2 describes a theoretical model of occupational choice under imperfect credit markets and establishes the main implications for the informal sector. Data and reduced form results are presented in Section 3. Structural estimation of the theoretical model as well as nonparametric evaluation are performed in Section 4. Specification analysis and various robustness checks are provided in Section 5. Section 6 quantifies the impact of counterfactual policy experiments, followed by concluding remarks in Section 7. Section 8 provides additional tables, figures and other technical details.

2 Model Description and Implications

In this section, I present a basic model that I estimate and use to evaluate and quantify the impact of economic policies toward the informal sector. The model is an extension of Evans & Jovanovic (1989). This model is particularly attractive because it delivers stylized facts that can be comparable to those observed in the informal sector (constrained firms, unconstrained firms and subsistence workers). The economy is populated by a continuum of agents who are potential entrepreneurs and who differ in their skill \( \theta \) and their level of initial wealth endowment \( z \). At the beginning of the period, agents choose their occupation: whether to work as subsisters or to start a small business (entrepreneurship/micro-entrepreneurship). Since entrepreneurship is costly, their occupation choices are based on their comparative advantage as entrepreneur according to their entrepreneurial ability and their access to capital. Access to capital is however limited by their initial wealth through an exogenous collateral constraint because of imperfect enforceability of capital rental contracts. It is assumed that each household operate only one production unit and devote their entire labor endowment to their activity. The unit fixed gross rate of capital rental is \( r \).

2.1 Subsistence activity

Households working in the subsistence tier require no capital or technology. We assume as in Lloyd-Ellis and Bernhardt (2000), (see also Jeong and Townsend 2007) that expected payoff from subsistence activity is a constant, i.e. independent of entrepreneurial skill \( \theta \).

In fact, Grimm et al. (2010) find extremely high marginal returns of more than 100% at very low levels of capital, which is not consistent with data from subsistence activities in West Africa.

I use the terms agents, individuals and households interchangeably. Likewise for the terms skill, talent, and ability.

As explained later in this section, this assumption can be relaxed to allow the subsistence sector to value entrepreneurial skill; this would lead to a higher critical ability threshold required to transit to
\[ \pi^S = \mu, \]  

(1)

where \( \mu \) is a productivity parameter, which is assumed constant along the subsistence sector. This formulation of the subsistence earning is consistent with the view of Todaro (1997) who characterizes the subsistence sector as an “easy-entry tier with excess capacity and competition driving income down to the average supply of labor of potential new entrants.”

Some examples of subsistence activities include street-vending (e.g., a peanuts stand or a fruit stand), unpaid family labour or unskilled labour with non-specific work relations. Such activities require relatively low set-up costs and relatively low or no skill requirement. Street-vending licenses are easily procured. Location rental fees are nominal. Paid employees are rare (Fields 1990).

2.2 Entrepreneurship

An entrepreneur with skill \( \theta \) uses capital \( k \) to produce goods according to the technology

\[ y = \theta k^\alpha \epsilon \]  

(2)

where \( \alpha \in (0, 1) \) is the elasticity of output with respect to capital. The term \( \epsilon \) is a productivity shock, independent from \( \theta \) and \( z \), with positive support, mean 1 and variance \( \sigma^2 \). At the end of the period, the entrepreneur’s net income is given by

\[ \theta k^\alpha \epsilon - rk \]

We assume, as in Evans & Jovanovic (1989), that households can borrow only up to some fixed multiple of their total wealth, \( \lambda z \), but no more.\(^5\) The maximum amount that can be invested in their entrepreneurial business is then equal to \( \lambda z \), with \( \lambda \geq 1 \) (so that one can always self-finance).\(^6\)

The entrepreneur’s optimal investment capital then solves for his expected profit maximization problem

\[ \max \{ \theta k^\alpha - rk \}, \quad \text{s.t.} \quad 0 \leq k \leq \lambda z, \]

entrepreneurship, but the theoretical predictions would be the same. Keeping it constant, however, is a simplification that renders the identification and estimation of the model easier.

\(^5\)Even in the case where collateral may not be explicitly required in an informal lending like in the rotating saving and credit associations (ROSCA), the level of wealthiness of the borrower may still influence the maximum amount that lenders would be willing to release. Thus, the maximum capital that individuals can borrow may still be seen as depending on their initial wealth.

\(^6\)The assumption that \( k \leq \lambda z \) can also be micro-founded or endogenized. Indeed, with a simple model of limited liability under asymmetric information \( à la \) Banerjee & Newman (1993) it can be shown that the optimal amount that a financial institution would be willing to lend to a borrower is a proportion of the borrower’s initial wealth. This is shown in the Appendix 8.5.
with an interior solution
\[ k^* = \left( \frac{\theta_0}{r} \right)^{1/(1-\alpha)}. \]  
(3)

This solution is feasible only if the entrepreneur is \textit{unconstrained}. This means that \( k^*(\theta) \) is lower than \( \lambda z \). Specifically, entrepreneurs will be unconstrained if, for a given \( z \), their ability \( \theta \) satisfy
\[ \theta \leq \frac{r}{\alpha} (\lambda z)^{1-\alpha} \equiv \bar{\theta}(z) \]  
(4)

Equivalently, the unconstrained condition also means that for a given \( \theta \), their initial wealth \( z \) must satisfy
\[ z \geq \frac{1}{\lambda} \left( \frac{\theta_0}{r} \right)^{1/(1-\alpha)} \equiv \bar{z}(\theta) \]  
(5)

Equation (4) means that for a given level of wealth, entrepreneurs will be unconstrained only if their ability is low enough. Thus, holding household wealth constant, borrowing constraints are more likely to bind for higher skilled households. As for Equation (5), it tells us that for a given level of talent, entrepreneurs are unconstrained only if their initial wealth is high enough. For a given ability level \( \theta \), \textit{constrained} entrepreneurs are those with initial wealth below the critical threshold \( \bar{z}(\theta) \). Or, equivalently, for a given initial wealth level \( z \) they are those with ability higher than \( \bar{\theta}(z) \). For these households, the maximization constraint will be binding so that they will invest \( \lambda z \) in their entrepreneurial business, even though they would like to invest more (see Figure 1).

The household expected profit in entrepreneurship therefore takes the following form

\[ \pi^E(z, \theta) = \begin{cases} 
(1 - \alpha) \theta^{1/(1-\alpha)} \left( \frac{\alpha}{r} \right)^{\alpha/(1-\alpha)} & \text{if } z \geq \bar{z}(\theta) \\
\theta(\lambda z)^{\alpha} - \lambda rz & \text{otherwise.} 
\end{cases} \]  
(6)

2.3 Implications

Since households know their skill \( \theta \) and initial wealth level \( z \) before choosing an occupation, they choose entrepreneurship only if their expected profit from doing so exceeds what they would get by staying in the subsistence. In other words, they become entrepreneur only if their comparative expected profit, \( \pi(z, \theta) = \pi^E(z, \theta) - \pi^S \), is non-negative.

Given Equations (1) and (6) obtained above, an expression for \( \pi(z, \theta) \) can be derived as follows : 

\[ \pi(z, \theta) = \begin{cases} 
(1 - \alpha) \theta^{1/(1-\alpha)} \left( \frac{\alpha}{r} \right)^{\alpha/(1-\alpha)} - \mu & \text{if } z \geq \bar{z}(\theta) \\
\theta(\lambda z)^{\alpha} - \lambda rz - \mu & \text{otherwise.} 
\end{cases} \]  
(7)
Clearly, the comparative gain function, \( \pi(z, \theta) \), does not depend on the initial asset \( z \) when the latter is higher than the critical wealth threshold \( \bar{z} \). This means that wealthier households decisions to start a business does not depend on their capacity to get starting capital once they have enough, but only on whether they have sufficient skills. For further analysis, I define the critical ability level

\[
\theta^* = \left( \frac{\mu}{1-\alpha} \right)^{1-\alpha} \left( \frac{\bar{r}}{\alpha} \right)^{\alpha} \tag{8}
\]

The quantity \( \theta^* \) is a cutoff talent level that determines whether a financially unconstrained household has the ability to start a firm. This cutoff does not depend on wealth.\(^7\) In principle, if there were no financial constraints, all households with talent above this threshold should choose to become entrepreneur. But because of imperfect credit markets, I instead have the following result.

**Proposition 1.** Consider households with initial wealth \( z \) and ability \( \theta \).

(i) For ability level \( \theta \), the function \( \pi(\cdot, \theta) \) defined in (7) is strictly increasing for all \( z \leq \bar{z}(\theta) \), and constant for all \( z > \bar{z}(\theta) \).

(ii) If \( \theta \leq \theta^* \), where \( \theta^* \) is defined in (8), then \( \pi(z, \theta) \leq 0 \) for all \( z \).

(iii) If \( \theta > \theta^* \), there exists a critical wealth threshold \( \tilde{z}(\theta) \in (0, \bar{z}(\theta)) \) such that, \( \pi(z, \theta) < 0 \) \( \forall z < \tilde{z}(\theta) \), and \( \pi(z, \theta) > 0 \) \( \forall z > \tilde{z}(\theta) \).

The behavior of the comparative gain function \( \pi(\cdot, \theta) \), as described in the above proposition, is illustrated in Figure 1.

![Figure 1: Illustration of the comparative profit function](image)

Proposition 1 (i) says that at a given skill level, poorer household’s comparative gain from entrepreneurship increase with their initial wealth endowment, whereas wealthier households comparative gain from starting a business is invariant with their initial wealth. As for Proposition 1 (ii), it tells us that households with relatively low entrepreneurial ability get lower profit from starting a firm no matter the amount of wealth they have, so that they always choose subsistence activity. In contrast, Proposition 1

\(^7\) \( \theta^* \) is obtained by setting the upper right hand side of Equation (7) to zero and solving for \( \theta \).
(iii) emphasizes that having high skills is not enough to become entrepreneur. A minimum level of initial wealth is also required for this purpose. This result is driven by the existence of borrowing constraints in the credit market, which is captured in the model by the finite parameter $\lambda$. To analyze the behavior of entrepreneurs who are indifferent between operating in either segments, we define the critical level of wealth

$$z^* = \frac{1}{\lambda} \frac{\mu}{\alpha} \frac{1}{1 - \alpha r}.$$  \hfill (9)

The wealth level $z^*$ does not depend on households characteristics.\(^8\) For wealth levels below $z^*$ the choice on the margin is between being a subsistence worker and starting a constrained firm. In this case, the household will start a firm if their ability satisfies

$$\theta > (\lambda z)^{-\alpha}[\mu + r \lambda z].$$

On the other hand, for households with initial wealth above $z^*$, their ability $\theta$ would still have to be larger than $\theta^*$ for them to start a business. They would then start an unconstrained firm, if their ability satisfy $\theta^* < \theta < \bar{\theta}$. If the ability of the household is higher than $\bar{\theta}$, they would start a constrained firm. The behavior of this economy thus exhibits selection conditions that are depicted in Figure 2 and summarized in the following proposition, with no further proof.

**Proposition 2.** Consider households with initial wealth $z$ and ability $\theta$.

(i) They choose subsistence activity if $\theta < \max\{\theta^*, (\lambda z)^{-\alpha}[\mu + r \lambda z]\}$

(ii) They start an unconstrained firm if $\theta^* < \theta < \frac{r}{\alpha} (\lambda z)^{1-\alpha}$

(iii) They start a constrained firm if $\theta > \max\{(\lambda z)^{-\alpha}[\mu + r \lambda z], \frac{r}{\alpha} (\lambda z)^{1-\alpha}\}$

(iv) They are indifferent between subsistence and entrepreneurship if

$$\theta = \begin{cases} 
(\lambda z)^{-\alpha}[\mu + r \lambda z], & \text{for } z \leq z^* \\
\theta^*, & \text{otherwise.}
\end{cases}$$  \hfill (10)

The absence of financial constraints in this economy would correspond to the situation where $\lambda = \infty$, or equivalently, $z^* = 0$, and the right hand sides of all the selection conditions stated in Proposition 2 above go to $\infty$. The selection into entrepreneurship then reduces to a simple comparison of the household ability $\theta$ with the cutoff level $\theta^*$ which does not depend on $\lambda$. Thus, without financial constraints, only ability determines who capital should flow to since high $\theta$ implies high marginal productivity of capital. Also, the nature of the selection into occupations as depicted in Figure 2 does not depend on the particular functional form of the production technology specified in Equation (2). Any technology which is increasing in ability at all levels of capital and labor and which

\(^8\) $z^*$ is the value of $\bar{z}$, given by Equation (5), evaluated at $\theta^*$
satisfy standard Inada conditions would yield similar behavior.\textsuperscript{9}

The model described above has some theoretical limitations. First, it does not incorporate the role of risk in the preferences so that the analysis is solely based on profits comparison across sectors. But it is possible to include risk aversion by defining a utility function over returns that account for the stochastic component of the entrepreneur’s production function. I abstract from this extension since the crux of the analysis is the allocation and efficiency of occupations due to skills and credit constraints, and data on production risk are not available. Second, the model assumes that output in the subsistence sector is fixed, for simplicity. Although this is a common assumption is several studies (see, e.g. Lloyd-Ellis and Bernhardt 2000, Jeong and Townsend 2007), one can imagine that skills may play a role to some extent. Including it will raise the critical ability required to become an entrepreneur, given by Equation (8), and possibly increase some of the policy impacts (as I discuss later in Section 6.2), but the main implications of the model would remain unchanged. Finally, the entrepreneur’s production function does not include paid labor in its specification. Since paid labor is rare in the data (only about 5% of the firms hire paid workers as described below) adding labor brings a complication that does not improve the empirical results nor does it change any of the model implications.

3 Data and Reduced form Results

Since its decision point in the enhanced highly indebted poor country initiative by the Bretton Woods institutions in 2000, its implementation of a new Investment Code and the ongoing review of its Industrialization Master Plan (see Mastri 2008), the Cameroon

\textsuperscript{9}For general assumptions on entrepreneurial production functions that work in our context, see Lucas 1978.
government has focused on small and medium-size firms as one important element of its policy to stimulate the economy. This implies that the informal sector which gathers the quasi-totality of such firms together with an amazing share of the workforce (as described below) is at the centre of the attention. This section describes some important features of the data used for this study and assesses the empirical relevance of the model predictions for the Cameroon Informal Sector. The data covers 8540 households of the 10 Cameroon provinces distributed in both urban and rural areas. I provide descriptive statistics of these data and use them to test some implications of the model from Probit and conditional quantile regressions. Since the types of activities and the financing opportunities of people living in urban areas may differ from those in rural areas (for example, rural regions tend to be more agricultural oriented and more populated by smaller banks and credit unions than urban regions), the analysis is performed for the whole sample as well as subsamples of urban regions, “Urban”, and rural regions, “Rural”. I also distinguish a subsample denoted “Metro” that includes Douala and Yaounde, the two main cities of the countries which according to the 2001 Cameroon Household survey, embed about 70% of the overall economic activities of the country.\footnote{ECAM II: Second Cameroon Household Survey administered by the National Institute of Statistics in 2001; available at www.statistics-cameroon.org.}

3.1 Descriptive statistics

The Cameroon National Survey of Employment and Informal Sector (EESI), from which the data are drawn for this empirical investigation, defines informal enterprises as “production units that do not have written formal accounts and/or are not registered with the tax authorities.” Our sample contains 4337 households whose activities belong to the informal sector according to the above definition. The so-defined informal sector accounts for the vast majority of activities and employs about 90% of the Cameroon workforce aged 15 and above. The employment shares of the informal workforce are 60% for agricultural activities and 40% for non-agricultural activities. Households heads are either informal wage earners, purely self-employed or employers with less than six employees.\footnote{Almost 90% of informal jobs come from business with generally less than 6 workers, mostly unpaid family aids} Most of the latter run only a single business and rely heavily on family workers. Only 5% of the businesses hired and paid anyone from outside their family and most of the activities surveyed were started less than 5 years earlier at the time of the survey. Non-agricultural activities are distributed along all the three major sectors of the economy: manufacturing, commerce and services.

One interesting aspect of the EESI is that it provides information about both households characteristics and the characteristics of the micro-enterprises that they possibly run. It is thus an overlap of both household and enterprises surveys, which makes the dataset a comprehensive one that is useful to carry out extensive analysis about households, associated firms and interactions. However, these data do not contain a variable that reports household’s ex-ante total wealth, a key component to our analysis. To deal
with this lack of information, I construct an aggregate index that represents the total wealth index of the household. This aggregate index is created by a Principal Component Analysis from some of the household belongings reported in the questionnaires. Because the analysis requires that this index be representative of the initial wealth of the household, that is, their wealth prior to starting their current activity, only items that were acquired prior starting their activity are accounted for in the computation.\footnote{Although the survey is not dynamic, there are many retrospective questions that can be used to evaluate the length of time a household have been staying in a house/location, unemployed, running a business, using some consumer durables, etc.}

A detailed description of the wealth index can be found in Appendix 8.2.\footnote{A better proxy for initial wealth would have been the amount of inherited wealth as in Blanchflower and Oswald (1998). But the present data do not contain this information.}

There is no clear consensus about the definition of “entrepreneur” in the economic literature, although many empirical studies of entrepreneurship in developed countries associate business ownership and/or self-employment to entrepreneurship (Evans and Jovanovic 1989, Blanchflower and Oswald 1998, Hurst and Lusardi 2004). However, in developing countries where the majority of informal sector actors are self-employed, a distinctive definition for informal entrepreneurship is imperative. Mondragón-Veléz and Peña (2010) thoroughly investigate this issue in the context of Columbia. In line with their findings, I define as entrepreneur an individual who owns a business and employs others even if the employees are unpaid family members. All others, i.e. purely self-employed (those who work just by themselves) as well as workers employed by informal firms or households are considered subsisters. The only exception is for some category of activities that use high start-up capital but can only be operated by a single person at a time (like taxi-drivers). This definition, which is a combination of the above-mentioned, is obviously still not perfect. Some employers might operate at such a low scale that considering them as entrepreneurs is inadequate. On the other hand, given technology and new business models, even a purely self-employed, without employees and with low physical capital could innovate and implement new products so that considering them as subsisters is misleading. Such cases are however infrequent or have insignificant influence in these data as sensitivity analysis with alternate definitions and through simulations further support the definition I am using in this context (see Section 5.3). Examples of micro-entrepreneurs include taxi-drivers, grocers, tailors, carpenters, car mechanics who own shops, etc. Examples of subsisters include street vendors, unpaid family workers, housekeepers, call box dealers, etc.\footnote{Unlike widely used surveys in the U.S. such as the SCF and PSID that ask separate questions to determine occupation and business ownership, the EESI data categorizes individuals as either employers (that is, self-employed individuals who own a business that hires paid or unpaid workers) or self-employed who work alone using a single question about their primary occupation.}

A household is classified as “Entrepreneurial” if at least one member of the household is an entrepreneur in the sense given above.

Table 1 summarizes the characteristics of households in the sample by occupation (Entrepreneurs, Subsisters). Entrepreneurs represent 9.6% of the total sample while sub-
sisters represent 90.4%. About 23.7% of individuals have a secondary activity. Among them, 3.5% do both enterprise and subsistence while the others do multiple subsistence work. The number of paid employee (i.e., hired outside of family) per informal enterprise is about 1.3 for an average wage is 36,330 CFA. Women are very active in the informal sector, especially in the rural areas. They represent more than half of the total informal economy and about 57% of the rural informal sector. However, they are more concentrated in the subsistence activity than they are in entrepreneurial businesses. Entrepreneurial households are on average slightly older but much more educated than households in the subsistence activity. Retrospective questions also allow to calculate transition rate across sectors. A comparison of current versus prior occupations show that about 1% of subsisters were previously entrepreneurs, and 19.8% of entrepreneurs were previously subsisters.

Table 1: Household Characteristics by occupation

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Whole</th>
<th>Rural</th>
<th>Urban</th>
<th>Metro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entr</td>
<td>Subs</td>
<td>Entr</td>
<td>Subs</td>
</tr>
<tr>
<td>Num. of obs.</td>
<td>424</td>
<td>3913</td>
<td>98</td>
<td>1270</td>
</tr>
<tr>
<td>% of sample</td>
<td>9.4%</td>
<td>90.6%</td>
<td>7.3%</td>
<td>92.7%</td>
</tr>
<tr>
<td>% of women</td>
<td>6.4%</td>
<td>93.6%</td>
<td>4.7%</td>
<td>95.3%</td>
</tr>
<tr>
<td>Av. household size</td>
<td>6.1</td>
<td>5.8</td>
<td>6.3</td>
<td>5.7</td>
</tr>
<tr>
<td>Av. age of head</td>
<td>37.3</td>
<td>35.2</td>
<td>38.0</td>
<td>35.8</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>0-5 years</td>
<td>30.3%</td>
<td>45.1%</td>
<td>43.8%</td>
</tr>
<tr>
<td></td>
<td>6-10 years</td>
<td>49.4%</td>
<td>44.7%</td>
<td>43.0%</td>
</tr>
<tr>
<td></td>
<td>11+ years</td>
<td>20.3%</td>
<td>10.2%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Av. income*</td>
<td>71.88</td>
<td>35.16</td>
<td>54.99</td>
<td>30.05</td>
</tr>
<tr>
<td>Av. wealth index</td>
<td>7.30</td>
<td>3.08</td>
<td>3.77</td>
<td>2.09</td>
</tr>
</tbody>
</table>

*In thousands of local currency (CFA); 1,000 CFA ∼ $2 US

Entrepreneurs are also wealthier than subsisters and there is an important gap between mean entrepreneurial earnings and subsistence income. On average, the former earn at least twice as much as the latter. Although the comparison of means are informative they may hide interesting features about the differences in household wealth and occupation-specific earnings. Thus, following the distributional approach, Figure 3 provides densities of earnings and initial wealth by occupation. The earning distribution of subsistence households is highly skewed to the left with the highest proportion of the group showing earnings below the sample mean. A similar pattern is observed for the distribution of initial wealth of this group of households. On the other hand, entrepreneurial households own a substantially higher share of the overall household wealth and income. Their earnings show a higher spread as well as a higher right-skewness and
There are very few options available through which entrepreneurial activities are funded. Approximately 90% of the total initial business investment (initial financing of the activity) come from personal savings, gifts from family and relatives, or inheritance. Loans represent the remaining 10% and come either from commercial banks or microfinance institutions, from family or friends, and from ROSCAs (rotating saving and credit associations). This fraction is even smaller for the Metro and Urban samples and represent 7% and 9%, respectively, against 16% for the Rural sample. Barriers hindering informal entrepreneurs from getting adequate access to credit are numerous. The main reasons reported by the respondents were high interest rates imposed by financial institutions and high collateral requirements. Both these are consistent with the presence of financial market imperfections (asymmetric information, limited enforcement, etc.). Among the respondents, 40% acknowledged the need to get better access to credit in order to start new business ventures or expand their existing businesses.

## 3.2 Reduced form results

Before estimating the structural parameters of the model, it is useful to perform some basic tests of the model implications. In particular, I test whether the data confirm that the probability of starting a business is increasing with observable skills and initial wealth, whether observable skills and initial wealth are positively correlated with entrepreneurial earnings and whether the correlation with wealth fades away for people at the top of the initial wealth distribution. I apply reduced forms techniques that make use of a richer set of variables than the one appearing in the theoretical model. I estimate a Probit model expressing the probability of becoming an entrepreneur as a function of initial wealth, education, experience, parent occupation, and other variables including demographic and geographic characteristics. I also use Quantile regression to evaluate the relationship between initial wealth and education and entrepreneurial earnings after controlling for other variables.
3.2.1 Probit estimates

One implication of the theoretical model is that a positive correlation between initial wealth and the probability of starting a business would imply that initial wealth determines the amount of capital required for business start-ups thus providing evidence of financial constraints. Table 2 reports estimates of the relationship between the probability of entering entrepreneurship and initial household wealth.

Table 2: Probability of Entering Entrepreneurship: Reduced Form Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole</th>
<th>Rural</th>
<th>Urban</th>
<th>Metro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Wealth</td>
<td>0.0449***</td>
<td>0.0239</td>
<td>0.0544***</td>
<td>0.0498***</td>
</tr>
<tr>
<td></td>
<td>(0.0092)</td>
<td>(0.0181)</td>
<td>(0.0110)</td>
<td>(0.0143)</td>
</tr>
<tr>
<td>Education</td>
<td>0.0335 ***</td>
<td>0.0477***</td>
<td>0.0271***</td>
<td>0.0284**</td>
</tr>
<tr>
<td></td>
<td>(0.0077)</td>
<td>(0.0155)</td>
<td>(0.0090)</td>
<td>(0.0128)</td>
</tr>
<tr>
<td>Parent</td>
<td>0.3005*</td>
<td>0.4171</td>
<td>0.2589</td>
<td>0.3451*</td>
</tr>
<tr>
<td></td>
<td>(0.1664)</td>
<td>(0.3484)</td>
<td>(0.1896)</td>
<td>(0.2057)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.0037</td>
<td>-0.0006</td>
<td>0.0056</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.0077)</td>
<td>(0.0051)</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>Age</td>
<td>0.057044 ***</td>
<td>0.0353*</td>
<td>0.0699***</td>
<td>0.0619***</td>
</tr>
<tr>
<td></td>
<td>(0.0125)</td>
<td>(0.0194)</td>
<td>(0.0159)</td>
<td>(0.0224)</td>
</tr>
<tr>
<td>Age²/100</td>
<td>-0.0619***</td>
<td>-0.0328</td>
<td>-0.079200***</td>
<td>-0.0738***</td>
</tr>
<tr>
<td></td>
<td>(0.0152)</td>
<td>(0.0226)</td>
<td>(0.0194)</td>
<td>(0.0273)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.3524***</td>
<td>-0.3867***</td>
<td>-0.3455***</td>
<td>-0.3445***</td>
</tr>
<tr>
<td></td>
<td>(0.0565)</td>
<td>(0.1092)</td>
<td>(0.0663)</td>
<td>(0.0888)</td>
</tr>
<tr>
<td>Married</td>
<td>0.1968**</td>
<td>0.1633</td>
<td>0.2078***</td>
<td>0.1894**</td>
</tr>
<tr>
<td></td>
<td>(0.0610)</td>
<td>(0.1190)</td>
<td>(0.0713)</td>
<td>(0.0951)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.0952</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
</tr>
<tr>
<td></td>
<td>(0.0638)</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
</tr>
<tr>
<td>Handicapped</td>
<td>-0.0659</td>
<td>-0.1054</td>
<td>-0.0489</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>(0.1567)</td>
<td>(0.2967)</td>
<td>(0.1852)</td>
<td>(0.2374)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.8729***</td>
<td>-2.4738***</td>
<td>-3.0100***</td>
<td>-2.7589***</td>
</tr>
<tr>
<td></td>
<td>(0.2368)</td>
<td>(0.3823)</td>
<td>(0.2979)</td>
<td>(0.4269)</td>
</tr>
</tbody>
</table>

| Number of Obs. | 4337 | 1368 | 2969 | 1494 |
| AIC:           | 2606.4 | 684.97 | 1932.9 | 1109.3 |

Standard errors in parenthesis  Significance codes: 1% ***; 5% ** 10%*

The estimate of the coefficient on wealth shows that household initial wealth is positively associated with the probability of starting a business and this association is statistically significant at the 1% level for the whole sample. This means that wealthier households can start a business with capital level close to the efficient one and hence realize a greater
profit than poorer households. Household initial wealth is however insignificant for the rural region. One possible explanation could be that while I am using the same wealth index for both rural and urban regions, determinants of wealth in the former might differ from the latter. The probit estimates also show positive interaction between education, age and marriage on the probability of starting a business for the whole sample and across geographic stratifications. However, age increases the probability of starting a business but at a negative rate as evidenced by the negative coefficient on age-squared. In other words, while age is positively associated with becoming an entrepreneur being too old is a countervailing disadvantage. It also appears that women are less likely to become entrepreneurs compared to men in the whole sample as well as in all geographic stratifications. The influence of parents’ occupation is only significant for the whole sample while the fact of being married does not affect the probability of starting a business in the rural regions.

3.2.2 Quantile regressions

Another implication of the model is that since wealthier people start their businesses with more efficient capital, the correlation between initial wealth and entrepreneurial earnings is positive. To investigate this issue, I run a quantile regression of log entrepreneurial earnings on log initial wealth after controlling for education, experience, and several demographic characteristics. Given the high heterogeneity of informal enterprises earnings (see Figure 3), the quantile regression is an appropriate tool that captures, better than the ordinary least squares regression, the specificities of the data as it permits to get estimates at different percentiles of the distribution.

Table 3 presents quantile regression estimates for the median, the first and the third quarters of the distribution of earnings. The results show the positive and statistically significant correlation of my measure of initial wealth and earnings at all quarters of the distribution of earnings. Earnings are also strongly correlated with education. A similar result is obtained from OLS regression whose overall estimates are similar to those of the median regression. However, some important features that are not captured by the OLS regression but are by the quantile regression should be emphasized. While experience has no effect on the earnings of the remaining distribution of households, it positively influences the earnings of the poorer households (those at the top 25% percentile). This means that poorer households have to compensate their low investment by accumulating more experience in order to get a decent profit. Also, unlike the other groups, being handicapped negatively influences the earnings of richer households (those of the 75% percentile). These results confirm that highly skilled entrepreneurs earn more and financially constrained entrepreneurs will start with a suboptimal amount of capital and will therefore earn smaller profit than unconstrained entrepreneurs.
Table 3: Effect of Wealth on Entrepreneurial Earnings: Reduced Form Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Quantile Regression Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>$\tau = .25$</td>
</tr>
<tr>
<td>Log Wealth</td>
<td>0.25184***</td>
<td>0.2223***</td>
</tr>
<tr>
<td></td>
<td>(0.04716)</td>
<td>(0.0181)</td>
</tr>
<tr>
<td>Log Education</td>
<td>0.2231***</td>
<td>0.2528***</td>
</tr>
<tr>
<td></td>
<td>(0.0639)</td>
<td>(0.0570)</td>
</tr>
<tr>
<td>Log Experience</td>
<td>0.08482</td>
<td>0.1009*</td>
</tr>
<tr>
<td></td>
<td>(0.05132)</td>
<td>(0.0523)</td>
</tr>
<tr>
<td>Log Age</td>
<td>0.4557***</td>
<td>0.5131***</td>
</tr>
<tr>
<td></td>
<td>(0.1604)</td>
<td>(0.17478)</td>
</tr>
<tr>
<td>Parent</td>
<td>0.41773*</td>
<td>0.3362</td>
</tr>
<tr>
<td></td>
<td>(0.2275)</td>
<td>(0.2329)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.74921***</td>
<td>-0.6725***</td>
</tr>
<tr>
<td></td>
<td>(0.08950)</td>
<td>(0.1101)</td>
</tr>
<tr>
<td>Married</td>
<td>-0.0611</td>
<td>0.0390</td>
</tr>
<tr>
<td></td>
<td>(0.0933)</td>
<td>(0.1207)</td>
</tr>
<tr>
<td>Urban</td>
<td>-0.01392</td>
<td>0.0694</td>
</tr>
<tr>
<td></td>
<td>(0.10420)</td>
<td>(0.1239)</td>
</tr>
<tr>
<td>Handicapped</td>
<td>-0.2796</td>
<td>-0.1937</td>
</tr>
<tr>
<td></td>
<td>(0.2443)</td>
<td>(0.1966)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.7077***</td>
<td>0.7728</td>
</tr>
<tr>
<td></td>
<td>(0.5603)</td>
<td>(0.6219)</td>
</tr>
</tbody>
</table>

Number of Obs. 424 424 424 424

Standard errors in parenthesis Significance codes: 1% ***; 5% **; 10%*

4 Structural Estimation and Nonparametric Evidence

In this section I derive the likelihood function of the model and use it to produce structural parameter estimates. I also perform nonparametric regression among the key variables of the model to verify the nature of their relationship. The goal is to use these structural estimates to run some counterfactual policy simulations later on.

4.1 Log-likelihood function of the model

Entrepreneurial talent $\theta$ is not observable by the econometrician. I therefore follow the literature and assume that it follows a parametric distribution (see Paulson et al.2006). Formally, I assume that ability $\theta$ is lognormally distributed and is ex-ante correlated with initial wealth and formal education. I also extend the literature by allowing ability to be also correlated with informal entrepreneurial training which can be captured by
parents’ occupation. This assumption is motivated by a number of empirical studies that found that the entrepreneurial ability of individuals increases if their parents were also entrepreneurs, partly because children may receive informal business skills from their parents (see, e.g. Lentz & Laband 1990, Parker & van Praag 2006). One can therefore write

\[ \ln \theta = \delta_0 + \delta_1 \ln z + \delta_2 \ln(1 + S) + \delta_3 P + \epsilon \]  

where \( z \) is the initial wealth, \( S \) are the years of education, and \( P \) is a dummy variable taking 1 if at least one parent was an entrepreneur and 0 otherwise. The error term \( \epsilon \) is assumed to be normally distributed with mean 0 and variance \( \sigma^2 \).

Recall from Section 2 that \( \pi(z, \theta) \) is the comparative profit from choosing entrepreneurship as opposed to subsistence activity. The allocation of households in entrepreneurship (\( E = 1 \)) and subsistence (\( E = 0 \)) can then be modeled as:

\[ E = \mathbb{I}\{\pi(z, \theta) \geq 0\} \]  

where \( \mathbb{I}\{\cdot\} \) is the indicator function that equals 1 if its argument is true and 0 otherwise. We then have

\[ \Pr\{E = 1\} = \Pr\{\pi(z, \theta) \geq 0\} = \Pr\{\pi(z, \theta) \geq 0 | z \geq \bar{z}\} \Pr\{z \geq \bar{z}\} + \Pr\{\pi(z, \theta) \geq 0 | z < \bar{z}\} \Pr\{z < \bar{z}\}, \]

where the latter display follows from the law of total probability. Substituting \( \bar{z} \) and \( \pi(z, \theta) \) by their expressions given in Formula (5) and Formula (7), respectively, we get

\[ \Pr\{E = 1\} = \Pr\left\{ (1 - \alpha)\theta \frac{1}{r} \left( \frac{\alpha}{r} \right)^{\frac{1}{1-\alpha}} \geq \mu \right\} \Pr\left\{ z \geq \frac{1}{\lambda} \left( \frac{\theta \alpha}{r} \right)^{\frac{1}{1-\alpha}} \right\} + \Pr\left\{ \theta (\lambda z)^{\alpha} \geq \lambda rz + \mu \right\} \Pr\left\{ z < \frac{1}{\lambda} \left( \frac{\theta \alpha}{r} \right)^{\frac{1}{1-\alpha}} \right\} \]

Taking the logs in the inequalities in the above terms yields

\[ \Pr\{E = 1\} = \Pr\left\{ \ln \theta > \ln \left( \frac{\mu}{1 - \alpha} \right)^{\frac{1}{1-\alpha}} \left( \frac{r}{\alpha} \right)^{\alpha} \right\} \Pr\left\{ \ln \theta < \ln \frac{r}{\alpha} (z \lambda)^{1-\alpha} \right\} + \Pr\left\{ \ln \theta > \ln (\mu + \lambda rz) (\lambda z)^{\alpha} \right\} \Pr\left\{ \ln \theta < \ln \frac{r}{\alpha} (z \lambda)^{1-\alpha} \right\} \]

If I now plug-in the distributional specification of \( \ln \theta \) given by Equation (11), I obtain the probability of becoming entrepreneur as a function of parameters and observables

\[ \Pr\{E = 1 | X\} = \Phi(h_1(\psi, X))\Phi(-h_3(\psi, X)) + \Phi(h_2(\psi, X))\Phi(h_3(\psi, X)) = H(\psi, X) \]
where $X = [1, z, S, P]'$ is the vector of covariates, $\psi = [\mu, \delta_0, \delta_1, \delta_2, \delta_3, \alpha, \sigma, \lambda, r]$ is the set of parameters, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. The functions $h_i(\cdot), i = 1, 2, 3,$ appearing in Equation (15) are defined by

$$h_1(\psi, X) = \frac{1}{\sigma} \left\{ \delta_0 - (1 - \alpha) \ln \frac{\mu}{1 - \alpha} - \alpha \ln \frac{r}{\alpha} + \delta_1 \ln(z) + \delta_2 \ln(1 + S) + \delta_3 P \right\},$$

$$h_2(\psi, X) = \frac{1}{\sigma} \left\{ \delta_0 + \alpha \ln \lambda - \ln(\mu + \lambda rz) + (\delta_1 + \alpha) \ln(z) + \delta_2 \ln(1 + S) + \delta_3 P \right\},$$

$$h_3(\psi, X) = \frac{1}{\sigma} \left\{ \delta_0 - (1 - \alpha) \ln \lambda - \ln \frac{r}{\alpha} + (\delta_1 - 1 + \alpha) \ln(z) + \delta_2 \ln(1 + S) + \delta_3 P \right\} \tag{16}$$

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

$$L_n(\psi) = \sum_{i=1}^{n} \left\{ E_i \ln H(\psi, X_i) + (1 - E_i) \ln(1 - H(\psi, X_i)) \right\} \tag{17}$$

The gross interest rate $r$ is exogenously fixed at 1.42, representing the observed average interest rate in the microfinance institutions in the country (see IDLO 2011). The maximum likelihood estimation is therefore performed over the set of parameters $\psi = [\mu, \delta_0, \delta_1, \delta_2, \delta_3, \alpha, \sigma, \lambda]$. These parameters correspond respectively to the subsistence parameter $\mu$, the constant term of the ability distribution, $\delta_0$: the interaction between wealth and ability, $\delta_1$: the interaction between education and ability, $\delta_2$: the interaction between parents occupation and ability, $\delta_3$: the productivity of capital in the production technology, $\alpha$: the standard deviation of the ability distribution, $\sigma$: and the proportion of wealth, including outside funds, that can be invested, $\lambda$.

### 4.2 Structural Results

The maximum likelihood estimates of the model are reported in Table 4. The subsistence parameter $\mu$ is estimated at 37.2 for the whole sample and is higher in Urban and Metro compared to rural areas. The constant term of log-ability expressed by $\delta_0$ is estimated at 3.36 for the whole sample and is also higher in the urban region compared to the rural region. The highest of the country is recorded at the Metro region that includes the two main cities, Douala and Yaounde. This coefficient can be regarded as the “natural” ability level, i.e. ability level that do not depend on the acquired observable characteristics. The estimated correlation between entrepreneurial ability and household initial wealth $\delta_1$ is statistically insignificant for the whole sample as well as for all the sample stratifications. This means that our measure of household initial wealth is not acting as a proxy to entrepreneurial ability. The parameter estimates for $\delta_2$ indicates that there is a correlation between years of schooling and entrepreneurial ability. This parameter is estimated to be 0.091 for the whole sample and means that each additional year of schooling can be associated with approximately 9.1% average increase in ability.

The parameter that relates parent occupation and entrepreneurial ability, $\delta_3$, is estimated to be 0.068 for the whole sample. Thus, having a parent who was an entrepreneur
Table 4: Structural Maximum Likelihood Estimates of the Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Whole</th>
<th>Rural</th>
<th>Urban</th>
<th>Metro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subsistence parameter</td>
<td>μ</td>
<td>37.206</td>
<td>33.232</td>
<td>39.068</td>
<td>43.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.7339)</td>
<td>(1.1704)</td>
<td>(0.9274)</td>
<td>(1.3505)</td>
</tr>
<tr>
<td>Log ability - constant</td>
<td>δ₀</td>
<td>3.3585</td>
<td>1.7835</td>
<td>3.5029</td>
<td>3.6320</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0118)</td>
<td>(0.0980)</td>
<td>(0.1269)</td>
<td>(0.0636)</td>
</tr>
<tr>
<td>Log ability - wealth</td>
<td>δ₁</td>
<td>0.0293</td>
<td>-0.0171</td>
<td>0.0094</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0630)</td>
<td>(0.0158)</td>
<td>(0.0757)</td>
<td>(0.0215)</td>
</tr>
<tr>
<td>Log ability - education</td>
<td>δ₂</td>
<td>0.0912</td>
<td>0.0980</td>
<td>0.0850</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0079)</td>
<td>(0.0118)</td>
<td>(0.0426)</td>
<td>(0.0076)</td>
</tr>
<tr>
<td>Log ability - parents</td>
<td>δ₃</td>
<td>0.0676</td>
<td>0.3673</td>
<td>0.0626</td>
<td>0.0725</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0148)</td>
<td>(0.0234)</td>
<td>(0.0258)</td>
<td>(0.0370)</td>
</tr>
<tr>
<td>Capital share</td>
<td>α</td>
<td>0.4989</td>
<td>0.5292</td>
<td>0.4172</td>
<td>0.3961</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.059)</td>
<td>(0.0228)</td>
<td>(0.0735)</td>
<td>(0.0193)</td>
</tr>
<tr>
<td>St. Dev for ability</td>
<td>σ</td>
<td>0.2950</td>
<td>0.1938</td>
<td>0.3692</td>
<td>0.3715</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0380)</td>
<td>(0.0330)</td>
<td>(0.0390)</td>
<td>(0.0130)</td>
</tr>
<tr>
<td>Capital constraint</td>
<td>λ</td>
<td>12.583</td>
<td>39.199</td>
<td>13.574</td>
<td>8.1375</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.3410)</td>
<td>(0.5023)</td>
<td>(0.4164)</td>
<td>(0.6964)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td></td>
<td>-1338.98</td>
<td>-344.812</td>
<td>-1006.53</td>
<td>-570.394</td>
</tr>
<tr>
<td>Number of Obs.</td>
<td></td>
<td>4337</td>
<td>1368</td>
<td>2969</td>
<td>1494</td>
</tr>
</tbody>
</table>

Asymptotic standard errors in parenthesis

Increases the entrepreneurial ability by 6.8%. In addition, this parameter is positive and significant across strata and is higher in the Rural region than elsewhere. This result shows that the occupation of parents influence individuals' entrepreneurial ability. This finding is similar to that of Lentz & Laband (1990) who use the National Federation and Independent Businesses (NFIB) data to show that half of self-employed proprietors are second-generation proprietors. The point estimate of 0.49 of the production parameter, α, means that a 10% increase in the capital devoted to a business leads to about 4.9% increase in income in the whole sample. The value of this parameter is even higher in the rural region. This suggests that in the rural regions, households operate at a small scale with high marginal returns. Returns to capital therefore appear to be high in the informal sector compared to those obtain for the formal sector. This result is consistent with those of many studies performed in Africa (Kremer et al. 2010; Udry & Anagol 2006; Schundeln 2004). High returns to capital can then be regarded as indicative of the unexploited potential of informal entrepreneurs. The coefficient σ, is estimated at about 0.3 for the whole sample and varies across geographical stratification. In particular, it is
considerably lower in the rural areas as there is likely to be a less diverse variety of skills.

The degree of financial constraints is measured by $\lambda$ and estimated at 12.6 for the whole sample, 39.2 for the rural sample and 13.6 for the urban sample. This implies that households of the rural areas are less constrained than those of the urban areas. The most constrained entrepreneurs are those who operate in the Metro areas, where the parameter is estimated at 8.1. This result can be tied to the probit estimates where the correlation between wealth and entrepreneurship is insignificant in the rural region (see Table 2). A plausible explanation may be found in the determinants of wealth across regions or in the role of social networks and family ties in the formation of investment capital. Since such ties are much stronger in the rural regions than in the urban regions it is likely that rural households benefit from alternative sources of capital from family, relatives or friends that their urban and metro counterparts are somewhat excluded from.

Some of my results can be compared with those that Paulson et al. (2006) found in their study of occupational choice in Thailand. Paulson et al. (2006) produced estimates of $\alpha$ that ranged from 0.23 to 0.79. The estimate of $\alpha$ for the Cameroon informal sector is within this range at 0.49. The Cameroon talent estimates are not directly comparable with the Thai estimates because the model I estimate allows talent to be correlated with parent’s occupation and Paulson et al. do not. Nevertheless, for Cameroon, the point estimate of $\delta_0$ is 3.35 whereas Paulson et al. estimate this parameter to be between 0.1 and 1.02. The correlation between talent and initial wealth, $\delta_1$, in both studies are similar and estimated at 0.29, although the estimate for Cameroon is not statistically significant. A major difference between the Cameroon findings and those of Thailand is on the estimates of $\delta_2$. The Cameroon data show a positive correlation between talent and schooling, estimated at 0.09 whereas Paulson et al. (2006) found that schooling is negatively correlated with talent at $-0.22$. Likewise, the Cameroon credit constraint parameter $\lambda$ is not directly comparable to the Thailand estimates because unlike Paulson et al. who actually use data on total household asset my measure of initial wealth is an index of household belongings. Nonetheless, they found that $\lambda$ ranges from 10.7 to 21 and my estimate of this parameter falls within this range at 12.6.

4.3 Nonparametric evidence

Given that initial household wealth and education appear to be key factors influencing entrepreneurial choice in the reduced form and structural results described above, I use nonparametric techniques to further test this evidence. I run a nonparametric regression of the probability of becoming an entrepreneur as a function of wealth, and as a function of wealth and education. Unlike the parametric methods, the nonparametric estimation does not impose any à priori functional form representation of the relationship between variables that may alter the predictions, but only presumes that some regularity conditions such as differentiability are satisfied. I use kernel regression to estimate the probability of starting a business as a function of initial wealth and years of schooling
As shown in the figure, there is a positive overall relationship between initial household wealth and entrepreneurship. The point estimates show a strictly increasing and very precise pattern of the relationship at a lower level of initial household wealth. For higher initial wealth levels however, the estimates display a somewhat flat relationship between wealth and the probability of entrepreneurship, which is consistent with the theoretical model predictions as depicted in the stylized figure given by Figure 1. However, the standard errors bands at this latter pattern are extremely large so that the observed relationship between higher levels of wealth and entrepreneurship is in fact imprecise. This imprecision might come from the fact that the data contains only scanty observations at higher initial wealth levels as evidenced by Figure 3. The estimates also exhibit a strictly positive nonparametric relation between years of schooling and the probability to start a business at all levels of the education distribution. These nonparametric results are consistent with those previously obtained from the reduced form and structural approaches.

5 Specification analysis and robustness checks

Since the estimated empirical model is just a simplification of the unknown true model, I perform a specification analysis to test the validity of the maximum likelihood estimates as well as various robustness checks in order to get a sense of how reliable they are for a counterfactual policy experiment. For the former, I use the Lavergne & Nguimkeu (2011)
model specification test statistic and for the latter I consider various specifications for the log entrepreneurial ability and also explore alternative definitions of entrepreneur.

5.1 Model specification test

Recall that the econometric model I am estimating is given by

\[ Pr\{E_i = 1|X_i\} = H(\psi, X_i) \] (18)

where \( H(\psi, X_i) \) is the probability of being an entrepreneur generated by the theoretical model of Section 3 for a household \( i \) with observable characteristics \( X_i \), given in Equation (15). Maximization of the sample log-likelihood function of this model have produced a maximum likelihood estimate \( \hat{\psi}_{ML} \) which is consistent with the true parameter value \( \psi \), and is asymptotically efficient, so long the model is correctly specified. If the model is misspecified, this estimator is inconsistent for \( \psi \) and the estimation results are invalid because the estimated model is too far from the data generating process. To test the model, one can recast the problem in a conditional moment restriction (CMR) framework. To do this, observe that since the random variables \( E_i \) takes only two possible values 0 and 1, the model can be rewritten as

\[ E\left[ E_i - H(\psi, X_i) | X_i \right] = 0 \] (19)

For a sample of independent observations \( \{(E_i, X_i), i = 1, \ldots, n\} \), the conditions that guarantee the identification and consistency of the MLE estimation of \( \psi \) are sufficient to estimate \( \psi \) from the CMR model (19) using smooth minimum distance (SMD) estimation method (see Lavergne & Patilea 2009).

Denote \( w_i = (E_i, X_i) \) and \( m(w_i, \psi) = E_i - H(\psi, X_i) \); The SMD estimator \( \hat{\psi}_h \) solves the minimization problem

\[ \min_{\psi} M_{n,h}(\psi) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} m'(w_i, \psi) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) m(w_j, \psi) K_{ij} \] (20)

with \( K_{ij}^h = h^{-q} K\left( \frac{X_i - X_j}{h} \right) \), where \( K(\cdot) \) is the multivariate gaussian kernel, \( h \) is the bandwidth parameter, and \( q \) corresponds to the dimension of \( X \), excluding the constant. The weighting factor \( W_n(.) \) is defined by

\[ W_n(x) = \frac{1}{n} \sum_{k=1}^{n} m(w_k, \hat{\psi}_{ML}) m'(w_k, \hat{\psi}_{ML}) b_n^{-q} K((x - X_k)/b_n) \]

with \( b_n \) taken as \( n^{-1/5} \). Solving (20) with two different bandwidths \( b = 1 \) (a fixed bandwidth) and \( h_n = cn^{-1/5} \) (a bandwidth that depends on the size \( n \) of the sample), where \( c \) is an arbitrary positive constant, yields two SMD estimators \( \hat{\psi}_b \) and \( \hat{\psi}_{h_n} \) which are respectively consistent and semiparametrically efficient when the model is correctly specified, and both asymptotically normal.
As proposed by Lavergne & Nguimkeu (2011), a Hausman-type specification test statistic for Model (19) can be defined by

$$\hat{T}_n = n(\hat{\psi}_b - \hat{\psi}_h)\hat{\Sigma}_n^{-1}(\hat{\psi}_b - \hat{\psi}_h)$$  \hspace{1cm} (21)

The matrix $\hat{\Sigma}_n$ is a suitable standardized matrix whose expression is given in Appendix 8.4. The test statistic $\hat{T}_n$ is asymptotically $\chi^2_p$ (where $p$ is the size of the parameter vector) distributed under the null of correct specification (see Lavergne & Nguimkeu 2011, Theorem 1), and takes significantly large values under the alternative of misspecification. Using numerical optimization algorithms, I computed the values of the specification test statistic using the data, for different values of $c$, namely, $c = 0.5; 1; 1.5$. For these values, $\hat{T}_n$ equals respectively to 15.102, 13.432, and 15.713, thus failing to reject the model at a 5% significance level.

The results of the bandwidth Hausman test show that the model correctly fits the data at a 5% statistical significance. However, this result should be interpreted with caution. In fact, given the evident imprecision of the estimates of the effects of initial wealth on entrepreneurship at high levels of initial wealth, as shown in the nonparametric estimation above (see Figure 4), it is hard to know whether the lack of rejection is simply a power issue. To further examine this possibility the power of the test is further assessed in the next paragraphs.

5.2 Allowing for different specifications of log ability

The power of the specification test can be assessed by examining its sensitivity to other specifications for the log ability parametrization. I consider two alternative specifications: one where the talent does not depend on any covariate, i.e. $\ln \theta = \delta_0 + \epsilon$, as in Evans & Jovanovic (1989), and one where the talent depends on $z$ and $S$ (as in Paulson et al. 2006) i.e. $\ln \theta = \delta_0 + \delta_1 z + \delta_2 \ln(1 + S) + \epsilon$. The latter specification also allows to make a direct comparison between my estimates and those of Paulson et al. (2005). The first alternative specification gives a worse overall fit than the estimated model and is strongly rejected by the test for all values of the bandwidth parameter (the p-values of the test are 0.0061, 0.0067 and 0.0065 respectively). The second specification is also rejected at a 10% level, though weakly. In fact the p-values obtained for the specification test statistic are 0.015; 0.018 and 0.021 respectively for the three bandwidths considered above. More importantly the parameter $\delta_1$ that relates wealth to ability is now statistically significant at the 10% level. This suggests that the inclusion of parent occupation in the distribution of skills not only significantly improves the overall fit of the empirical model but also takes away a significant portion of the correlation between ability and wealth. The results of this estimation can be found in Table 7 in the Appendix. Except for the parameter $\lambda$ (which is obtained differently as explained above) the other parameter estimates obtained here can be directly compared to Paulson et al.(2006) results from Thailand data. The estimate of $\alpha$ for the Cameroon informal sector is now 0.67 and is still within the range obtained by Paulson et al. The point estimate of $\delta_0$ is 3.62, way
above Paulson et al.’s range of 0.1 and 1.02. The major difference between these results and those of Thailand remains on the estimates of \( \delta_2 \) which is significantly positive and estimated at 0.098, in contrast to Paulson et al. where it was significantly negative. These results confirm that the bandwidth Hausman test statistic applied above has the power to detect misspecifications of log ability parametrization in this model.

5.3 Identification of Entrepreneurs

I return to the issue of whether the initial assignment of entrepreneurial and subsistence status to the sample households has content. I start by comparing the estimated fraction of individuals in each occupation category relative to the true data. The estimated partition of the wealth-ability space into subsisters, constrained entrepreneurs and unconstrained entrepreneurs using the maximum likelihood estimates is depicted in Figure 5. This graph is obtained by evaluating the indifference curves given by the functions \( \theta \mapsto (\lambda z)^{-\alpha}[\mu + r\lambda z]; \theta \mapsto \frac{r}{\alpha}[(\lambda z)^{1-\alpha}; \text{ and } \theta \mapsto \theta^*, \text{ that appear in the stylized version of the model summarized in Figure 2. Ignoring the actual occupation of the households, I then reallocate the households in the estimated wealth-ability space according to the criteria stated in Proposition 2. The graph shows that the structural model delivers sizes of occupations that are comparable to the observed assignment of entrepreneurial and subsistence status from the data.

The overall fractions of households by occupation predicted by the estimated model is closed to the actual fraction of households by occupation. Indeed, while the actual composition is 90.4% for subsisters and 9.6% for entrepreneurs the estimated model predicts 88.7% of subsisters and 11.3% of entrepreneurs among which about 9.5% are constrained and 1.8% are unconstrained (see percentages appearing in Figure 5).

I also consider alternative definitions of entrepreneurs used in the literature and I re-estimate the model based on these definitions. One is to consider self-employed as
entrepreneurs (Evans and Jovanovic 1989, Blanchflower and Oswald 1998) and the other is to consider business-owners as entrepreneurs (Cagetti and De Nardi 2006, Hurst and Lusardi 2004). The first provides an extremely poor model fit and is clearly not applicable in the informal economy context for the reasons already evoked. The second yields a relatively better fit, but still less than my original assignment, probably because important and widespread activities like taxi-driving are wrongly associated to subsistence in this definition. I finally evaluate my original definition through simulations of the model by constructing 75 samples of the data where entrepreneurial status is randomly assigned regardless of the actual occupation of the household. The fraction of randomly assigned entrepreneurs is fixed at the proportion of entrepreneurs in the original sample. I re-estimate the model for each of these simulated samples and the overall fit, as measured by the log likelihood, substantially declines. Figure 9 pictures the values of the overall fits obtained from various definitions of entrepreneurship including the definition I am using (Combined), self-employment (Self-Employed), business ownership (Business Owner), and the values from the 75 randomly assigned samples (Random). These results show that the definition of entrepreneur used in this study is quite robust.

5.4 Other robustness checks

It is also useful to check whether the estimated parameter $\lambda$ reflects the degree of financial constraint observed in the data. Unfortunately it is difficult to directly associate the estimated $\lambda$ to any feature of the data not only because constrained entrepreneurs are not actually observed (only entrepreneurs are), but also because some of the loans, especially those from informal sources, may not be collateralized. In practice, loan-to-collateral values are usually very low and often less than the collateral used to secure it, therefore consistent with a $\lambda$ of less than 2. At the same time there are many unsecured loans (based on reputation, trust or social ties, for example), implying an infinite $\lambda$. A combination of these two situations as it is the case in the data should yield a fairly large estimated $\lambda$ as obtained in the results (see Table 4). Nevertheless, consistency can still be checked by comparing the magnitude of this coefficient across stratifications. The estimates of $\lambda$ show that people in the Metro are more constrained than people in the general Urban areas who in turn are more constrained than people in the Rural areas, which is consistent with the data.

The above specification analysis and robustness checks show that the model is not at odds with the data. As such, it has significant implications on the nature of credit market imperfections and allocation of skills in the Cameroon informal sector. In particular, it shows that skills are an important factor for business creation and enterprise earning, yet observable individual skills (such as education) do not matter to lenders in their assessment of customers’ amount of loan or ability to repay, as does the collateral. It also consolidates the non-obvious predictions of the model that initially wealthy people

---

16 In fact, the proportion of self-employed in the data is 89%, too high to be a realistic definition for entrepreneur
without minimum skills remain subsisters, and among people with the same initial wealth
the higher skilled person is more likely to be constrained because the optimal size of his
business is larger. A natural and interesting question that arises is therefore to examine
what would be the impact on the economy if people were able to improve their business
skills through training or if potential entrepreneurs were allowed to borrow without the
collateral requirements.

6 Counterfactual Policy Experiment

In this section, I use the structural estimates of the model to evaluate the impact of some
policies on the informal sector. Since both the theory and the estimated model suggest
that entrepreneurial ability and access to credit are important factors influencing occu-
pations and earnings in the informal sector, I consider the following strategies. I examine
both the effects of a micro-lending program as well as a business training program on
the proportion of entrepreneurs, the heterogeneity among entrepreneurs captured by the
ratio of constrained versus unconstrained enterprises, and entrepreneurial and overall
average earnings. For both experiments, the starting point is the initial state of the
economy given by the estimated occupation sizes depicted in Figure 5.

6.1 Microfinance Policy

Microfinance is a credit program targeting small-scale entrepreneurial activities of the
poor who may otherwise lack access to credit because of absence of collateral. Following
Buera et al. (2011), I quantify the effect of microfinance by relaxing individual capital
rental limit as follows:

\[ 0 \leq k \leq \max\{\lambda z, z + F\} \]  

(22)

where \( F \) represents the additional capital provided by microfinance. Entrepreneurs there-
fore choose to borrow from the traditional financial institutions subject to the collateral-
ized borrowing limit \( \lambda z \) as discussed in our previews sections, or to use Microfinance to
get additional funding on top of their initial capital, \( z + F \). This topped-up funding is a
lump-sum amount that is independent of their initial wealth or entrepreneurial ability.
Practically, the microfinance policy may be initiated by a government-sponsored agency
that guarantees small loans for business start-ups, such as the US Small Business Admin-
istration (SBA) or the British Enterprise Allowance Program (BEAP). Other examples
from developing countries include the NABARD in India (a government rural develop-
ment bank that supports small-scale saving and internal lending from self help groups),
the BRI in Indonesia (Bank Rakayat Indonesia, that promotes microloans to small and
medium-scale enterprises), or the PCFC in Philippines (People’s Credit and Finance
Corporation, that is mandated by law to supplement loans to the poor through whole-
sale funds to retail microfinance institutions) (See Kaboski & Townsend 2010, Buera et
al. 2011, for a review).

In Cameroon, a microfinance program, called the National Microfinance Program Sup-
port Project (NMPSP) have been in the government’s agenda since 2003, and is described
in its poverty reduction strategy paper (IMF 2003). Although the infrastructures in place and those projected promise a reasonable take-up of the microfinance (see IMF 2003) its implementation is still ineffective. This section therefore provides an evaluation which can give some more empirical insights about the role of this policy on the economy and hopefully urge decision-makers.

The impact of this Microfinance policy on the Cameroon informal sector within the context of the model is depicted in Figure 6. The impacts are quantified for a wide range of the additional capital $F$, expressed as a multiple of the average household initial wealth index estimated at 5.20 (see Table 5 in the appendix). The left panel of Figure 6 shows the variation in the fraction of entrepreneurs (left vertical axis) and that of the ratio of constrained versus unconstrained entrepreneurs (right vertical axis). The initial value of $F$, i.e. $F = 0$, corresponds to the actual state of the economy, as produced by structural estimates in the previous sections. We observe two clear patterns. First, the fraction of entrepreneurs substantially increases with increasing values of $F$, from about 11% to reach a steady proportion close to 20%, representing almost twice the initial fraction. The maximum size of entrepreneurs is reached when the available funding is up to 3.5 times the value of the average household initial wealth, $\bar{z}$. This shows that, because of financial constraints, the country is “missing” about 9% of informal entrepreneurs, that is, about the same fraction currently available in the raw data. Second, the ratio of constrained over unconstrained entrepreneurs slightly increases with smaller increases in $F$, then steadily decreases with higher values of $F$. This is because for smaller values of $F$, many talented households will move from subsistence to entrepreneurship, but will remain constrained since the capital at their disposal may not be enough to optimally implement their project. Thus the ratio of constrained over unconstrained will increase. However, for larger amounts of $F$, many talented households will move from subsistence to directly run unconstrained firms while constrained entrepreneurs will become unconstrained as well. The ratio of constrained over unconstrained will then drastically decrease.

The right panel of Figure 6 shows that entrepreneurial earnings and total informal
sector earnings increase with $F$. We observe an increase of about 15% of entrepreneurial earnings when $F$ grows from 0 to 5, and an increase of about 30% of total earnings for the same increase in $F$. While both quantities increase, it is worth noticing that the gap between total earnings and entrepreneurial earnings consistently decreases with increasing values of $F$. This means that the earnings differentials in the informal sector get smaller as people get more access to credit. These findings are consistent with several other studies that have shown how misallocations of capital that arise from credit constraints, limited loan enforcement or institutions and policy distortions can cause important total factor productivity (TFP) gaps across firms or between countries and how appropriate policy could remedy the situation (see Restuccia and Rogerson 2008; Hsieh and Klenow 2009; Banerjee and Moll 2010).

The results obtain in this experiment are based on the assumption that the available infrastructure and cultural behavior allow for a perfect take-up rate and that people who borrow from microfinance fully repay their loans at the end of the period. Because I am estimating a partial equilibrium model, I am thus completely abstracting from the expected costs incurred by people who may default the repayment of their loan as well as the costs involved with running microfinance. These results should therefore be regarded as the upper bound of the benefits of microfinance. However, given the efforts for infrastructure deployment undertaken by the Cameroon government (see IMF 2003) and the extremely low or insignificant rates of default encountered in microfinance as found in various empirical studies (see Field and Pande 2008; McIntosh 2008) it is expected that the results presented here would not be significantly affected even if the cost of this microfinance policy was incorporated via a more general framework. The simulations therefore suggest that microfinance could be a suitable tool not only to reduce poverty and increase entrepreneurship (by increasing job creation and earnings in the informal sector) but also to lower the heterogeneity of occupations and earnings observed in the informal sector (by reducing misallocation of capital and skills).

6.2 Business Training Programs

The above experiment focuses on access to physical capital and examines the effect of relaxing the financial constraints on the entrepreneurial choice and earnings of those individuals who are sufficiently talented. However, another central implication of the model is that subsistence may arise simply because individuals lack human capital and therefore have limited entrepreneurial skills (see Proposition 1 (ii) ). It is therefore important to examine to what extent a change in human capital would affect the distribution of occupations and earnings. In this section I investigate this alternative by evaluating the impact of business training programs, which have been documented by recent studies as being an important policy option to improve the skills of entrepreneurs as well as the performance of their firms (see Calderon, Cuhna and De Georgi 2013,\footnote{All quantities represented in this simulation figure are averages.}
Although business training workshops are often initiated by firms for their workers, a nationwide business training program open to subsisters could be initiated by the Cameroon government, non-governmental organizations or international organizations (see, e.g., the International Labor Organization’s Start and Improve Your Business (SIYB) program). There are many community level associations and mutual fund networks in Cameroon that could serve as channels or platform for such initiatives (e.g., the Cameroon Cooperative Credit Union League network (CAMCCUL), the Mutuelles Communautaires de Croissance (MC2), etc.).

The impact of a business training program policy in the model can be quantified by allowing individuals’ ability level to shift as a result of an improvement in their human capital due to increased training participation.

\[ \theta' = \theta + \eta, \]

where \( \theta \) is the initial ability level and \( \eta \) is the increase in ability inculcated by the training.\(^\text{18}\) Once again, in this setting, I am abstracting from the costs incurred by financing and operating the program and more importantly the take-up rate is assumed perfect, though it would generally be lower in practice (see McKenzie and Woodruff 2012). The results should therefore be also interpreted as the maximum gain of the business training program policy. The impact of this policy on the Cameroon informal sector in the context of this model is depicted in Figure 7.

\(^{18}\)It would be interesting to calculate the number of training sessions/hours required to generate a unit increase in average talent, but this is not in the scope of this framework. Rather, I am supposing that there must exists some amount of training that would generate an increase of \( \eta \) in the average talent and I’m evaluating the impact of such increment on entrepreneurial choice and earnings.
of fractions of the initial average ability level. The left panel of Figure 7 shows that both the fraction of entrepreneurs and the ratio of constrained over unconstrained entrepreneurs increase with increasing values of \( \eta \) (left and right axes, respectively). Two main messages arise from this pattern. First, the increase in ability due to training permits individuals with a minimum amount of wealth to start an entrepreneurial business. This leads to a flow from subsistence activity to entrepreneurship, which explains the increase in the fraction of entrepreneurs. However, since this flow is purely instilled by increased skills rather than capital, most of them get stuck with constrained businesses so that the ratio of constrained entrepreneurs over unconstrained entrepreneurs increases as well. An increase of 9% of informal entrepreneurs (like in the microfinance experiment) is driven by and increase in \( \eta \) from 0 to 0.15\( \bar{\theta} \), corresponding to the situation where the total ability gain from training is 15% the initial average ability. Meanwhile, this shift implies a drastic increase of about 80% in the constrained versus unconstrained ratio of entrepreneurs, since the share of the latter is not influenced by an increase in talent.

The right panel of Figure 7 shows an increase in both entrepreneurial earnings and total earnings with respect to \( \eta \). This is because increased skills improve entrepreneurial profits which in turn drive total earnings as well. However, unlike in the scenario observed in the microfinance policy, the increase is flatter for total earnings than it is for entrepreneurial earnings and the gap between these two quantities grows with increasing levels of training. This suggests that a policy aimed to increase skills without a corresponding increase in the availability of starting capital may result to higher earnings differential within the informal sector. The increase in earnings corresponding to the implementation of a training program that shifts individuals ability by 0.15\( \bar{\theta} \) is about 18% for entrepreneurs and only about 3% for the whole informal sector.\(^{19}\)

7 Concluding Remarks

This paper provides an econometric analysis of the heterogeneity observed in the informal sector, whose main characteristics are its diversity and segmentation featuring massive proportion of subsisters who coexist with a handful of microentrepreneurs. I use a simple theoretical model of occupational choice where agents differ with their initial wealth and ability, to analyze the implications of financial constraints and skills on earnings, occupational patterns and entrepreneurship in the informal sector. The implications of the model are tested using reduced form estimates including probit and quantile regressions over multivariate controls, as well as nonparametric regressions. The results suggest that initial wealth and education are the main factors that strongly influence both the probability of becoming entrepreneur and entrepreneurial earning. These results show evidence of the existence of limited skills as well as financial constraints in the economy.

\(^{19}\)Note, however, that this effect is underestimated if subsistence earning is allowed to vary with skills as well, since business training will also increase the average subsistence earning, though at a lower extent. At any rate, the earning differentials resulting from this policy will still be larger compared to microfinance.
The theoretical model is also structurally estimated using maximum likelihood. The results suggest that there are high returns to capital of about 50% in the informal sector. The capital constraints, however, appear to be weaker in rural regions than in urban and metro areas. The adequacy of the empirical model with the data is checked through a model specification test which confirms that the model is consistent with the data at a 5% statistical significance level. The potential of the theoretical model to generate differences in earnings and occupational patterns that match the data from Cameroon is then exploited to evaluate and quantify the impact of economic policies on the informal economy using its estimated version. Counterfactual experiments show that a micro-lending program aiming at providing more access to credit to the poor, as well as a business training program aiming at enhancing human capital, substantially increases the fraction of microentrepreneurs and reduces the proportion of poor in the economy. Roughly, microfinance can reduce the size of subsistence by up to 10% while almost doubling the fraction of entrepreneurs in the informal sector. Moreover, when abstracting from possible defaults costs, microfinance can improve the total earnings capacity of the informal sector by up to 30%, and has the potential of lowering heterogeneity, by reducing misallocations of skills and capital. The effects of a business training program policy are also quantified through counterfactual simulations where the take-up rate is assumed perfect. A business training that shifts the average skilled individual’s ability by up to 15% also doubles the fraction of entrepreneurs and increases entrepreneurial earnings by 18%. However the increase in the earnings capacities of the whole informal sector due to this policy is only about 3%. This suggests that while both programs have beneficial effects on entrepreneurship and earnings, they are not enough on their own. Combining both policies should produce better improvement of the economy.

To the best of my knowledge, the structural test of the role of financial constraints and availability of skills in explaining heterogeneity and occupational patterns and providing a ground for policy interventions in the informal economy performed in this study is new to the literature. There are however some limitations as well as directions in which this research can be improved and extended. First, because the formal sector is absent in this study, the model does not allow to derive a general equilibrium through which the phenomenon of informal sector can be quantified over the rest of the economy. For example, although it is reasonable to believe that the increase of small-scale enterprises due to microfinance or business training, as implied by the model, would have extensive beneficial consequences on the demand for capital goods for production use in the formal industrialized sector, our framework does not allow to quantify such benefits. Second, because the model is static and the data is cross-sectional, the analysis fails to capture the possibly dynamic sorting of households across sectors as a consequence of dynamic productivity shocks, which could potentially cloud some of the results. These considerations are left for further research.
8 Appendix

8.1 Additional tables and figures

Figure 8: Importance of the Informal Sector in Developing Countries
Informal employment as % of non-agricultural employment in selected regions and countries, various years (1995-2000)

Table 5: Household Characteristics in the Cameroon informal sector

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whole</td>
</tr>
<tr>
<td>Number of obs.</td>
<td>4337</td>
</tr>
<tr>
<td>% of sample</td>
<td>100%</td>
</tr>
<tr>
<td>% of women</td>
<td>54.3%</td>
</tr>
<tr>
<td>Av. household size</td>
<td>6.0</td>
</tr>
<tr>
<td>Av. age of head</td>
<td>35.4</td>
</tr>
<tr>
<td>Years of schooling</td>
<td></td>
</tr>
<tr>
<td>0-5 years</td>
<td>43.7%</td>
</tr>
<tr>
<td>6-10 years</td>
<td>45.2%</td>
</tr>
<tr>
<td>11+years</td>
<td>11.1%</td>
</tr>
<tr>
<td>Av. income*</td>
<td>58.61</td>
</tr>
<tr>
<td>Av. wealth index</td>
<td>5.205</td>
</tr>
</tbody>
</table>

*In thousands of local currency (CFA); 1,000CFA ~ $2US  Source: Own calculations

8.2 Aggregate index for Household initial wealth

A Principal Component Analysis is used to create the aggregate wealth index used as household initial wealth in the empirical analysis. The idea of the PCA is as follows. We
Table 6: The cost of regulation: Requirement to start a formal business

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of Procedures</th>
<th>Duration (Days)</th>
<th>Cost as Percentage of GNI/capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD</td>
<td>6</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>South Asia</td>
<td>9</td>
<td>46</td>
<td>45.4</td>
</tr>
<tr>
<td>East Asia and the Pacific</td>
<td>8</td>
<td>51</td>
<td>47.1</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>10</td>
<td>39</td>
<td>51.2</td>
</tr>
<tr>
<td>Latin America and the Caribbean</td>
<td>11</td>
<td>70</td>
<td>60.4</td>
</tr>
<tr>
<td>Subsahara Africa</td>
<td>11</td>
<td>63</td>
<td>225.2</td>
</tr>
</tbody>
</table>

Source: UN-Habitat (2006)

Table 7: Tests of alternative log ability specifications

<table>
<thead>
<tr>
<th>Specification of log talent</th>
<th>Parameter estimates</th>
<th>Bandwidth Tests (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta_0$</td>
<td>$\delta_1$</td>
</tr>
<tr>
<td>E &amp; J</td>
<td>6.530</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>—</td>
</tr>
<tr>
<td>P. et al.</td>
<td>3.621</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.051)</td>
</tr>
</tbody>
</table>

Figure 9: Model overall fit from various definitions of entrepreneurship

have data on $p$ variables for $n$ individuals. The $i$th principal component (or principal factor) is a linear combination of the $p$ variables $X_1, X_2, \ldots, X_p$,

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{ip}X_p$$

that has the largest possible variance, subject to the condition that $a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2 = 1$. The principal factor can be extracted from this method through the following steps:
1. Define the matrix $X$ (whose columns are the $X_i$, $i = 1, \ldots, p$).
2. Subtract the mean of each column of $X$.
3. Calculate the sample covariance matrix $\Sigma$ of $X$.
4. Find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and the corresponding eigenvectors $a_1, a_2, \ldots, a_p$.
The eigenvectors should be scaled (by their euclidean norm) so that they are unit vectors. The coefficients of the $i$th principal component are then given by $a_i$ while $\lambda_i$ is its variance.
5. Assuming that the eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, choose the eigenvector that is associated with the highest eigenvalue, that is $a_1$. The principal component then corresponds to the eigenvector $a_1$,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

In particular, $\text{Var}[Z_1] = \lambda_1$ which is the highest possible variance.

To construct this single index, I use some household items in the questionnaire that relate to the number of durable goods in the household: number of vehicles, number of TVs, number of radios, number of DVD/Video-CD, number of fridges, number of freezers, number of gas cookers, number of fans, number of sewing machines, a dummy variable that is equal to 1 if the household has air conditioning, number of mobile phones, number of computers, number of electric irons, number of houses owned by the household. We consider only items that were acquired by the households prior their entry to their current activity. The resulting index computed from the data explains 31% of the variance in households durables.

8.3 Proof of Proposition 1
(i) The function $\pi(\cdot, \theta)$ is continuous on $\mathbb{R}_+$ and differentiable on $\mathbb{R}_+^*$ and we have:

$$\frac{\partial \pi(z, \theta)}{\partial z} = \begin{cases} 
0 & \text{if } z \geq \bar{z} \\
\alpha \theta \lambda^\alpha z^{\alpha-1} - \lambda r & \text{if } 0 < z < \bar{z}
\end{cases} \quad (23)$$

Recall that $\bar{z} = \frac{1}{r} \left( \frac{\theta \alpha}{r} \right)^{1/(1-\alpha)}$. We can then rewrite (23) as :

$$\frac{\partial \pi(z, \theta)}{\partial z} = \begin{cases} 
0 & \text{for } z \geq \bar{z} \\
\lambda r \left( \frac{z^{1-\alpha}}{\bar{z}^{1-\alpha}} \right) - 1 & \text{for } 0 < z < \bar{z}
\end{cases} \quad (24)$$

It then follows that $\frac{\partial \pi(z, \theta)}{\partial z} \geq 0$, for all $z$.
(ii) From part (i) above, we know that the function $\pi(\cdot, \theta)$ is increasing and has its
maximum value at \( \bar{z} \). We can write

\[
\pi(\bar{z}, \theta) = \theta (\lambda \bar{z})^\alpha - \lambda \bar{z} - \mu = (1 - \alpha) \theta^{1/(1-\alpha)} \left( \frac{\alpha}{r} \right)^\alpha/(1-\alpha) - \mu = (1 - \alpha) \left( \frac{\alpha}{r} \right)^\alpha/(1-\alpha) \left[ \theta^{1/(1-\alpha)} - \theta^{1/(1-\alpha)} \right]
\]

For \( \theta \leq \theta^* \), we have \( \pi(\bar{z}, \theta) \leq 0 \), and hence \( \pi(z, \theta) \leq 0 \) for all \( z \).

(iii) Suppose \( \theta > \theta^* \), then from Expression (25) above, we see that \( \pi(\bar{z}, \theta) > 0 \). On the other hand, we have \( \pi(0, \theta) = -\mu < 0 \). By the Intermediate Value Theorem, there exists \( \bar{z} \in (0, \bar{z}) \) such that \( \pi(\bar{z}, \theta) = 0 \). By the monotonicity of \( \pi(\cdot, \theta) \) obtained in part (i), it follows that \( \pi(z, \theta) < 0 \) \( \forall z < \bar{z} \) and \( \pi(z, \theta) > 0 \) \( \forall z > \bar{z} \). ■

8.4 Standardizing matrix in the Lavergne-Nguimkeu test statistic

The expression of \( \hat{\Sigma}_n \) used in the test statistic given in Equation 21 can be obtained as

\[
\hat{\Sigma}_n = \hat{V}_b^{-1} \Delta \hat{b} \hat{V}_b^{-1} - \hat{V}_0^{-1}
\]

where the respective matrices estimators \( \hat{V}_b, \Delta \hat{b}, \) and \( \hat{V}_0 \) are given by

\[
\frac{1}{n(n-1)} \sum_{i \neq j} \nabla^m(w_i, \hat{\psi}_b) \hat{W}_n^{-1/2}(X_i) \hat{W}_n^{-1/2}(X_j) \nabla^m(w_j, \hat{\psi}_n) b^{-q} K \left( \frac{X_i - X_j}{b} \right),
\]

\[
\frac{1}{n} \sum_i \nabla^m(w_i, \hat{\psi}_n) \hat{W}_n^{-1}(X_i) f_n(X_i) \nabla^m(w_i, \hat{\psi}_n)
\]

\[
\frac{1}{n(n-1)(n-2)} \sum_{i \neq j, k \neq k} \nabla^m(w_i, \hat{\psi}_b) \hat{W}_n^{-1/2}(X_i) \hat{W}_n^{-1/2}(X_k) \nabla^m(w_k, \hat{\psi}_b) f_n(X_j)
\]

\[
\frac{1}{n} \sum_{i \neq j} h^{-q} K((X_i - X_j)/h) \] is the leave-one-out kernel estimator of \( f(X_i) \)

8.5 Endogenizing financial constraints

I show here how the proportionality factor \( \lambda \) that is exogenously imposed in the theoretical model to capture the existence of financial constraints due to imperfect enforceability of capital rental can arise endogenously. Suppose that borrowers may voluntarily default. In that case they keep their entrepreneurial income \( \theta k^\alpha \epsilon \) but lose their collateral \( z \). However, with probability \( \phi \) they can be caught, in which case the fraction of wealth \( \beta z \) is forfeited. Thus, borrowers who choose to default receive a payoff of \( \theta k^\alpha \epsilon - \phi \beta z \) while those who choose to repay receive \( \theta k^\alpha \epsilon + r(z - k) \). Since the lenders are only willing to lend to those who will choose to repay, the incentive compatibility constraint is \( \theta k^\alpha \epsilon + r(z - k) \geq \theta k^\alpha \epsilon - \phi \beta z \). This implies that lenders will only rent to households whose wealth is sufficiently high, that is, \( z \geq \frac{r}{\beta \phi - k} \). Equivalently, this means that the capital that will be invested by the household satisfies \( k \leq \left( 1 + \frac{\beta \phi}{r} \right) z \).
References


