Involuntary Entrepreneurship – Evidence from Thai Urban Data[*]

Alexander Karaivanov[†]   Tenzin Yindok
Simon Fraser University   Colgate University

February, 2017

Abstract

We structurally estimate a model of occupational choice between entrepreneurial and non-entrepreneurial alternatives. Unlike much of the existing literature, we explicitly model and distinguish between "involuntary" entrepreneurship, that is, running an own business out of necessity vs. running an own business by choice. Involuntary entrepreneurship arises for agents who prefer (would earn higher income in) the non-entrepreneurial occupation (e.g., wage work) but cannot access it, with some probability that we estimate, due to low education, qualifications or labor market frictions. We also incorporate a credit constraint and analyze its interaction with the labor market constraint. We estimate the model via GMM using the 2005 Townsend Thai urban survey. We find that approximately 17% of all households running businesses are classified as involuntary entrepreneurs. Involuntary entrepreneurs earn lower income and are more likely among low-wealth and low-schooling households. We use the estimated model to quantify and distinguish the misallocations in occupational choice and investment from the credit and labor market constraints. We also evaluate the effects of relaxing the constraints and the impact of a microfinance policy on the rate of total and involuntary entrepreneurship and on household income, on average and stratified by wealth and schooling. The results suggest large potential income gains, especially for poorer households. Relaxing the credit constraint mostly alleviates misallocations in investment, while the misallocation from involuntary entrepreneurship is only significantly reduced by addressing the labor market constraint.

1 Introduction

Ever since Smith, Knight and Schumpeter, entrepreneurship, or running one's own business, has been viewed by most economists as an important engine of innovation and economic growth. Many tax and other government policies are explicitly designed to help small businesses grow and prosper. On the other hand, self-employment is particularly widespread in developing countries – for example, the World Bank Development Indicators data show that self-employment accounts for more than 80% of total employment in the poorest countries. How can we reconcile the notion of entrepreneurship as a driver of growth and innovation with the fact that it is so prevalent in very poor countries, often with low or negative GDP growth?

As Banerjee and Duflo (2007) put it, "...it is important not to romanticize the idea of these penniless entrepreneurs" and add "...Are there really a billion barefoot entrepreneurs, as the leaders of microfinance institutions and the socially minded business gurus seem to believe? Or is it just an optical illusion, stemming from a confusion about what we call an entrepreneur?" (Banerjee and Duflo, 2011).

Obviously, the way to solve the apparent contradiction about the role of entrepreneurship in the economy is to acknowledge that entrepreneurs are not all alike. Some people start own businesses purely on their own volition, sometimes quitting a wage job to do so. Others, however, become self-employed involuntarily or out of necessity, as their only option to earn some income and survive. Clearly the potential policy implications differ for these two categories of entrepreneurs – while some may need tax rebates, others may need social insurance or marketable job skills and qualifications.

The point that entrepreneurs are not all alike is easy to make, however, it is much harder to distinguish in the data which business owners fall in which category and to quantify the resulting misallocation in the economy. Most of the existing empirical literature adopts a reduced form approach and uses an ad-hoc criterion to distinguish between the two categories of entrepreneurs. For example, one could compare individuals who left a paid job to start a business vs. all others (Block and Wagner, 2010) or those who run an own-account business vs. those who employ other people (de Mel et al., 2012). Self-identified data on involuntary entrepreneurs is rare, the exception being the Global Entrepreneurship Monitor (GEM) survey which finds that, on average, 17% of the respondents in high-income countries and about 33% in low and middle-income countries in 2005 chose the second option in the question: "Are you in this start-up/firm to take advantage of a business opportunity or because you have no better choices for work?" (Minniti et al., 2005).[1]

Along with the empirical literature, there is a large literature on occupational choice between wage work and starting a business (Banerjee and Newman, 1993; Piketty, 1997; Aghion and Bolton, 1997; Evans and Jovanovic, 1989; Lloyd-Ellis and Bernhardt, 2000; Paulson et al., 2006; Karaivanov, 2012; Buera, 2009; Nguimkeu, 2014 among others). In all these papers the key assumption is that economic agents freely choose, out of all possible options, the occupation they prefer the most. Typically this means picking the occupation that maximizes (expected) income. Many of the models allow for market imperfections which shape the agents' optimal choices by affecting the payoffs of the different occupations but all occupations are always considered and can be chosen by all individuals. This modeling assumption is hard to reconcile with the data presented earlier which suggest that some individuals would ideally choose a different occupation (e.g., wage work instead of running a business) if it were available to them.

---

[1]The reported GEM 2005 numbers for Thailand and the USA are 24% and 12% respectively.

We build and estimate with Thai urban data a structural occupational choice model that explicitly allows for the possibility that some individuals may have a restricted choice set of occupations. In particular, in our model some agents run a business due to lack of access to wage work.[2] This restricted access can be motivated either by low education, lack of qualifications, or other similar barriers to finding paid work; or as the outcome of informational, matching or other frictions in the labor market.

Specifically, we extend and nest as a special case the classic occupational choice model of Evans and Jovanovic (1989). In that model, individuals who differ in their initial wealth and 'entrepreneurial ability' choose between running a business and wage work. They can borrow up to a fixed fraction of their initial wealth to invest in the business, representing a credit market constraint. Entrepreneurship is chosen over wage-work if the net income from running a business is larger than the income from wage work. We extend this basic framework by adding a probability with which an agent with given observable characteristics does not have access to the wage labor market. This gives rise to *involuntary entrepreneurship* if, in the absence of the choice constraint, the agent would have maximized his income as a wage worker. We specify the labor market (occupational choice) constraint by a parameter governing the tightness of the constraint. In the estimation stage, this allows the data to reveal whether the labor market constraint is negligible or significant and therefore whether our extension to the basic income-maximization model matches better the observed occupational choices in the data. Additionally, our structural approach allows us to quantify the fraction of involuntary entrepreneurs in the economy and their distribution over observables such as initial wealth and years of schooling. In a robustness check we also consider an alternative specification of the labor market constraint by assuming a fixed cost of entry into the non-entrepreneurial occupation.[3]

We use data from the Townsend Thai Project initial household survey (urban area) from 2005 (NORC, 2008). The data cover six Thai provinces (Chachoengsao, Lopburi, Srisaket, Buriram, Phrae and Satun) and surveys households in municipal areas considered urban or semi-urban. The data include detailed retrospective information on the households' assets, income, businesses, lending and borrowing, as well as individual level demographic and occupation variables. In the sample, 66% of all households are classified as entrepreneurs or 'business households' based on answering "yes" to the question whether any household member has an own business. Among the business owners, about 60% are traders (e.g., vendors of prepared food) while 33% run a business involving services (tailor, laundry, restaurant, repair shop, taxi, etc.). Among the non-business households, 93% earn the majority of their gross annual income from wages. In the robustness analysis (Section 6) we also consider an alternative definition of entrepreneurship, based on the major source of income.

We estimate the model parameters structurally via the generalized method of moments (GMM), by matching observed and model-predicted occupational choices and income levels in different stratifications by household initial wealth and education. Entrepreneurial ability is modeled as a source of unobserved heterogeneity. We match eleven moments in total (seven occupational choice moments and six income moments) and estimate nine structural parameters.

Our baseline estimation results indicate that nearly 11% of all households in the sample, or 17% of all households who report running a business, are classified as involuntaryentrepreneurs. The predicted propensity

---

[2]Unemployment is ruled out as a viable choice, for example, due to lack of social safety nets. Another typical option from the literature, subsistence agriculture is not applicable to the urban environment from which our data originate.

[3]This specification can be also interpreted as isomorphic to allowing for non-pecuniary benefits of running one's own business, for example see Hamilton (2000) or Hurst and Pugsley (2011).

of involuntary entrepreneurship at the GMM estimates varies across the different households from as high as 60% to as low as 0%. Involuntary entrepreneurship is decreasing in the household's principal earner's years of schooling and in initial wealth. Almost half of the involuntary entrepreneurs are estimated to be among the households with both initial wealth and schooling below the median. We find that the credit constraint is more likely to bind for voluntary entrepreneurs (it binds for 57% of them) than for involuntary entrepreneurs (23%). The reason is that voluntary entrepreneurs have higher entrepreneurial ability on average and hence are more likely to be credit constrained for a given wealth level. Voluntary entrepreneurs are estimated to earn significantly higher yearly income on average (554 thousand Baht) compared to involuntary entrepreneurs (83 thousand Baht) and to households not running a business (195 thousand Baht).

Simulating the model at the GMM estimates, we quantify the misallocations in terms of occupational choice and capital use among business owners, evaluate the incidence of misallocations across households with different observables, and disentangle the effects of the credit and labor market constraints. Our results imply 10.8% *excess* (involuntary) entrepreneurs relative to the first best due to the labor market constraint and 1.5% *less* entrepreneurs due to the credit constraint. Entrepreneurship is higher than in the first best among households with low schooling, due to the labor constraint, and lower than the first best for households with high schooling but low wealth, due to the credit constraint. On the intensive (capital use) margin, only 48% of the total capital used in the first best is used in the presence of the constraints, of which 1.6% is used by involuntary entrepreneurs. The investment misallocations are most severe for low-wealth households. Holding wealth constant, the investment of voluntary entrepreneurs is more misallocated (constrained) relative to the first best than that of involuntary entrepreneurs.

We also study three counterfactuals using the estimated model. First, we consider the elimination of the labor market constraint – that is, the counterfactual in which each agent is always able to choose their income-maximizing occupation (e.g., as in Evans and Jovanovic, 1989). Naturally, all else equal, relaxing the labor constraint reduces the rate of entrepreneurship in the economy since only the voluntary entrepreneurs remain. The average income in the economy goes up by 1.8% when the occupational misallocations are eliminated but the income gains are unevenly spread over the income distribution with agents at the 10th income percentile receiving a 6% income gain versus only 1% income gain at the 90th income percentile. Eliminating the labor market constraint is weakly beneficial for all households by construction but has important composition effects: it lowers the average income of ex-post non-business households because of the entry of the relatively less-skilled former involuntary entrepreneurs and raises the mean income and productivity of ex-post entrepreneurs.

In a second counterfactual we relax the credit constraint by doubling the credit tightness parameter λ which determines the maximum capital level that can be borrowed and invested in a business. Like eliminating the labor constraint, this counterfactual is Pareto improving for all.[4] At our GMM estimates, we find that relaxing the credit constraint has only a minor effect on the rate of involuntary entrepreneurship among those running a business (it falls from 16.6% to 16.2%). However, relaxing the credit constraint has significant impact on incomes by enabling some entrepreneurs to invest more. Mean income goes up by almost 5%, accompanied with gains across the income distribution and largest among poorer households (9% at the 10th income percentile). Voluntary entrepreneurs gain about the same as the average agent, while involuntary entrepreneurs and

---

[4]We assume that the households are part of a 'small open economy' and the interest rate is not affected by relaxing the credit constraint locally.

non-entrepreneurs register only minor income gains.

In a final counterfactual we introduce the option for agents in the model to take a microfinance loan of up to 10% of the median gross income in the data, M (20 thousand Baht). The loan has the same interest rate as in the baseline economy, so it effectively raises the credit limit from λz to $\lambda z + M$ where z is initial wealth. At our GMM estimates, we find that this microfinance policy has a relatively small effect on the rate of involuntary entrepreneurship (it falls from 16.6% to 15.8% of all entrepreneurs) but it raises the overall rate of entrepreneurship from 65.2% to 66.3%. The effect of the microfinance policy on household income is more significant. Average income goes up by 3% but households at the bottom of the income distribution benefit more from the ability to expand their businesses or select into a higher-income occupation – the estimated income gain is 16.5% at the 10th income percentile. We also find that the policy effects are very unevenly distributed, with the largest gains (up to 75% income increase) observed for the households with both very low wealth and schooling.

Review of the literature

Much of the existing empirical work looks at 'voluntary' vs. 'involuntary' entrepreneurs by using an ad-hoc definition based on available data. For example, Block and Wagner (2010) find a 16% earning premium in Germany for individuals who start a business after voluntarily leaving their previous job, compared to those who start a business after losing their previous job. Using data from six ex-USSR countries, Earle and Sakova (2000) find that own-account workers would earn more as employees and conclude that at least some of them are choice constrained. In Sri Lanka, de Mel et al. (2010) find that along a wide range of dimensions (parental and childhood background, labor history, measures of ability and risk-attitude), the majority of own-account entrepreneurs resemble more wage-workers than larger firm owners. Schoar (2010) differentiates between entrepreneurs who start a business as a means for providing subsistence income and 'transformational' entrepreneurs who create businesses going beyond subsistence and create jobs. The author argues that the two types differ in their objectives and skills and consequently in how they respond to policies. In particular, the paper highlights the differences between the two types in the context of microfinance and its failure in creating an entrepreneurship revolution since the structure of MFI's is not well-suited for transformational entrepreneurs. The main policy recommendation is removing the bottlenecks that limit the growth of transformational entrepreneurs, such as expanding financing or relaxing entry regulations and labor market constraints.

A few papers analyze entrepreneurship in the framework of income maximization while at the same time allowing for a labor market friction, as we do here. For example, Falco and Haywood (2013) estimate the returns to observable characteristics in self-employment vs. wage work in Ghana. They assume that job queueing may exist in the wage market modeled as an entry cost possibly depending on unobservable worker characteristics. The authors focus on obtaining consistent estimates for the return to observables and unobservables in each sector and so their results are not directly comparable to ours.

Gunther and Launov (2012) model observed income as a finite mixture of incomes from a segmented labor market. Accounting for selection, they model earnings in each segment as a linear function of demographic variables.[5] Using a 1998 Ivorian household survey, they conclude that the informal sector is made up of at least two latent segments and show that 44% of informal sector workers are predicted to maximize their earnings in a

---

[5]In their sample, 52.6% of those between the age of 15 and 65 years are inactive. In contrast, we use household level business ownership and all our households are occupied in at least one income-earning activity.

different labor market segment than the one they are engaged in. This is interpreted as evidence that involuntary employment is significant in the urban labor market. Our paper differs in that, instead of using a statistical approach, we propose a structural economic model of involuntary entrepreneurship based on maximizing behavior subject to constraints. We are also able to distinguish between labor market-constrained and credit-constrained households.

Our paper also differs from two recent working papers on entrepreneurship in a structural model setting, respectively by Banerjee et al. (2015) and Donovan (2015). Banerjee et al. use data from a microfinance randomized trial in India and define two types of entrepreneurs: 'gung-ho entrepreneurs' (GE) defined as those who already owned a business before the intervention, and 'reluctant entrepreneurs' (RE), defined as those without a business prior to the intervention. Their definition thus differs from our endogenous determination of voluntary vs. involuntary entrepreneurship within the structural model. The authors estimate a model of technology choice in which REs only have access to a decreasing returns to scale technology, while GEs can also access another technology with large fixed costs but higher return. Using data on various outcome variables separately for the GEs and REs in the treatment and control neighborhoods, they find that most of the impact from the treatment is driven by the GEs who expand their businesses in contrast to REs for whom most policy effects are insignificant. Unlike here, Banerjee et al.'s focus is not on determining who and how many the involuntary entrepreneurs are (an ex-ante definition is used) but on quantifying the heterogeneity in policy outcomes.

Donovan (2015) defines 'subsistence entrepreneurs' similarly to us, as business owners who would accept a wage job if offered but, in contrast, focuses on the role of unemployment and search frictions. In his model subsistence entrepreneurship arises as a result of low unemployment benefits and financial market imperfections. He studies the impact of the resulting talent misallocation on firm size and cross-country TFP differences. The model is calibrated and assessed (but not estimated structurally) with data from Mexico, finding that subsistence business owners earn lower profit conditional on observables and are more likely to have been fired from their previous jobs.

Buera et al. (2014) study the aggregate and distributional impacts of microfinance in a dynamic model of occupational choice with financial frictions. They allow for a stochastic shock that can lead to an agent drawing zero productivity and hence 'forced' into entrepreneurship. The authors find markedly different general equilibrium results (higher interest rate after a microfinance intervention) in the presence of 'forced' entrepreneurs compared to in a benchmark model in which agents have the complete occupational choice set. Their extended model generates a large mass of poor, low-productivity entrepreneurs who earn less than the market wage and endogenously different saving rates between entrepreneurs and workers. The authors' focus is on analyzing the equilibrium effects of microfinance and not on estimating the proportion of involuntary entrepreneurs. Nevertheless, their results underline the importance of accounting for involuntary entrepreneurship when studying occupational choice and credit policies.

2 Model

2.1 Preferences, endowments and technology

Consider a large number of households (agents) who are risk-neutral and have strictly increasing preferences over expected income. The agents differ in their initial endowments of a single investment good, z where $z \geq 0$. They also differ in two productive characteristics: $x \in [0, \bar{x}]$ which can be thought of as qualifications / schooling or, more generally, 'labor market characteristics'; and $\theta \in [\theta_{\min}, \bar{\theta}]$ which will be interpreted as entrepreneurial talent or ability.

There are two occupations (technologies). The first is a business or 'entrepreneurship' technology, E which requires capital investment $k > 0$ and one agent to operate and yields output[6]

$$q^E(\theta) = \theta k^\alpha$$

where $\alpha \in (0, 1)$. There is no minimum scale or fixed costs to start up a business.

The second occupation or technology does not require capital and yields

$$q^A(x) = \mu(1 + x)^\gamma.$$

The parameter $\mu > 0$ corresponds to what a person with labor market characteristics $x = 0$ would earn while $\gamma \geq 0$ governs the sensitivity of q^A to increases in x. We interpret occupation A as a non-business occupation, that is, the alternative to entrepreneurship. It may include wage work or other similar activities, the income from which increases in x.

2.2 Credit market

As in Evans and Jovanovic (1989), hereafter EJ (1989), assume that the agents have access to a financial intermediary via which they can save or borrow at the fixed gross interest rate $r \geq 1$. The credit market is imperfect – due to a limited enforcement problem the maximum amount of capital k that an agent can invest is λz, where $\lambda > 0$ is a parameter capturing the tightness of credit constraints.[7] A very large λ corresponds to perfect credit markets while $\lambda = 0$ corresponds to a missing credit market (only saving is possible).

Agents employed in the A occupation do not need capital, so they save their initial wealth z which results in income of:

$$y^A(x, z) = \mu(1 + x)^\gamma + rz$$

Agents employed in the E occupation (entrepreneurs) either save or borrow at the rate r, depending on their desired investment k. Their income is

$$y^E(\theta, z) = \theta k^\alpha + r(z - k)$$

---

[6]We can allow output to be stochastic as in Evans and Jovanovic (1989) but because of risk neutrality all that matters for the analysis is expected output. One can therefore interpret all output and income variables in the model as expected values over stochastic technology shocks.

[7]The upper bound λz can be micro-founded by a limited enforcement friction, see for example Paulson et al. (2006).

If an agent has a sufficiently large wealth z, the credit constraint $k \leq \lambda z$ would not bind and she would be able to invest the first-best amount of capital (to be determined below). In contrast, if an agent has relatively low wealth, she will be credit-constrained and invest λz even though at $k = \lambda z$ the marginal product of capital exceeds the cost of funds r. The credit market constraint thus leads to a misallocation of capital (under-investment).

2.3 Involuntary entrepreneurship

In EJ (1989) agents always pick the occupation (E or A) which yields higher expected income. That is, absent any constraints on her choice set, an agent would choose the occupation which attains $\max\{y^E(\theta, z), y^A(x, z)\}$. Here, we depart from EJ by assuming that, depending on the agent's characteristic x (schooling, labor market skills), the agent's access to occupation A is restricted, with some probability that we will estimate. For instance, agents with lower x find it harder to find wage work; government or private sector jobs may require diplomas, qualifications, certificates, etc. We interpret this as a labor market constraint. In Section 6.3 we also consider an alternative specification of the labor market constraint in the form of a fixed cost of entering the non-business occupation A.

Specifically, let P_x be the probability with which an agent with labor market characteristics x *does not have access* to occupation A in the current period. That is, with probability P_x the agent only has access to the entrepreneurial occupation E, while with probability $1 - P_x$ she can *choose* between E or A. If occupation E is what this agent would have chosen to maximize her income, then the labor market constraint is not binding for her. However, if $y^A(x, z) > y^E(\theta, z)$ for this agent, then she will be an "involuntary" entrepreneur – someone who engages in the E occupation because no other alternatives are available.

Assume that

$$1 - P_x = \left(\frac{1 + x}{1 + \bar{x}}\right)^{\eta} \tag{1}$$

where $\bar{x}$ is the largest possible value of x and $\eta \geq 0$ is a parameter governing the tightness of the labor market constraint for different values of x. The special case $\eta = 0$ corresponds to $P_x = 0$ for all x, that is, all agents are able to choose freely between both occupations. This corresponds to the Evans and Jovanovic (1989) model in which there are no involuntary entrepreneurs. In contrast, the case $\eta < 1$ corresponds to the constraint becoming less tight quickly for relatively low x, while $\eta > 1$ corresponds to the case when the constraint is relaxed only for relatively large values of x. The economic interpretation of (1) is that agents with higher schooling or other labor market skills x are more likely to have access to both occupations in any given moment of time.

2.4 Investment and occupational choice

Remember that for an agent with ability θ and initial wealth z, income from entrepreneurship is

$$y^E(\theta, z) = \theta k^{\alpha} - r(z - k).$$

If the credit constraint $k \leq \lambda z$ is not binding, an agent with initial wealth z and ability θ would optimally invest the first-best (unconstrained) capital amount,

$$k_u(\theta) \equiv \arg\max_{k}\{\theta k^{\alpha} - rk\} = \left(\frac{\theta \alpha}{r}\right)^{\frac{1}{1-\alpha}} \tag{2}$$

Note that $k_u(\theta)$ is increasing in θ which implies that higher-ability entrepreneurs would like to invest more. The first-best investment $k_u(\theta)$ does not depend on the entrepreneur's initial wealth z. Intuitively, in the absence of credit constraints all businesses should be capitalized at the efficient level that equalizes marginal product with marginal cost regardless of the business owner's wealth.

In the presence of credit constraints, however, the first-best investment is only feasible if $k_u(\theta) = (\frac{\theta\alpha}{r})^{\frac{1}{1-\alpha}} \leq \lambda z$. Call $B(z)$ the threshold level of entrepreneurial talent θ at which $k_u(\theta) = \lambda z$, that is,

$$B(z) \equiv \frac{r}{\alpha}(\lambda z)^{1-\alpha} \tag{3}$$

For given initial wealth z, the value $B(z)$ is the maximum level of talent θ at which an agent is financially unconstrained and able to invest $k_u(\theta)$. For given z, the credit constraint is therefore more likely to bind for more talented entrepreneurs. If $\theta > B(z)$, since the marginal product of capital exceeds the marginal cost, the agent would optimally invest the maximum possible amount λz which is less than $k_u(\theta)$.

We therefore obtain,

$$y^E(\theta, z) = \begin{cases} \theta(k_u(\theta))^\alpha + r(z - k_u(\theta)) & \text{if} \quad \theta \leq B(z) \\ \theta(\lambda z)^\alpha + r(z - \lambda z) & \text{if} \quad \theta > B(z) \end{cases}$$

or equivalently,

$$y^E(\theta, z) - rz = \begin{cases} (1-\alpha)\theta^{\frac{1}{1-\alpha}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} & \text{if} \quad \theta \leq B(z) \\ \theta(\lambda z)^\alpha - \lambda rz & \text{if} \quad \theta > B(z) \end{cases}$$

Alternatively, an agent in the non-business occupation A would earn,

$$y^A(z, x) = \mu(1 + x)^\gamma + rz.$$

The following result captures the main occupational choice trade-off when there is no labor market constraint, as in Evans and Jovanovic (1989).

Proposition 1

Define the income differential between entrepreneurship and the alternative occupation as

$$\Delta(z, \theta, x) \equiv y^E(\theta, z) - y^A(z, x).$$

An agent with initial wealth z and characteristics θ and x who has access to <u>both</u> occupations E and A would optimally choose entrepreneurship, E if

$$\Delta(z, \theta, x) \geq 0 \Leftrightarrow \begin{cases} \theta \geq A(x) & \text{if } \theta \leq B(z) \\ \theta \geq C(x, z) & \text{if } \theta > B(z) \end{cases} \tag{4}$$

where $A(x) \equiv (\frac{\mu}{1-\alpha})^{1-\alpha}(1 + x)^{\gamma(1-\alpha)}(\frac{r}{\alpha})^\alpha$, $B(z) \equiv \frac{r}{\alpha}(\lambda z)^{1-\alpha}$, and $C(z, x) \equiv (\lambda z)^{-\alpha}[\mu(1 + x)^\gamma + r\lambda z]$.

Proof: see Appendix A.

2.5 The probability of entrepreneurship

We follow the literature and assume that entrepreneurial ability θ is known to the agents in the model but is unobservable to the econometrician. That is, we treat θ as source of unobserved heterogeneity in the empirical work. In contrast, initial wealth z and the labor market characteristics x are known to all. Thus, for a given distribution of θ and given z and x the model implies a probability that an agent chooses to be an entrepreneur (occupation E) or not (occupation A). In Section 4 we compute and use these predicted probabilities to estimate the structural parameters of the model based on the observed occupational status of households in the Thai urban data. In addition, for any z and x, our structural model implies a probability (fraction) of *involuntary* entrepreneurs, which is unobserved in the data.

Suppressing the arguments in the expressions $A(x)$, $B(z)$, $C(z,x)$ and $\Delta(\theta, z, x)$ to save on notation and using P to denote probabilities, Proposition 1 implies,

$$\begin{aligned} P(\Delta \geq 0) &= P(\Delta \geq 0 | \theta > B)P(\theta > B) + P(\Delta \geq 0 | \theta \leq B)P(\theta \leq B) \\ &= P(\theta \geq C | \theta > B)P(\theta > B) + P(\theta \geq A | \theta \leq B)P(\theta \leq B) \\ &= P(\theta \geq C \wedge \theta > B) + P(\theta \geq A \wedge \theta \leq B) \end{aligned} \tag{5}$$

For any given wealth z, labor market characteristics x and model parameters, the exact ordering by magnitude among $A(x)$, $B(z)$ and $C(z,x)$ is completely determined.

To compute the probabilities in (5) we need an assumption on the distribution of unobserved heterogeneity, θ (entrepreneurial talent). We follow Paulson et al. (2006) and assume,

$$\ln \theta = \delta_0 + \delta_1 \ln z + \delta_2 \ln(1 + x) + \varepsilon \tag{6}$$

$$\text{where } \varepsilon | z, x \sim N(0, \sigma)$$

The interpretation is that entrepreneurial ability may be correlated with initial wealth z and the observable labor market characteristics x (in the baseline estimation we proxy x by the years of schooling of the household's principal earner) but we also allow a random ability component (shock), ε. The distributional parameters $\delta_0, \delta_1, \delta_2$ and σ are estimated together with the model's structural parameters $\alpha, \gamma, \lambda, \mu, \eta$.

Let $\mathbf{1}_{B>A}$ denote the indicator function which equals one if $B > A$ for given (x,z) and zero otherwise. It is easy to show that, for any (x,z), the inequality $B > A$ is mathematically equivalent to the inequality $B > C$, that is, $\mathbf{1}_{B>A} \Leftrightarrow \mathbf{1}_{B>C}$. Denoting the conditional expectation of log ability by

$$\bar{\theta}(z,x) = E(\ln \theta | z, x) = \delta_0 + \delta_1 \ln z + \delta_2 \ln(1 + x),$$

using Proposition 1 and (5), we obtain the following result.

Lemma 1

For an agent with observable characteristics (z,x) who has access to <u>both</u> occupations E and A, the probability (likelihood) of choosing entrepreneurship equals,

$$P(\Delta(\theta, z, x) \geq 0) = \mathbf{1}_{B>A}(1 - \Phi(a)) + (1 - \mathbf{1}_{B>A})(1 - \Phi(c)) \tag{7}$$

Table 1: **Voluntary and involuntary entrepreneurship**

	$1_E = 0$	$1_E = 1$
$\triangle \geq 0$	N/A	voluntary entrepreneur
	$P(\Delta \geq 0, 1_E = 0) = 0$	$P(\Delta \geq 0, 1_E = 1) = P(\Delta \geq 0)$
$\triangle < 0$	Non-entrepreneur	involuntary entrepreneur
	$P(\Delta < 0, 1_E = 0) = P(\Delta < 0) - P_x P(\Delta < 0)$	$P(\Delta < 0, 1_E = 1) = P_x P(\Delta < 0)$

where $a \equiv \frac{\ln A(x) - \bar{\theta}(z,x)}{\sigma}$ *and* $c \equiv \frac{\ln C(z,x) - \bar{\theta}(z,x)}{\sigma}$.

2.6 Involuntary entrepreneurship

Denote by 1_E the indicator function for choosing entrepreneurship in the model conditional on observables x and z. Using the Law of total probability and suppressing the conditioning on x, z to simplify the notation, we have:

$$P(1_E = 1) = P(1_E = 1|\Delta \geq 0)P(\Delta \geq 0) + P(1_E = 1|\Delta < 0)P(\Delta < 0)$$

where $P(\Delta \geq 0)$ is given by (7) in Lemma 1 and $P(\Delta < 0) = 1 - P(\Delta \geq 0)$.

Observe that $P(1_E = 1|\Delta \geq 0) = 1$, since any agent who earns higher income by being entrepreneur ($\Delta \geq 0$) would choose occupation E (it is always available). Also, $P(1_E = 1|\Delta < 0) = P_x$ where P_x was defined in (1) in Section 2.3 – with probability P_x an agent with characteristics x is constrained on the labor market and hence enters occupation E even though $\Delta < 0$, that is, he has higher potential income in the unavailable occupation A. Therefore, for any given z and x, the probability (predicted rate) of entrepreneurship in the model is

$$P_E(z,x) \equiv P(1_E = 1) = P(\Delta \geq 0) + P_x P(\Delta < 0) \tag{8}$$

The overall probability of entrepreneurship conditional on z and x, $P_E(z,x)$ is the sum of two terms. The first term, $P(\Delta \geq 0)$ corresponds to the probability (rate) of entrepreneurship that would arise if all agents could choose occupation E based solely on income maximization, as typically assumed in the literature, for example EJ (1989). The second term,

$$P_I(z,x) \equiv P_x P(\Delta < 0)$$

is the additional probability/rate of entrepreneurship relative to the income-maximization model, which we interpret as the probability (rate) of *involuntary entrepreneurship*. Table 1 summarizes the analysis.

The probability P_x of having access to the non-business occupation is a function of labor market characteristics x. The parameter η determines how schooling affects the tightness of the labor market (occupational choice) constraint. The parameter γ on the other hand determines how x affects directly the non-business income of a household. Agents with low x are more likely to be involuntary entrepreneurs through the effect of η, but less likely to be involuntary entrepreneurs since their non-business incomes are also lower through γ. The overall effect of the labor market constraint thus depends on the relative sizes of these two parameters.

3 Data and Reduced Form Evidence

We use data from the Townsend Thai Project's 2005 Urban Annual Survey.[8] The main outcome variable of interest is household business ownership. We measure business ownership in the data in terms of whether a household reports that they own at least one business at the time of the survey. That is, we construct a binary variable equal to 1 if a household reports owning a business and zero otherwise. The corresponding variable in the model is $\mathbf{1}_E$. We also consider an alternative definition of business ownership in the robustness Section 6.

Initial household wealth (the variable z in the model) is measured as the total value in 2005 Thai baht of land holdings, household durables and agricultural assets owned by a household *five years prior to* the survey. The reason for the back-dating is to avoid possible simultaneity problems between occupational status and current wealth. Recall that in the model initial wealth z affects the investment potential of a household. We are therefore assuming that the level of pre-existing (year 2000) wealth measure we construct is free of reverse causality. Also, note that our model allows initial wealth z to be correlated with entrepreneurial ability θ and therefore we can capture, in a reduced form, the possibility that more talented agents may save more in anticipation of becoming business owners.

We proxy the model variable x interpreted as the education, qualifications or other characteristics determining one's potential labor market income by the years of schooling of the principal earner in the household. To identify the principal earner we use data on individual occupations and work type within the households. For business households, the principal earner is defined as the member whose occupation and worker type matches the reported business type (for households running more than one business, the principal earner is defined as the owner of the largest business in terms of assets). For non-business households, the principal earner is defined as the wage-earning member (for households with multiple wage-earners, the principal earner is the member earning the highest monthly wage income). We also consider an alternative definition of x in the robustness Section 6.

Finally, in the structural estimation we also use annual gross household earned income, defined as household income excluding remittances, government program transfers and interest income. The model counterparts are $q^E(\theta) = \theta k^\alpha$ and $q^A(x) = \mu(1+x)^\gamma$ for business and non-business households respectively.

The sample we use in the estimation is constructed as follows. We exclude all households in the top one percentile of the initial wealth distribution, all households with zero initial wealth or zero gross income, and all households for which the principal earner could not be identified.[9] Table 2 shows that 66% of the households in our final sample report running a business. Using the income data, we also see that running a business and wage work are the two most important sources of income for households. More than half of all households in our sample derive the majority of their annual gross income from running a business and nearly 42% of all households do so from wages. Only a small fraction (2.9%) derive the major part of their income from farming (rice, other crops, and livestock-raising).

---

[8] Details are available at `cier.uchicago.edu`.

[9] Because of data limitations we were not able to identify a principal earner for about 15% of all surveyed households.

Table 2 – Occupation and Source of Income

Self-reported business ownership	Number	Percent
yes	786	66.1
no	403	33.9
total	1,189	100

Major source of annual gross income	Number	Percent
business	632	53.2
wage	496	41.7
farming	34	2.9
other	27	2.2
total	1,189	100

Notes: The sample excludes the top percentile of the wealth distribution, households with zero income, and where a principal earner could not be identified.

Table 3 presents summary statistics of the key variables of the data. We see that business households have statistically significantly larger mean wealth and annual gross incomes than non-business households. The annual gross income of households that run businesses also has much larger standard deviation than that of non-business households. The principal earners in non-business households have higher years of schooling, are younger and more likely to be male, compared to the principal earners in business households. There is no statistically significant difference in household size between the two types.

Table 3 – Summary statistics

variable	business	non-business	all
wealth 5 years ago ('000 Baht), mean*	620.5	469.4	569.3
standard deviation	(814.8)	(682.3)	(775.5)
median	*335.1*	*235.1*	*305.0*
annual gross income ('000 Baht), mean*	513.6	164.7	395.3
standard deviation	(1313)	(132.5)	(1075)
median	*276.8*	*126.0*	*200.8*
years schooling of principal earner, mean*	7.3	9.8	8.1
standard deviation	(4.0)	(4.7)	(4.5)
age of principal earner, mean*	49.4	41.2	46.6
standard deviation	(11.0)	(13.1)	(12.3)
male (gender of principal earner), mean*	0.45	0.59	0.50
standard deviation	(0.50)	(0.49)	(0.50)
household size, mean	4.28	4.35	4.30
standard deviation	(1.90)	(1.83)	(1.87)
sample size	786	403	1189
sample proportion	66.1%	33.9%	100%

Notes: The sample excludes the top percentile of the wealth distribution, households with zero income and where a principal earner could not

be identified. Mean and standard deviations (in parentheses) reported for all variables, median (in italics) for monetary values. Wealth and income are in thousands of 2005 Thai baht. (*) difference-in-means test between business and non-business is significant at the 1% level.

Table 4 reports the coefficient estimates obtained by a probit regression of business ownership (a binary variable equal to one if a household reports owning a business) on initial wealth (assets five years prior to the survey), years of schooling, and additional household characteristics, as defined earlier. The results indicate that both the household initial wealth and the principal earner's schooling are correlated with the probability of business ownership in a statistically significant way. For both initial wealth and schooling the association with business ownership is positive and with a diminishing rate. Households with female or older principal earners, and with larger household size are more likely to be business owners. We view these results as a validation of our modeling assumptions that initial wealth and years of schooling matter for business ownership. We consider the effects of age and gender on the model parameters estimates in robustness checks using sample stratifications (see Section 6).

Table 4 – Probit for household business ownership

variable	coefficient estimate
initial wealth (mln Baht)	0.431***
	(0.137)
initial wealth squared	-0.070***
	(0.033)
schooling of principal earner	0.197***
	(0.044)
schooling squared	-0.015***
	(0.002)
age of principal earner	0.028***
	(0.004)
male (gender of principal earner)	-0.450***
	(0.084)
household size	0.040**
	(0.023)
provincial dummies – included	
sample size	1189

Notes: The dependent variable is an indicator for whether a household reports owning a business in 2005. Standard errors are reported in parentheses. The regression includes provincial dummies and an intercept. * p<0.10, ** p<0.05, *** p<0.01.

4 Structural Estimation

We have a sample of N households, $i = 1, ..., N$ with data on their initial wealth, z_i, years of schooling of the principal earner, x_i and occupational status E_i (with $E_i = 1$ if the household runs a business and zero otherwise), as defined in Section 3. We estimate the structural parameters (technology, credit and labor market access) as well as the distributional parameters of entrepreneurial ability θ via the generalized method of moments (GMM)

by matching a list of entrepreneurship probabilities and income moments predicted by the model to their data counterparts, for the observed x_i and z_i.

The nine estimated parameters are: α – the elasticity of business revenue with respect to investment; λ – the credit constraint tightness; γ – the elasticity of non-business income with respect to x; η – the parameter in P_x governing the labor market constraint; μ – a scaling parameter for non-business income; δ_0 – the conditional mean of log entrepreneurial talent; δ_1 and δ_2 – the elasticities of log talent with respect to initial wealth and schooling and σ – the standard deviation of log-talent. Call the vector of all estimated parameters $\phi \equiv (\alpha, \lambda, \gamma, \eta, \mu, \delta_0, \delta_1, \delta_2, \sigma)$. We calibrate the gross interest rate r to 1.06, which corresponds to the median rate of interest on household loans in our data.

4.1 GMM – matched moments and computation

The model parameters are estimated using GMM by minimizing the percentage deviation between a set of moments in the model and their respective sample analogs. Specifically, given parameters ϕ, denote the model-predicted moments by $h_j(z, x, \phi)$ for $j = 1, ..., J$ and their respective sample analogs by h_j^d. Definitions of all J moments we use are provided in Table 5 below. Define the percentage deviation of the model predicted moment from its sample analog as

$$q_j(z, x, \phi) \equiv \frac{h_j(z, x, \phi) - h_j^d}{h_j^d}, \quad j = 1, ...J$$

Construct $\mathbf{q}(z, x, \phi)$ as the $J \times 1$ vector of percentage deviations between the model-predicted moments and their sample analogs. The GMM estimates are computed by minimizing the criterion function $\mathbf{q}(z, x, \phi)'\mathbf{q}(z, x, \phi)$ over the parameters ϕ. We use an optimization routine robust to local extremes initialized at the results from an extensive grid search over the parameter space.[10]

In our baseline specification we match the eleven moments listed in Table 5 below by choice of the nine parameters ϕ. The first seven moments correspond to the probabilities (proportions) of business ownership in different sub-samples defined based on the terciles of years of schooling (x) and initial wealth (z). The model-predicted probability (proportion) of business ownership for some subset of the observed initial wealth levels $z_i \in Z$ and years of schooling $x_i \in X$ is

$$h_j(z, x, \phi) = \frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \in Z, x_i \in X\}} P(\mathbf{1}_E = 1 | z_i, x_i, \phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \in Z, x_i \in X\}}}$$

where the probability or entrepreneurship $P(\mathbf{1}_E = 1 | z_i, x_i, \phi)$ is computed using (8). These moments, for different subsets Z and X, are labeled $j = 1, ..7$ in Table 5. Their sample analogs in the data are the actual observed fractions of business owners (those with $E_i = 1$) with characteristics $z_i \in Z$ and $x_i \in X$.

The remaining four matched moments, labeled $j = 8, .., 11$ in Table 5, correspond to the average expected gross incomes of business and non-business households in the whole sample or when stratified by initial wealth

---

[10]We first perform an extensive grid search over approximately 20,000 parameter configurations. We then use the Matlab global optimization routine *particleswarm* starting with an initial population of the 20 best-fitting parameter vectors from the grid search.

and schooling. For example, the average expected gross income of business households in the model is

$$h_8(z, x, \phi) = \frac{\sum_{i=1}^{N} E(q^E|\mathbf{1}_E = 1, z_i, x_i, \phi))}{\sum_{i=1}^{N} P(\mathbf{1}_E = 1|z_i, x_i, \phi)}$$

where the expectation is taken over the entrepreneurial ability random component ε and the average is across households. The expected business and non-business gross incomes in the model, $E(q^E|\mathbf{1}_E = 1, z_i, x_i, \phi)$ and $E(q^A|\mathbf{1}_E = 0, z_i, x_i, \phi)$ for any z_i, x_i are computed in Appendix B. The sample analogs of the income moments are obtained by replacing $P(\mathbf{1}_E = 1|z_i, x_i, \phi)$ with the households' observed occupational status, E_i and replacing $E(q|\mathbf{1}_E = o, z_i, x_i, \phi)$ for $o = \{0, 1\}$ by the actual observed incomes, q_i^E and q_i^A of business or non-business households in the data (see Table 5, moments 8–11).

Table 5 – Matched moments

moment	model	sample analog
1. Average probability of entrepreneurship	$\frac{1}{N} \sum_{i=1}^{N} P(\mathbf{1}_E = 1|z_i, x_i, \phi)$	$\frac{1}{N} \sum_{i=1}^{N} E_i$
2. Probability of entrepreneurship, $x \leq x_{t1}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_{t1}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_{t1}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_{t1}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_{t1}\}}}$
3. Probability of entrepreneurship, $z \leq z_{t1}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}\}}}$
4. Probability of entrepreneurship, $x > x_{t3}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i > x_{t3}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i > x_{t3}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i > x_{t3}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i > x_{t3}\}}}$
5. Probability of entrepreneurship, $z > z_{t3}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}\}}}$
6. Prob. of entrepreneurship, $z \leq z_{t1}, x \leq x_{t1}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}, x_i \leq x_{t1}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}, x_i \leq x_{t1}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}, x_i \leq x_{t1}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_{t1}, x_i \leq x_{t1}\}}}$
7. Prob. of entrepreneurship, $z > z_{t3}, x > x_{t3}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}, x_i > x_{t3}\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}, x_i > x_{t3}\}}}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}, x_i > x_{t3}\}} E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i > z_{t3}, x_i > x_{t3}\}}}$
8. Average gross income, entrepreneurs	$\frac{\sum_{i=1}^{N} E(q^E|\mathbf{1}_E=1,z_i,x_i,\phi))}{\sum_{i=1}^{N} P(\mathbf{1}_E=1|z_i,x_i,\phi)}$	$\frac{\sum_{i=1}^{N} q_i^E E_i}{\sum_{i=1}^{N} E_i}$
9. Average gross income, non-entrepreneurs	$\frac{\sum_{i=1}^{N} E(q^A|\mathbf{1}_E=0,z_i,x_i,\phi)) P(\mathbf{1}_E=0|z_i,x_i,\phi)}{\sum_{i=1}^{N} P(\mathbf{1}_E=0|z_i,x_i,\phi)}$	$\frac{\sum_{i=1}^{N} q_i^A (1-E_i)}{\sum_{i=1}^{N} (1-E_i)}$
10. Average gross income, entrepreneurs, $z \leq z_m$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_m\}} E(q^E|\mathbf{1}_E=1,z_i,x_i,\phi)) P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_m\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_m\}} q_i^E E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{z_i \leq z_m\}} E_i}$
11. Average gross income, entrepreneurs, $x \leq x_m$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_m\}} E(q^E|\mathbf{1}_E=1,z_i,x_i,\phi)) P(\mathbf{1}_E=1|z_i,x_i,\phi)}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_m\}} P(\mathbf{1}_E=1|z_i,x_i,\phi)}$	$\frac{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_m\}} q_i^E E_i}{\sum_{i=1}^{N} \mathbf{1}_{\{x_i \leq x_m\}} E_i}$

We use GMM to structurally estimate the model. Using GMM is computationally fast and allows us to derive and use analytical expressions for both the model-predicted occupational choice and the incomes of business and non-business households (see Appendix B). In earlier drafts we also estimated the model parameters via maximum likelihood by using only occupational choice data (as in Paulson et al., 2006 or Karaivanov, 2012), however, the resulting simulated business and non-business incomes at the MLE estimates (available upon request) were an order of magnitude off from their data counterparts.[11] Since a major part of the paper focuses on evaluating the misallocations and the effects on household income/welfare from (relaxing) the credit and labor market constraints, we consider incorporating income data in the estimation as essential.

4.2 Results

Table 6 reports the GMM parameter estimates. The return to capital in entrepreneurial income, α is estimated at 0.23, implying that a 10 percent increase in capital k would lead to an approximately 2.2% percent increase in the income of unconstrained entrepreneurs, all else equal. The estimate of the credit constraint parameter λ is 0.23, which implies that for a household with initial wealth z equal to the median, the maximum business investment it can make is about 70,000 Baht.[12] As a comparison, the median business assets in the data is about 19,700 Baht which is about 6.5% of median initial wealth. The parameter γ, estimated to be 0.75, determines how schooling affects the non-business income of a household – for example, an increase in the years of schooling from 4 to 5 raises non-business income by 18%. The labor market constraint parameter η is estimated to be 0.41. At the modal years of schooling $x = 4$, this implies a 41% probability (using (1) and $\bar{x} = 17$) that an agent is constrained in her income-maximizing occupational choice. Entrepreneurial talent θ is found to be weakly positively correlated with both initial wealth and years of schooling (the estimates of δ_1 and δ_2 are positive).

Table 6 – GMM estimates

Parameter		estimate	standard error
return to capital in business income	α	0.227	0.058
credit constraint parameter	λ	0.233	0.455
return to schooling in non-business income	γ	0.747	0.075
tightness of the labor constraint	η	0.407	0.173
non-business income parameter	μ	28.5	4.7
talent – constant	δ_0	3.42	0.48
talent – elasticity w.r.t. initial wealth	δ_1	0.129	0.053
talent – elasticity w.r.t. schooling	δ_2	0.168	0.123
talent – standard deviation	σ	0.956	0.159

Notes: Standard errors are calculated from 99 bootstrap samples.

---

[11]We found it analytically intractable to derive the joint likelihood of occupational choice and income for business and non-business households in our model.

[12]In the model λ can also proxy for the liquidity or collateralizability of household wealth as defined (land, household durables and agricultural assets). An estimate less than one could thus be interpreted as households not being able to completely use their wealth to finance their businesses.

Table 7 reports several model predictions evaluated at the GMM estimates. We compute these statistics by simulating data from the model at the GMM parameter estimates which is done by drawing 100 random values from the distribution of the shock ε for each $i = 1, ...N$. We then average, first over ε for each household $i = 1, ...N$, and then over the reported stratification of households to compute the different statistics in Table 7.

The proportion of involuntary entrepreneurs (business households) from all households in our sample is 10.8%. In other words, 16.6% of all business owners in the sample are classified as involuntary entrepreneurs. The remainder, 54.4% of all households or 83.4% of all business owners are classified by the model as voluntary entrepreneurs. Approximately 51% of all entrepreneurs are estimated to be credit constrained – that is, their investment k equals λ times their initial wealth z, and they invest less than their unconstrained optimum. The fraction of credit constrained is large among the voluntary entrepreneurs (57%), while much fewer (23%) of involuntary entrepreneurs are credit constrained. The reason is that voluntary entrepreneurs have higher entrepreneurial ability θ on average, and hence larger unconstrained capital levels, $k_u(\theta)$. Indeed, in the simulated data from the model the average log talent ($\log\theta$) at the GMM estimates is 5.2 for voluntary entrepreneurs versus 3.5 for involuntary entrepreneurs and 3.7 for non-entrepreneurs.

Table 7 – Model predictions at the GMM estimates

Model statistic	Value
entrepreneurs, % of all agents	65.2
involuntary entrepreneurs, % of all agents	10.8
involuntary entrepreneurs, % of all entrepreneurs	16.6
voluntary entrepreneurs, % of all agents	54.4
voluntary entrepreneurs, % of all entrepreneurs	83.4
credit constrained, % of all entrepreneurs	51.3
credit constrained, % of voluntary entrepreneurs	56.8
credit constrained, % of involuntary entrepreneurs	23.4

The next table (Table 7b) breaks down the distribution of voluntary and involuntary entrepreneurs in the model by initial wealth, z and years of schooling, x (both taken from the data). The reported percentages in Table 7b use the same model-simulated data at the GMM estimates used in Table 7. We see that the majority (57.4%) of voluntary entrepreneurs have wealth above the median. This is intuitive since larger wealth makes it less probable that an entrepreneur will be credit constrained and hence prefer the alternative occupation. This effect is emphasized for schooling above the median, since in that case the alternative income is larger and thus the households needs higher z to be able to invest a sufficient amount and earn higher income as entrepreneurs. The distribution of voluntary entrepreneurs with years of schooling below vs. above the median is closer to uniform (56% vs. 44%). The smallest fraction of voluntary entrepreneurs is estimated among households with wealth below the median and schooling above the median. Intuitively, they are the most likely to be credit constrained and also have larger potential non-business income.

Looking at the involuntary entrepreneurs (panel B in Table 7b), we see that a large majority (over 70%) have years of schooling below the median (6 years) and also more than 60% have wealth below the median. There are two reasons for this. First, from our assumptions, the labor market constraint which forces households into involuntary entrepreneurship is more restrictive for lower schooling x. Second, having lower wealth z

makes it more likely that one would be credit constrained if one chose to start a business, and hence prefer the alternative occupation. Indeed, in the simulated data at the GMM estimates 70% of all credit-constrained involuntary entrepreneurs (not reported in the table) have both wealth and schooling below the median while none of the credit-constrained involuntary entrepreneurs have wealth above the median.

Table 7b – Model predictions, distribution of entrepreneurs by type

A. Percent of voluntary entrepreneurs with

	wealth, $z \leq$ median	wealth, $z >$ median	total
schooling, $x \leq$ median	29.2	27.1	56.3
schooling, $x >$ median	15.4	28.3	43.7
total	44.6	55.4	

B. Percent of involuntary entrepreneurs with

	wealth, $z \leq$ median	wealth, $z >$ median	total
schooling, $x \leq$ median	45.6	25.0	70.6
schooling, $x >$ median	15.6	13.8	29.4
total	61.2	38.8	

Figure 1 shows the estimated relationship between log initial wealth and the probability (rate) of entrepreneurship and illustrates how our model differs from the standard income-maximization occupational choice model of Evans and Jovanovic (EJ, 1989). The left panel shows the relationship between initial wealth and entrepreneurship overall – it is positive but there is a lot of noise. In contrast, the relationship between initial wealth and *voluntary* entrepreneurship is strongly positive with less noisiness (the middle panel). This is the familiar picture from EJ (1989) and others, interpreted as indicative of the presence of financial constraints. We see that the relationship between initial wealth and entrepreneurship is made weaker by the estimated negative relationship between initial wealth and involuntary entrepreneurship (the right panel).

Figure 1: Probability of entrepreneurship as function of wealth
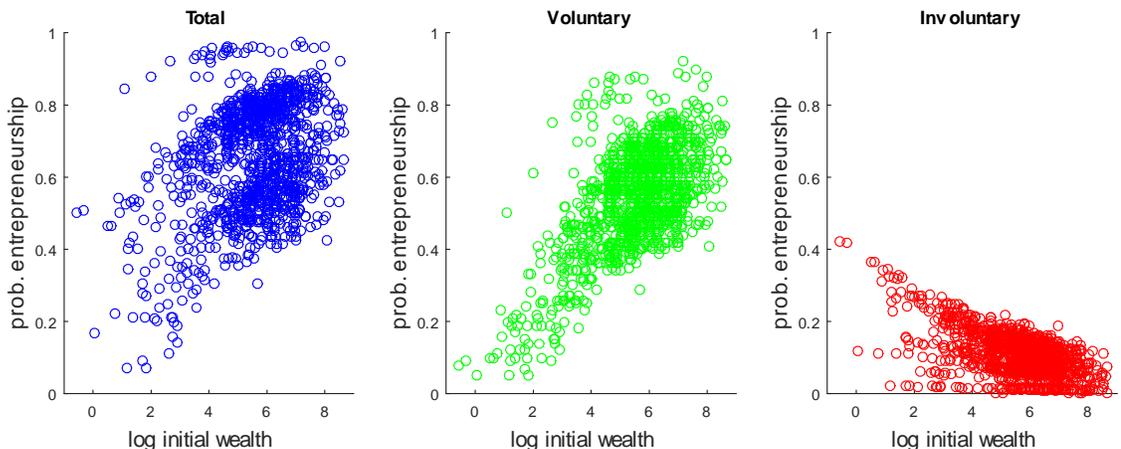
4.3 Model fit

We next assess the model fit to the data at the GMM parameter estimates. In Table 8 we report the model fit for the 11 chosen moments, as defined in Table 5, that we match (target) in the GMM estimation by minimizing the criterion function over the nine parameters ϕ. We see that the seven moments based on the percentage of entrepreneurs (lines 1-7) are are all matched well, within 5% deviation of their counterparts in the data. The four income moments (lines 8-11) are matched even closer – all are within 0.4% of the data counterparts.[13]

Table 8 – Model fit: matched moments at the GMM estimates

	moment	model	data	% deviation
1.	% entrepreneurs	65.2	66.1	-1.38
2.	% entrepreneurs, x in bottom tercile	78.7	79.5	-1.03
3.	% entrepreneurs, z in bottom tercile	59.5	58.9	1.03
4.	% entrepreneurs, x in top tercile	50.5	52.0	-2.90
5.	% entrepreneurs, z in top tercile	69.0	71.9	-4.13
6.	% entrepreneurs, z and x in bottom terciles	74.1	72.5	2.31
7.	% entrepreneurs, z and x in top terciles	57.0	54.3	4.90
8.	average gross income – entrepreneurs	512.5	513.6	-0.22
9.	average gross income – non-entrepreneurs	164.8	164.7	0.08
10.	avg. gross income – entr., z below median	349.7	350.3	-0.17
11.	avg. gross income – entr., x below median	387.1	385.7	0.38
	GMM criterion value (sum of squared deviations)			$5.9(10^{-3})$

Notes: x_m= median x, z_m= median z; income levels are in thousands Baht

In Table 9, we next assess the model fit on additional moments corresponding to other important dimensions that we did not target in the GMM estimation. A good fit within these moments can be interpreted as an additional validation of the model with data that are not used directly in the estimation.[14] Table 9 indicates that the model fits well (within 6% deviation) in most of these additional dimensions (lines 1-11 in Table 9). The model is relatively farthest from the data in matching the incomes of business households with wealth and schooling both below or both above the median (both are under-predicted, see lines 12 and 13).

---

[13]The overidentifying restrictions are however rejected with a J-statistic of 11.6. The test statistic magnitude is driven mostly by the 7th moment (% entrepreneurs in the top tercile of z and x).

[14]Moments 10 and 11 in Table 9 can be constructed from the matched moments in Table 8 and are reported only for completeness.

Table 9 – Model fit: non-matched moments at GMM estimates

|	moment	model	data	% deviation
1.	% entrepreneurs, x below median	77.3	74.8	3.36
2.	% entrepreneurs, z below median	61.5	60.3	1.86
3.	% entrepreneurs, x above median	55.3	57.4	-3.62
4.	% entrepreneurs, z above median	71.2	71.9	-0.95
5.	% entr., z below median, x below median	73.0	68.9	5.84
6.	% entr., z above median, x above median	62.2	63.8	-2.52
7.	% entr., z below median, x above median	46.3	49.0	-5.51
8.	% entr., z above median, x below median	83.0	82.5	0.64
9.	average gross income – all	390.8	395.3	-1.15
10.	avg. gross income – entr., z above median*	638.3	651.0	-1.94
11.	avg. gross income – entr., x above median*	669.7	680.6	-1.60
12.	avg. gross income – entr., z and x below med.	294.2	335.5	-12.3
13.	avg. gross income – entr., z and x above med.	778.8	858.0	-9.22

Note: income levels are in thousands Baht; *these moments can be obtained from moments in Table 8.

Figure 2 clarifies further the findings from Table 9 about exactly where the model matches well or less well the probability/fraction of entrepreneurship relative to the data. The Figure plots lowess regression lines and confidence intervals around the data (dashed lines). Since the initial wealth distribution is very skewed we use a percentile scale on the horizontal axis for better visualization. We see that, at the GMM parameters the model matches well the overall level and slope of the lowess fit of the data (both with respect to initial wealth and schooling). However, the model is unable to fully match the data at very low levels of wealth (it under-predicts entrepreneurship[15]) and for very low or very high levels of schooling (it over-predicts entrepreneurship).

Figure 2: Probability of Entrepreneurship – Model vs. Data

---
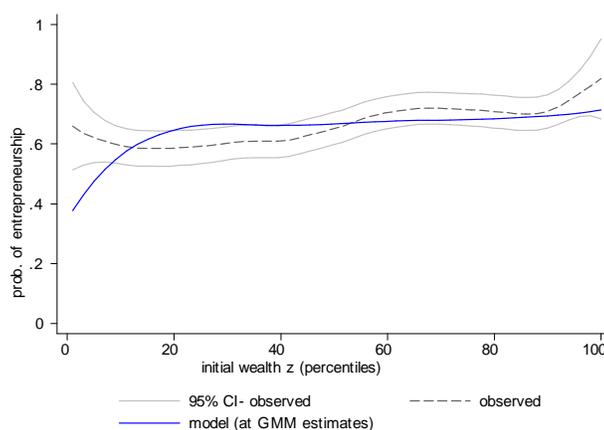
[15]Related to this result, Lee (2016) proposes a model aiming to explain the observation that many households with zero or negative net worth start businesses in the USA. The author shows that allowing for unsecured credit with an interest rate premium in addition to collateralized debt in the EJ (1989) setting raises the model-predicted probability of entrepreneurship at low asset levels closer the observed rate in the data.
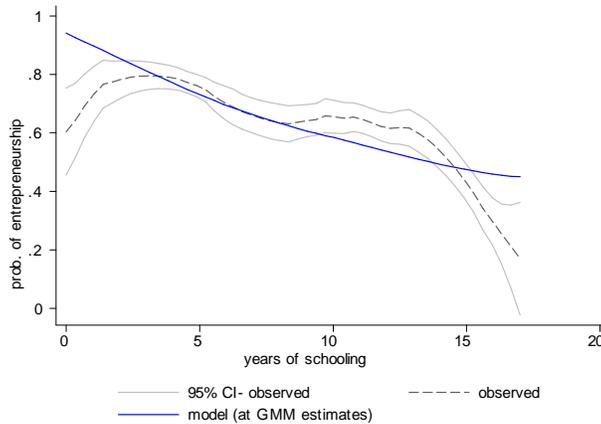
4.4 Misallocations – sources, levels and distribution

In this section we further explore the model predictions at the GMM estimates by examining the misallocations stemming from the estimated labor market and credit market imperfections relative to the first best. Essentially, there are two allocation tasks in the model. First, households differing in their entrepreneurial ability θ and labor market characteristics (schooling), x are allocated across the two occupations. Second, capital k is allocated among the households who run businesses. The labor and credit market constraints can cause misallocations in both of these dimensions. On the extensive (occupational choice) margin, a constrained household may end up in the suboptimal occupation. This misallocation could be either reflected in involuntary entrepreneurship, due to the labor market constraint, or in a severely credit-constrained household choosing the non-business occupation. On the intensive (capital utilization) margin, an entrepreneur (either voluntary or involuntary) can face a binding credit constraint and hence use a suboptimally low amount of investment k relative to the unconstrained level $k_u(\theta)$.

We evaluate the degree of both the occupational choice and capital use misallocations in the estimated model, as well as the incidence of the misallocations across households with different observable characteristics – initial wealth z and schooling, x. We also disentangle the effects of the labor and credit constraints. The misallocations are defined relative to the first best (unconstrained) benchmark. In our model, the first best corresponds to setting $\eta = 0$, that is, no labor market constraint; and having $\lambda \rightarrow +\infty$ (10^8 is used in the computation), that is, no credit constraint. All other parameters are held fixed at their GMM estimates.

Figure 3, the top panel ("estimated model vs first best") plots the differences between the predicted probability of entrepreneurship in the estimated model (with both labor and credit constraints present) and in the first best, across the households with different initial wealth z and schooling x taken from the data. Warm colors (red, orange, yellow) mean more predicted entrepreneurs relative to the first best while cool colors (blue, cyan) mean less entrepreneurs relative to the first best. If there was no misallocation, all estimated model vs. first best differences should equal zero (depicted in green). We see, however, that the labor and credit constraints lead to both 'over-supply' of entrepreneurs among some households and 'under-supply' among others. Specifically, for low values of schooling, there is a higher model-predicted rate of entrepreneurship (by up to 20 percentage
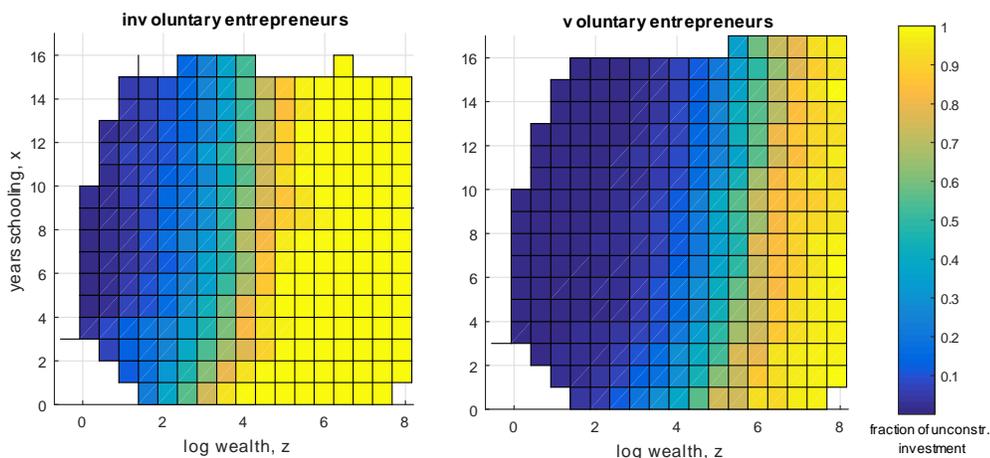
Figure 3: Misallocation in entrepreneurship

Note: Positive numbers (warm colors) mean more entrepreneurs relative to the first best, negative numbers (cool colors) mean less entrepreneurs relative to the first best.
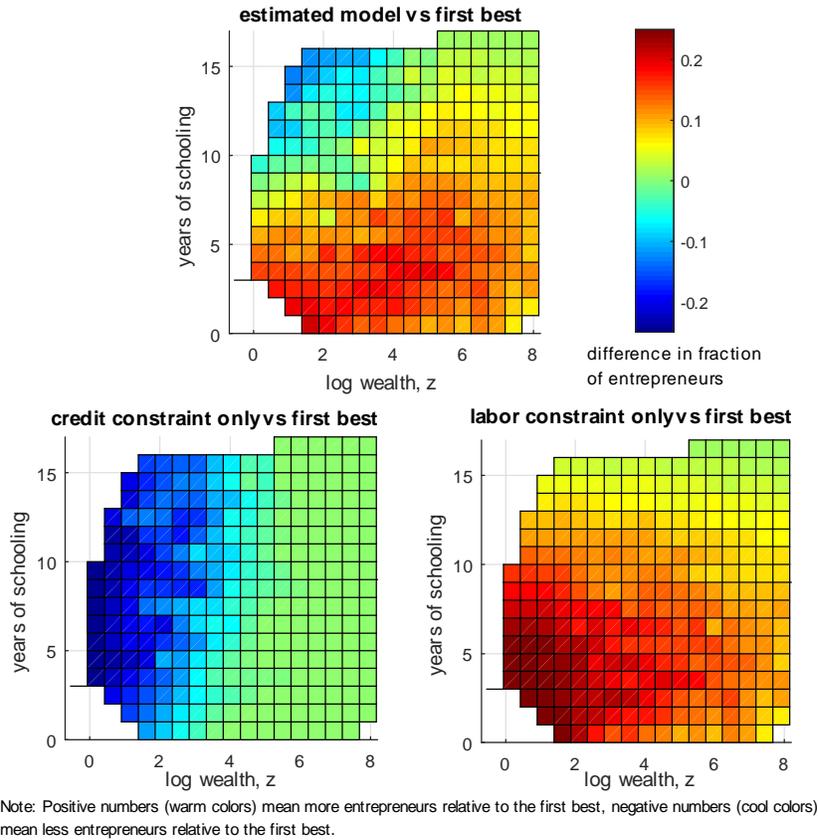
points) than in the first best. This is due to involuntary entrepreneurship, as the labor market constraint is assumed more likely to bind for low schooling. The differences are larger for low wealth levels, where the involuntary entrepreneurship effect is compounded by a tighter credit constraint. In contrast, for households with high schooling but low wealth, the model predicts *less* entrepreneurship (by up to 16 percentage points) than there would be in the first best – this is due to the credit constraint. For high schooling and high wealth (the top right corner) there is no misallocation because both constraints are not binding for such households.

The bottom two panels of Figure 3 decompose the overall difference in the expected rates of entrepreneurship between the estimated model and the first best by evaluating the misallocation effects stemming from the labor and credit constraints separately. In the bottom left panel ("credit constraint only vs first best") we set $\eta = 0$ (no labor market constraint) but keep the credit constraint parameter λ at its GMM estimate. As should be expected, the credit constraint alone results in a weakly lower rate of entrepreneurship compared to the first best throughout. This is most pronounced (by up to 26 percentage points) for low-wealth households but it has no effect on high wealth households who can invest at the unconstrained amount. The misallocation magnitude ("missing" entrepreneurs) is larger for higher levels of schooling since it is estimated as positively correlated with entrepreneurial ability. In the bottom right panel ("labor constraint only vs first best"), we set $\lambda = 10^8$ which eliminates credit constraints and keep η at its GMM estimate. In contrast to the credit constraint effect, now the direction of the misallocation in the rate of entrepreneurship relative to the first best is the opposite – the labor market constraint results in an excess amount (up to 30 percentage points) of entrepreneurs. This is

Figure 4: Misallocation in investment

what we call involuntary entrepreneurship. The degree of misallocation is the highest for households with low schooling and low wealth, both of which are also positively correlated with low entrepreneurial talent.

We next analyze the misallocations on the intensive margin (investment). Figure 4 illustrates the level and distribution, over observed initial wealth and schooling, of the investment misallocations among entrepreneurs, relative to the unconstrained (first best) investment level. Specifically, we plot the ratio (integrated over θ) of actual investment level k to the unconstrained investment level k_u for voluntary and involuntary entrepreneurs in the model at the GMM estimates. We see that, for both groups of entrepreneurs, the investment misallocation is the most severe for low wealth households. Holding wealth constant, the investment of voluntary entrepreneurs is more misallocated (constrained) relative to the first best than that of involuntary entrepreneurs. The reason is that voluntary entrepreneurs have higher ability θ on average.

We finish by summarizing the aggregate implications of the misallocations at the extensive and intensive margins. Table 10 (lines 1-3) reports the occupational choice misallocations stemming from the labor and credit constraint. We already saw that at the GMM estimates, 10.8% of households are classified as involuntary entrepreneurs. Table 10 indicates that the major cause for this 'excess entrepreneurship' misallocation is the labor market constraint, accounting for 10.5% of the 10.8% (compare the 'model' with 'labor constraint only' columns). On the other hand, the aggregate number of 'missing' voluntary entrepreneurs due to the credit constraint is estimated at 1.5% of all households (55.9% – 54.4%). Overall, these two effects add up to 9.3% more entrepreneurs in the estimated model relative to the first best.

Table 10 – Misallocation aggregates

	model	credit constraint only	labor constraint only	first best
1. % voluntary entr.	54.4	54.4	55.9	55.9
2. % involuntary entr.	10.8	0	10.5	0%
3. % total entr.	65.2	54.4	66.4	55.9
4. k used by vol. entr.	46.5 (47.8)	46.5 (47.8)	100 (100)	100 (100)
5. k used by invol. entr.	1.6 (8.4)	0 (0)	1.9 (10)	0 (0)
6. k used by all entr.	48.1 (41.2)	46.5 (47.8)	101.9 (85.8)	100 (100)

Lines 4-6 of Table 10 show the aggregate level of misallocation in capital use. These numbers do include the compositional effects on the extensive margin and so they should be interpreted together with Figure 4. Normalize total capital used in the first best as $K^{fb} = 100$ and normalize capital per entrepreneur in the first best as $k^{fb} = 100$. Lines 4–6 in Table 10 then report the (percentage of) total capital and capital per entrepreneur (in the brackets) relative to the corresponding first best levels. The 'model' column shows that, at our GMM estimates, only about 48% of the total capital amount in the first best (41% per business household) is used. Of this total, 1.6% is used by involuntary entrepreneurs. Shutting down the labor constraint reduces capital use to 46.5% of the first best total, quantifying the aggregate impact of the estimated credit constraint on voluntary entrepreneurs. On the other hand, facing the labor market constraint alone results in *over-utilization* of capital by 1.9% relative to the first best total (but not per person) as capital is used inefficiently by involuntary entrepreneurs.

5 Counterfactuals and Welfare Analysis

5.1 Relaxing the labor or credit constraints

Involuntary entrepreneurship arises in the model if both of the following conditions are true: (i) the household does not have access to the alternative occupation (for example, a wage job), which we can interpret as a labor market constraint/friction and (ii) household income is maximized in the alternative occupation. The labor constraint is important for condition (i), while the credit constraint affects (ii). In this section we evaluate and disentangle the effects of the labor and credit constraints on entrepreneurship (total, voluntary and involuntary) and on household income. Since the households are assumed risk-neutral, changes in household income can be directly interpreted as welfare effects.

In the first counterfactual, we set the labor constraint parameter η to zero while keeping all other parameters at their GMM estimates. This means that involuntary entrepreneurship is completely eliminated – all households have free occupational choice as, for example, in EJ (1989). This counterfactual also affects average income in the economy since previously involuntary entrepreneurs are now able to choose the non-business occupation which is income maximizing for them. The voluntary entrepreneurs are not affected by the relaxation of the labor constraint.

The results reported in Table 11 are computed from the model-simulated data at the GMM estimates. Panel A shows that the elimination of the labor constraint reduces the rate of entrepreneurship to 54.4%. In

Panel B we also compute the mean, median and percentiles or the expected *net income* in the estimated model (the column labeled 'baseline') and the resulting income change from relaxing the labor constraint ('change from baseline'). Net income, as opposed to gross, is what households compare to make their occupational choice. The expected net income is defined as $E(q^E - rk + (r-1)z)$ for entrepreneurs, that is, output minus the cost of capital plus interest income, where the expectation is taken over the talent shock ε. Similarly, define net income as $q^A + (r-1)z$ for non-entrepreneurs. Table 11 (Panel B, 'baseline') shows that mean net income is the highest for voluntary entrepreneurs and the lowest for involuntary entrepreneurs. This is intuitive since involuntary entrepreneurs are more productive in the non-business occupation.

Relaxing the labor constraint increases households incomes throughout the income distribution (Panel B, 'free occ. choice'), at the mean, median and different percentiles. The income changes include the effects of mobility within the income distribution as a result of the counterfactual. For example, an ex-ante involuntary entrepreneur who is now free to enter the non-business occupation could move from the 10th to the 30th income percentile, etc. We observe that relaxing the labor constraint has the strongest effect at the 10th income per-centile (+6.1%) where households are most likely to be involuntary entrepreneurs in the baseline. We also see a large positive effect on the mean entrepreneurial income (16% increase) accompanied with a fall in the mean income of non-business households (-6.1%). The latter effect should not be confused with a negative impact on non-business income. Clearly, no one loses from the relaxation of the labor market constraint, since everyone's income weakly increases (one can either stay in one's current occupation or switch to a preferred one).[16] In-stead, the reason for the fall in mean non-business income is due to a composition effect – some unproductive entrepreneurs (with low talent θ and low schooling x) exit the business occupation and enter the non-business occupation. Finally, relaxing the labor constraint also affects the number of constrained entrepreneurs (those with $k = \lambda z$). The simulated data show that the percent of constrained entrepreneurs increases from 51.3% in the baseline to 56.8% (not reported in the table). The reason is that without the labor constraint all entrepreneurs are voluntary and have higher ability θ on average.

---

[16]Of course, this is only unambiguously true when abstracting from general equilibrium effects.

Table 11 – Relaxing the labor or credit constraints

	baseline	free occ. choice ($\eta = 0$)	relaxed credit (2λ)
A. Occupational choice			
entrepreneurs	65.2%	54.4%	65.7%
of which voluntary	83.4%	100%	83.8%
of which involuntary	16.6%	0%	16.2%
B. Net income	*baseline*	*change from baseline*	*change from baseline*
mean, all	378.0	+1.8%	+4.7%
10th percentile	188.0	+6.1%	+9.2%
30th percentile	281.6	+2.6%	+6.6%
median	353.6	+1.8%	+5.7%
70th percentile	433.5	+1.3%	+4.2%
90th percentile	595.1	+1.1%	+3.0%
mean, entrepreneurs	475.5	+16%	+5.1%
mean, voluntary entr.	553.9	no change	+4.8%
mean, involuntary entr.	82.6	n.a.	+0.6%
mean, non-business	195.3	-6.1%	+0.3%

The second counterfactual we study is relaxing the credit constraint which we analyze by doubling the estimate of λ from the estimated baseline (from 0.23 to 0.46), keeping all other parameters at their GMM estimates. In view of all the evidence for credit constraints in developing countries we consider this exercise more informative than completely eliminating the credit constraint. Relaxing the credit constraint has a minor effects on involuntary entrepreneurship (its share falls from 16.6% to 16.2%) and on entrepreneurship overall (it increases from 65.2% to 65.7%) – see Figure 11, Panel A. This reinforces our finding in Section 4.4 that the labor market constraint is more important in causing involuntary entrepreneurship.

Table 11, Panel B ('relaxed credit') shows, however, that relaxing the credit constraint can have significant impact on households' incomes by mitigating the misallocations in capital utilization. The increase in the mean net income (+4.7%) is more than double the corresponding increase (+1.8%) from relaxing the labor constraint, with the impact on income being larger across the income distribution. Households at the 10th income percentile experience the largest income/welfare gains (+9.2%) as they can invest amounts closer to their first-best capital levels. The voluntary entrepreneurs gain about the same (+4.8%) as the average agent, while the involuntary entrepreneurs and non-entrepreneurs have only minor income gains, the former since they are mostly constrained by talent, the latter due to the small compositional shift in the economy towards entrepreneurship. Looking at the number of credit constrained households in the simulated data (not reported in the table), unsurprisingly we see a large drop from 51.3% to 33.7% in the fraction of constrained entrepreneurs. Among the voluntary entrepreneurs, the fraction of credit constrained falls from 56.8% to 37.5% while the corresponding impact among involuntary entrepreneurs is a decrease from 23.4% to 14%.

Figure 5 illustrates the distribution of income/welfare gains from each of the two counterfactuals stratified by households' log initial wealth, z and years of schooling, x. We use the simulated data from the model to compute the change in expected income (integrated over the talent shock ε) of each households with characteristics (z_i, x_i) from the data, before vs. after relaxing each constraint. The Figure shows that relaxing the labor

constraint leads to very large income gains for low wealth individuals (up to 40%). These gains are on average monotonically decreasing in initial wealth and (except for very low x values) in the years of schooling, as it is less likely that one would have been an involuntary entrepreneur for high z and x.

In contrast, the income gains from relaxing the credit constraint are non-monotonic over initial wealth, with the households with intermediate wealth levels gaining the most. The reason is that they are most likely to be credit constrained entrepreneurs. The income gains from relaxing the credit constraint decline in schooling on average, since households with larger values of x are more likely to have higher ability θ and hence less likely to have been constrained.

Figure 5: Expected Income Gains – Relaxing the Labor or Credit Constraints
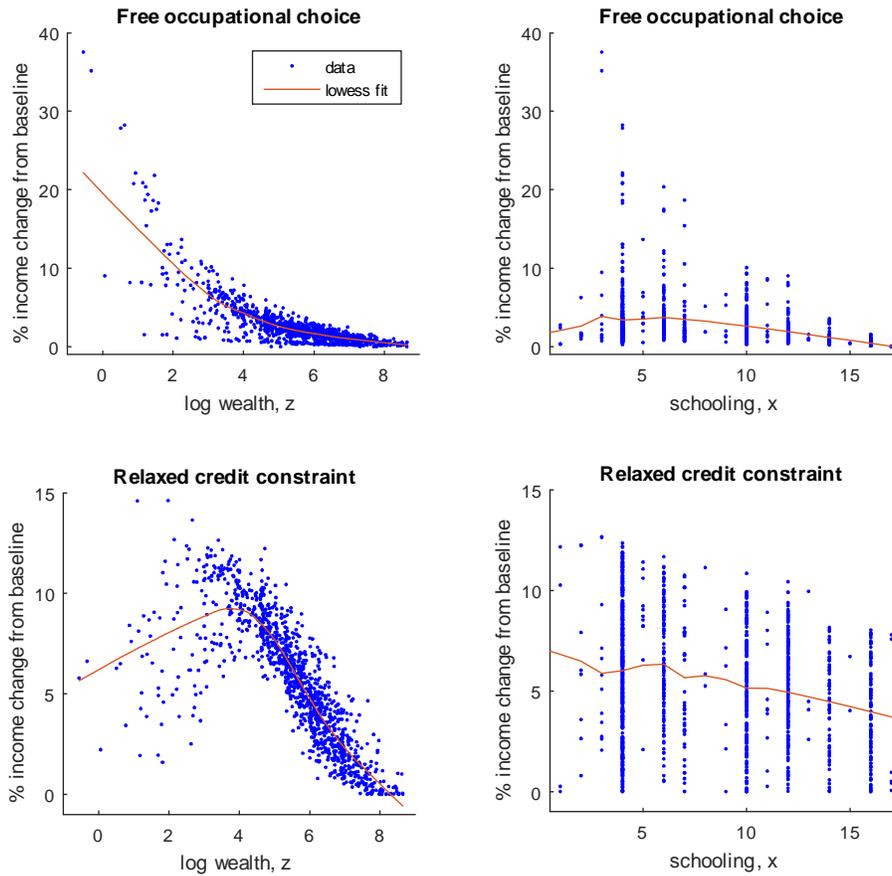
5.2 Microfinance

We next consider the counterfactual of offering households the option to borrow and invest in their business *up to* an additional M dollars. This counterfactual can be interpreted as a microfinance program, with the requirement that loans be only used to buy/rent business capital at the current interest rate r. We analyze the effects of this policy on the rate of involuntary entrepreneurship and household income. All model parameters are held at the baseline GMM estimates. We set the maximum microfinance loan to 10% of the median gross

income in our sample, $M = 20,000$ Baht.

Households choose k to solve

$$\max_k \theta k^\alpha - rk \quad \text{subject to} \quad k \le \lambda z + M \quad \text{(MF)}$$

and would optimally choose to run a business if their income from entrepreneurship when using the investment level k solving problem (MF) is higher than their alternative income from the non-business occupation which is unaffected by the policy. Clearly, all households who are initially not credit constrained are not affected by this policy while all constrained households have an incentive to participate (borrow).

Table 12 shows the microfinance policy effects on occupational choice and household income overall and across different groups of households. The fraction of entrepreneurs overall goes up by about 1 percentage point, from 65.2% to 66.3%. Within the larger number of businesses, the policy induces more voluntary entrepreneurship (+0.8%) while the rate of involuntary entrepreneurs falls from 16.6% to 15.8%.

In terms of household income, Table 12, Panel B shows that the microfinance policy raises average income by 3.4% but the income/welfare gains are unevenly spread among the households. The poorest, those at the 10-th income percentile benefit the most from the availability of additional credit (a 14% income increase post vs. pre-policy), while the richest households, those at the 90-th income percentile benefit only marginally as they are more likely to have been unconstrained ex-ante.

The mean income of entrepreneurs goes up by 3% for two reasons – first, the additional credit relaxes the credit constraint and allows some entrepreneurs to earn more and second, there is a compositional shift from involuntary to voluntary entrepreneurs. The mean non-business income also goes up slightly (+1.3%) as some agents with low schooling exit the occupation.

Table 12 – Microfinance policy evaluation

	baseline	policy
A. Occupational choice		
entrepreneurs	65.2%	66.3%
of which voluntary	83.4%	84.2%
of which involuntary	16.6%	15.8%
B. Net income	*baseline*	*change from baseline*
mean, all	378.0	+3.4%
10th percentile	188.0	+16.5%
30th percentile	281.6	+5.6%
median	353.6	+2.6%
70th percentile	433.5	+1.7%
90th percentile	595.1	+1.0%
mean, entrepreneurs	475.5	+3.0%
mean, voluntary entr.	553.9	+2.2%
mean, involuntary entr.	82.6	+1.3%
mean, non-business	195.3	+0.6%

The effects of the microfinance loan policy on income are further illustrated on Figure 6, stratified by log

initial wealth, z and years of schooling, x. We use the simulated data from the model and compute the change in expected income (integrated over the shock ε) of each households with characteristics (z_i, x_i) from the data, before and after the policy. We see that the microfinance policy benefits poorer households significantly (income gains of up to 60 percent relative to the baseline). The gains quickly decrease for wealthier households since they are less likely to have been credit constrained ex-ante and benefit from the access to microfinance. The income gains are smaller but more evenly spread by years of schooling. This is due to the interaction of wealth and schooling in the data. The bottom panel of the figure shows that the households who gain the most from the policy are those with the lowest wealth and schooling. Low-wealth agents with high schooling do not gain much, as they are more likely to be engaged in the non-business occupation. Only the involuntary entrepreneurs among them stand to gain from the microfinance policy.

Figure 6: Microfinance – income gains by wealth and schooling
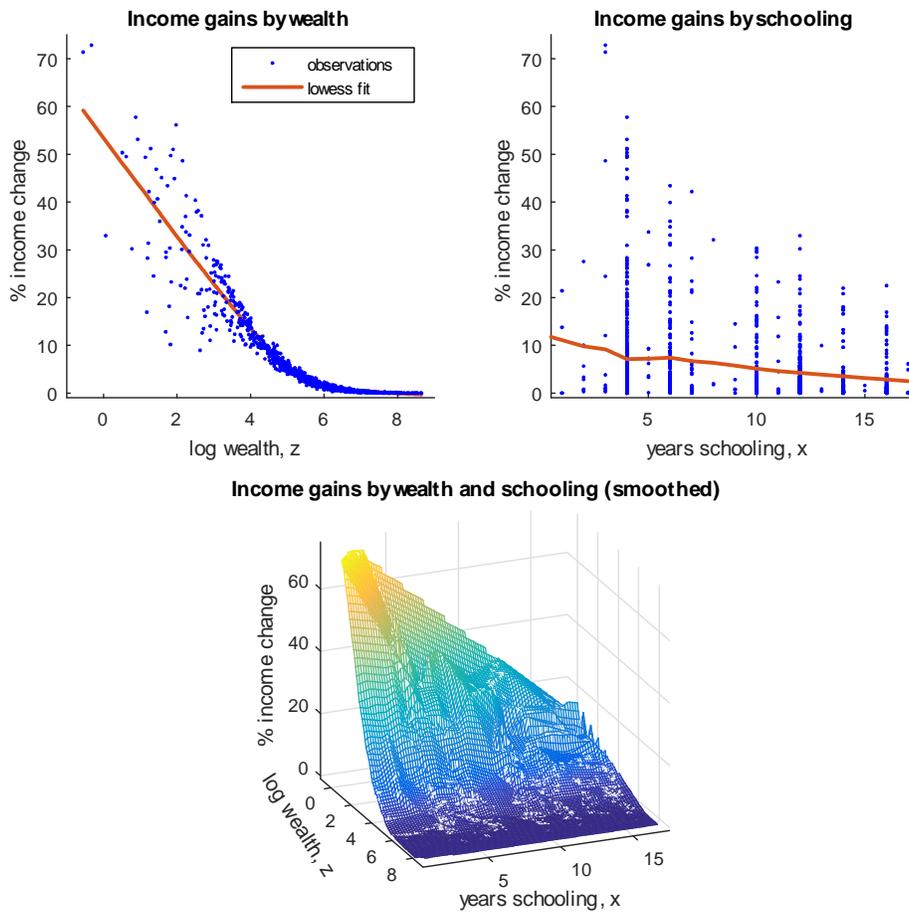
The main difference between the results of the microfinance counterfactual and the counterfactual of relaxing the credit constraint by doubling the credit constraint parameter λ is that with microfinance the gains in income (welfare) are monotonically decreasing in household wealth. The reason is that under the microfinance

policy poorer households (with low z) receive a relatively larger increment in their ability to borrow compared to wealthier households, as the maximum loan size M is assumed uniform. In contrast, when the credit constraint is directly targeted by increasing λ (for example, this could be interpreted as better enforcement or better property rights enabling posting more collateral) the effect is non-monotonic as explained above.

6 Robustness

6.1 Alternative definitions

We study the sensitivity of our results to the definitions of business ownership and labor market characteristics x. Column (2) in Table 13 reports the GMM estimates when we define business ownership by whether a household derives the *majority* of their gross income from business. With this narrower definition, the rate of business ownership in the sample is reduced to 50%, compared to 65% in the baseline (households that report owning a business) – compare column (2) with column (1) in Table 13. The alternative definition of business ownership also results in smaller estimated probability of involuntary entrepreneurship, 5.7% among all households in the sample or 11.4% among the business households. A possible interpretation is that the proportion of income drawn from a business may be correlated with entrepreneurial talent, such that households that rely more on business ownership as their major income source have higher ability and therefore are less likely to be involuntary entrepreneurs.

Column (3) in Table 13 uses years of schooling of the head of the household as a proxy for labor market characteristics x, instead of the principal earner's years of schooling used in the baseline. The probability of involuntary entrepreneurship is estimated as 6.9% on the whole or 10.5% among business households. The reduction in the estimated rate of involuntary entrepreneurship could be because household heads have lower schooling on average compared to the principal earners and so their implied alternative income is lower.

In column (4) of Table 13, we re-define labor market characteristics, x as a composite index of schooling and age. Specifically, we perform a principal component analysis using the principal earner's years of schooling and the difference between the maximum age and the principal earner's age (normalized by 4 to match the years of schooling range) and define x to be the first principal component, in which the loading on schooling is estimated to be 74%. With this broader definition of labor market characteristics we find a slight reduction in the estimated number of involuntary entrepreneurs to 15% of all business households. The GMM parameter estimates are also close to those in the baseline.

6.2 Estimation on subsamples stratified by gender and age

We next estimate the model on different subsamples stratified by the gender and age of households' principal earners to see whether and how much the estimated rate of involuntary entrepreneurship differs by these characteristics. Table 13, columns (5)–(8) report the results. In columns (5) and (6) in which the sample is stratified by gender, we see that the estimated rate of involuntary businesses is significantly lower in the 'male' sample (11.7% of all businesses), as opposed to 19.8% in the 'female' sample. The credit constraint (the parameter λ) is also estimated to be tighter in the 'female' sample. The labor market constraint (the parameter η) is also tighter for households with female principal earners. These findings, together with the fact that the observed

rate of business ownership in the data is higher for the 'female' sample (70% vs. 61% for the 'male'), suggest that the misallocations due to involuntary entrepreneurship and credit constraints are more pronounced among households with female principal earners. This is of potential policy significance.

In columns (7) and (8) of Table 13, we stratify the sample by the age of the principal earner in the household – below or above the median age. We find that 'younger' households are less constrained in both the credit and labor market (higher λ and lower η estimates) compared to 'older' households. The rate of involuntary entrepreneurship is also significantly lower among younger households (16% vs. 24.2% of all businesses). This suggests more pronounced misallocations in both investment and occupational choice for households with older principal earners.

Table 13 – Robustness

parameter \ specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
return to capital, α	0.23	0.41	0.40	0.26	0.30	0.11	0.23	0.33	0.26
credit constraint, λ	0.23	0.43	0.28	0.28	0.17	0.06	1.32	0.20	0.27
return to schooling, γ	0.75	0.72	0.13	0.83	0.95	0.48	0.99	0.60	0.74
labor market constraint, η	0.41	0.15	0.18	0.43	0.25	0.56	0.27	0.66	n.a.
non-business income, μ	28.5	38.7	120	22.0	18.5	49.9	14.3	46.8	29.1
talent – constant, δ_0	3.42	3.36	3.66	3.53	3.27	4.14	2.09	4.11	3.34
talent – wealth, δ_1	0.13	-0.03	-0.06	0.10	0.11	0.13	0.21	-0.02	0.11
talent – schooling, δ_2	0.17	0.25	0.31	0.09	0.18	0.02	0.39	0.12	0.15
talent – std. deviation, σ	0.96	0.73	0.60	0.94	1.03	1.02	1.02	0.81	0.98
entry cost, c									*49.6*
entrepreneurs, % of all	65.2	50.3	65.2	63.7	60.6	69.7	50.0	75.3	64.8
involuntary, % of all	10.8	5.7	6.9	9.5	7.1	13.8	8.0	18.2	14.3
involuntary, % of all entr.	16.6	11.4	10.5	15.0	11.7	19.8	16.0	24.2	22.0
GMM criterion value	.006	.003	.002	.008	.010	.010	.065	.011	.007

Notes: (1) baseline; (2) alternative definition of business households based on major source of income; (3) alternative definition of labor characteristics, x – head of household's years of schooling; (4) alternative definition of labor characteristics, x – first principal component of age and schooling (5) subsample, male principal earner; (6) subsample, female principal earner; (7) subsample, principal earner with age below median; (8) subsample, principal earner with age above median; (9) entry cost specification of the labor constraint.

6.3 Alternative labor market constraint specification

Finally, we also consider an alternative specification for the labor market (occupational choice) constraint. In the baseline model, a household with observable market characteristics x faces a probability of not finding a job in the wage labor market. Suppose instead that households must pay a fixed amount c to enter the non-business occupation. In the estimation, we allow c to be either positive or negative. That is, we allow the data to determine whether entry into the alternative occupation is costly or beneficial for the agents.

Given c, an agent with initial wealth z and labor characteristics x would choose to run a business if

$$y^E(\theta, z) > y^A(x, z) - c$$

32

or, equivalently,

$$(1-\alpha)\theta^{\frac{1}{1-\alpha}}(\tfrac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} > \mu(1+x)^\gamma - c \quad \text{if } \theta \le B(z)$$
$$\theta(\lambda z)^\alpha - \lambda rz > \mu(1+x)^\gamma - c \qquad \text{if } \theta > B(z)$$

The resulting expressions for the probability of entrepreneurship and expected income are derived in Appendix C.

We estimate the model with the alternative specification of the labor market constraint. The results are reported in column (9) of Table 13. Reassuringly, the estimates of the nine common parameters are very close to those in the baseline model. The parameter c is estimated to be positive (a *cost* of entry into the non-business occupation) and equals 49.6 thousand Baht which is approximately 30% of the average income of non-business households in the data.

With this alternative specification of the labor market friction, we define involuntary entrepreneurship as the difference between the rate of entrepreneurship for our estimated $c > 0$ and the rate of entrepreneurship at $c = 0$ (no entry cost, as in EJ, 1989), holding all other parameters fixed at their GMM estimates from column (9) in Table 13. The estimated rate of involuntary entrepreneurship is larger in the entry cost specification – about 22% of all businesses, relative to 17% in the baseline. A possible reason is that, unlike in our baseline specification, the entry cost specification assumes that the labor market friction is uniform across the households. In addition the baseline specification achieves a better fit.

Note that a positive cost of entry into the non-business occupation is isomorphic to assuming an additional non-pecuniary benefit of running a business, that is, an agent starts a business if her income from running, $y^E(\theta, z)$ it plus an additional benefit, c exceeds the alternative income $y^A(x, z)$. However, we think that our preferred interpretation of c as labor market entry cost (constrained occupational choice) is more plausible in the Thai setting in view of the evidence reviewed in the introduction about many people in developing countries running businesses out of necessity.

7 Conclusions

The classical theory of occupational choice is predicated on the observed choice being always better than the alternative. In this paper, we model and empirically explore the idea that some observed occupational choices can be involuntary, especially in the context of a developing country. We structurally estimate via GMM the possibility that some agents do not have access to labor market (wage) employment, nesting the standard model of income-maximizing occupational choice as a special case. Specifically, we define involuntary entrepreneurs as business owners who would maximize their income in a non-business occupation (for example, wage employment), but who are not able to access that occupation due to frictions in the labor market. Our baseline structural estimation results classify about 17% of all business owners in our 2005 Thai urban data as involuntary entrepreneurs, with other robustness runs indicating a range from as low as 11% to as high as 22%.

We use the structurally estimated model to quantify the extent and distribution of occupational and investment misallocations across households with different observable characteristics. We find significant misallocations on both the occupational choice (extensive) margin and the investment (intensive) margin. The misallocations go in both directions (too many or too few entrepreneurs, too much or too little capital used) depending on the interaction between the labor and credit constraints for the different households. Broadly

speaking, credit constraints suppress entrepreneurship and investment while labor market constraints cause an excess of involuntary entrepreneurs.

We also evaluate the effects of relaxing the credit and labor constraints and the impact of a microfinance policy on the rate of entrepreneurship (voluntary and involuntary) and on household income, on average and stratified by wealth and schooling. Our results suggest that there are large potential income gains, especially for poorer households, from relaxing either the labor market or credit constraint or from providing access to micro-credit, however the fraction of involuntary entrepreneurs can only be significantly reduced and their incomes increased by addressing the labor market constraint.

A limitation of our approach which is also present in much of the occupational choice literature, is that our model is essentially static and hence does not fully capture dynamic decision-making by households. An extension to a dynamic model, with assets and capital accumulation, could capture the non-linear relationship between wealth and entrepreneurship over time as in Buera (2009). Second, the counterfactual analysis was done in partial equilibrium, e.g., assuming that the effects are local and wages and interest rates do not change. Policy interventions might have different effects when general equilibrium effects are incorporated (e.g., Kaboski and Townsend, 2011; Poschke, 2013 or Buera et al., 2014). For example, relaxing credit constraints might not only raise entrepreneurs' incomes but also increase labor demand and wages (unmodeled here). If the labor market friction is not affected, the higher wage may increase involuntary entrepreneurship. On the other hand, if the increase in labor demand relaxes the labor market constraints, then the misallocations due to involuntary entrepreneurship can be reduced. Further analysis of these effects can be important. Third, while we emphasize and quantify the importance to labor market frictions for involuntary entrepreneurship, our way of modeling these frictions has been very stylized. Further research on the microfoundations of the relevant labor market constraints in developing country settings remains necessary.

References

[1] Aghion, P. and P. Bolton. 1997. A theory of trickle-down growth and development. *Review of Economic Studies* 64(2): 151-172.

[2] Banerjee, A. and E. Duflo. 2007. The economic lives of the poor. *Journal of Economic Perspectives* 21(1): 141-168.

[3] Banerjee, A., E. Breza, E. Duflo and C. Kinnan. 2015. Do credit constraint limit entrepreneurship? Heterogeneity in the returns to microfinance. working paper.

[4] Banerjee, A. and E. Duflo. 2011. *Poor economics: a radical rethinking of the way to fight global poverty.* New York: Public Affairs.

[5] Banerjee, A. and A. Newman. 1993. Occupational choice and the process of development. *Journal of Political Economy* 101(2): 274-298.

[6] Block, J. and M. Wagner. 2010. Necessity and opportunity entrepreneurs in Germany: characteristics and earnings differentials. *Schmalenbach Business Review* 62: 154-174.

[7] Buera, F. 2009. A dynamic model of entrepreneurship with borrowing constraints: theory and evidence. *Annals of Finance* 5(3), 443-464.

[8] Buera, F., J. Kaboski and Y. Shin. 2014. The Macroeconomics of Microfinance. *NBER Working Paper* 17905.

[9] de Mel, S., D. McKenzie and C. Woodruff. 2010. Who are the microenterprise owners? Evidence from Sri Lanka on Tokman v. de Soto. *International Differences in Entrepreneurship*. J. Lerner and A. Schoar (eds.): 63-87.

[10] Donovan, K. 2015. Financing, Subsistence Entrepreneurs and Talent Misallocation. working paper.

[11] Earle, J. and Z. Sakova. 2007. Business startups or disguised unemployment? Evidence on the character of self-employment from transition economies. *Labor Economics* 7: 575-601.

[12] Evans, D. and B. Jovanovic. 1989. An Estimated Model of Entrepreneurial Choice under Liquidity Constraints, *Journal of Political Economy* 97(4): 808-27.

[13] Falco, P. and L. Haywood. 2013. Entrepreneurship versus joblessness: explaining the rise in self-Employment. *Discussion Papers of DIW Berlin 1334*, DIW Berlin, German Institute for Economic Research.

[14] Gunther, I. and A. Launov. 2012. Informal employment in developing countries. *Journal of Development Economics* 97(1): 88-98.

[15] Hamilton, B. 2000. Does entrepreneurship pay? An empirical analysis of the returns of self- employment. *Journal of Political Economy*, 108(3): 604-631.

[16] Hurst, E. and B. Pugsley. 2011.What Do Small Businesses Do? *Brookings Papers on Economic Activity* 2011(2): 73–118.

[17] Kaboski, J. and R. Townsend. 2011. A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5): 1357-1406.

[18] Karaivanov, A. 2012. Financial constraints and occupational choice in Thai villages. *Journal of Development Economics* 97(2): 201-220.

[19] Lee, J. 2016. On the relationship between household wealth and entrepreneurship. working paper.

[20] Lloyd-Ellis, H. and D. Bernhardt. 2000. Enterprise, inequality and economic development. *The Review of Economic Studies* 67(1): 147-168.

[21] Minniti, M., W. Bygrave, and E. Autio. 2005. *Global Entrepreneurship Monitor, 2005 Executive Report*. GEM Consortium. Global Entrepreneurship Research Association, London Business School.

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

[23] NORC at the University of Chicago, Thai Fiscal Policy Office. 2008. Townsend Thai Project Initial House-hold Survey 2005 (Urban Area) [Computer file]. 1st Data Distributor Version: NORC at The University of Chicago, Fiscal Policy Office [Producer]. Bangkok, Thailand: The University of Chicago-UTCC Research Center [distributor].

[24] Paulson, A., R. Townsend and A. Karaivanov. 2006. Distinguishing limited liability from moral hazard in a model of entrepreneurship. *Journal of Political Economy* 144(1): 100-144.

[25] Piketty, T. 1997. The dynamics of the wealth distribution and the interest rate with credit rationing. *Review of Economic Studies* 64(2): 173-189.

[26] Poschke, M. 2013, Who becomes an entrepreneur? Labor market prospects and occupational choice. *Journal of Economic Dynamics and Control* 37: 693-710.

[27] Schoar, A. 2010. The divide between subsistence and transformational entrepreneurship, in *Innovation Policy and the Economy*. Vol. 10, J. Lerner and S. Stern, eds.,p. 57-81. University of Chicago Press.

Appendix A – Proofs

Proof of Proposition 1

Using the definitions of $y^E(\theta, z)$, $y^A(z, x)$ and $\Delta(z, \theta, x)$ from the main text, we obtain,

$$\Delta(z, \theta, x) \geq 0 \Leftrightarrow \begin{cases} (1-\alpha)\theta^{\frac{1}{1-\alpha}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} - \mu(1+x)^\gamma \geq 0 & \text{if } \theta \leq B(z) \\ \theta(\lambda z)^\alpha - r\lambda z - \mu(1+x)^\gamma \geq 0 & \text{if } \theta > B(z) \end{cases} \qquad (9)$$

which, in terms of the agent's entrepreneurial ability θ, is equivalent to,

$$\Delta(z, \theta, x) \geq 0 \Leftrightarrow \begin{cases} \theta \geq (\frac{\mu}{1-\alpha})^{1-\alpha}(1+x)^{\gamma(1-\alpha)}(\frac{r}{\alpha})^\alpha & \text{if } \theta \leq B(z) \\ \theta \geq (\lambda z)^{-\alpha}[\mu(1+x)^\gamma + r\lambda z] & \text{if } \theta > B(z) \end{cases}$$

Proof of Lemma 1:

Using (5), we have, since $\mathbf{1}_{B>A} = \mathbf{1}_{B>C}$ for all x, z,

$$\tilde{P}_E = P(\Delta \geq 0) = \mathbf{1}_{B>A}[P(\frac{\varepsilon}{\sigma} > \frac{\ln B - \bar{\theta}}{\sigma}) + P(\frac{\ln A - \bar{\theta}}{\sigma} \leq \frac{\varepsilon}{\sigma} \leq \frac{\ln B - \bar{\theta}}{\sigma})] +$$
$$+ (1 - \mathbf{1}_{B>A})[P(\frac{\varepsilon}{\sigma} \geq \frac{\ln C - \bar{\theta}}{\sigma}) + 0]$$

Let $\Phi(\cdot)$ be the standard Normal cdf. We then obtain,

$$P(\Delta \geq 0) = \mathbf{1}_{B>A}\left\{1 - \Phi\left(\frac{\ln B - \bar{\theta}}{\sigma}\right) + \Phi\left(\frac{\ln B - \bar{\theta}}{\sigma}\right) - \Phi\left(\frac{\ln A - \bar{\theta}}{\sigma}\right)\right\} +$$
$$+ (1 - \mathbf{1}_{B>A})\left\{1 - \Phi\left(\frac{\ln C - \bar{\theta}}{\sigma}\right)\right\} =$$
$$= \mathbf{1}_{B>A}(1 - \Phi(a)) + (1 - \mathbf{1}_{B>A})(1 - \Phi(c)).$$

Appendix B – Derivation of the income moments

For an agent with entrepreneurial ability θ and initial wealth z, gross income conditional on entrepreneurship is $q^E(\theta) = \theta(k^*)^\alpha$ where $k^* = \min\{k_u(\theta), \lambda z\}$. We can thus write expected gross income conditional on entrepreneurship (and conditional on the observables x and z but this is suppressed to save on notation) as:

$$
\begin{aligned}
E(q^E|\mathbf{1}_E = 1) &= \int q^E(\theta) f(\theta|\mathbf{1}_E = 1) d\theta = \int q^E(\theta) \frac{f(\theta, \mathbf{1}_E = 1)}{P(\mathbf{1}_E = 1)} d\theta = \\
&= \int q^E(\theta) \frac{f(\theta, \Delta \geq 0) + P_x f(\theta, \Delta < 0)}{P(\mathbf{1}_E = 1)} d\theta = \\
&= \frac{P(\Delta \geq 0)}{P(\mathbf{1}_E = 1)} \int q^E(\theta) \frac{f(\theta, \Delta \geq 0)}{P(\Delta \geq 0)} d\theta + \frac{P_x P(\Delta < 0)}{P(\mathbf{1}_E = 1)} \int q^E(\theta) \frac{f(\theta, \Delta < 0)}{P(\Delta < 0)} d\theta = \\
&= \frac{P(\Delta \geq 0)}{P(\mathbf{1}_E = 1)} \int q^E(\theta) f(\theta|\Delta \geq 0) d\theta + \frac{P_x P(\Delta < 0)}{P(\mathbf{1}_E = 1)} \int q^E(\theta) f(\theta|\Delta < 0) d\theta = \\
&= \frac{P(\Delta \geq 0)}{P(\mathbf{1}_E = 1)} E(q^E|\Delta \geq 0) + \frac{P_x P(\Delta < 0)}{P(\mathbf{1}_E = 1)} E(q^E|\Delta < 0)
\end{aligned}
$$

where the probability $P(\Delta \geq 0)$ was computed in Lemma 1 and $P(\Delta < 0) = 1 - P(\Delta \geq 0)$.

Call $a \equiv \frac{\ln A - \bar\theta}{\sigma}$, $b \equiv \frac{\ln B - \bar\theta}{\sigma}$ and $c \equiv \frac{\ln C - \bar\theta}{\sigma}$. We then have

$$
\begin{aligned}
E(q^E|\Delta < 0) &= E(q^E|\Delta < 0, \theta > B) P(\theta > B|\Delta < 0) + E(q^E|\Delta < 0, \theta \leq B) P(\theta \leq B|\Delta < 0) \\
&= E(q^E|\Delta < 0, \theta > B) \frac{P(\theta > B, \Delta < 0)}{P(\Delta < 0)} + E(q^E|\Delta < 0, \theta \leq B) \frac{P(\theta \leq B, \Delta < 0)}{P(\Delta < 0)} \\
&= E(q^E|\Delta < 0, \theta > B) \frac{P(\theta > B, \theta < C)}{P(\Delta < 0)} + E(q^E|\Delta < 0, \theta \leq B) \frac{P(\theta \leq B, \theta < A)}{P(\Delta < 0)} \\
&= E(q^E|\Delta < 0, \theta > B) \frac{(1 - \mathbf{1}_{B>C})(\Phi(c) - \Phi(b))}{P(\Delta < 0)} + E(q^E|\Delta < 0, \theta \leq B) \frac{\Phi(\min(a, b))}{P(\Delta < 0)}
\end{aligned}
$$

and

$$
\begin{aligned}
E(q^E|\Delta \geq 0) &= E(q^E|\Delta \geq 0, \theta > B) P(\theta > B|\Delta \geq 0) + E(q^E|\Delta \geq 0, \theta \leq B) P(\theta \leq B|\Delta \geq 0) \\
&= E(q^E|\Delta \geq 0, \theta > B) \frac{P(\theta > B, \Delta \geq 0)}{P(\Delta \geq 0)} + E(q^E|\Delta \geq 0, \theta \leq B) \frac{P(\theta \leq B, \Delta \geq 0)}{P(\Delta \geq 0)} \\
&= E(q^E|\Delta \geq 0, \theta > B) \frac{P(\theta > B, \theta \geq C)}{P(\Delta \geq 0)} + E(q^E|\Delta \geq 0, \theta \leq B) \frac{P(\theta \leq B, \theta \geq A)}{P(\Delta \geq 0)} \\
&= E(q^E|\Delta \geq 0, \theta > B) \frac{1 - \Phi(\max(b, c))}{P(\Delta \geq 0)} + E(q^E|\Delta \geq 0, \theta \leq B) \frac{\mathbf{1}_{B>A}(\Phi(b) - \Phi(a))}{P(\Delta \geq 0)}
\end{aligned}
$$

and where:

1. $E(q^E|\Delta < 0, \theta > B) = E((\lambda z)^\alpha \theta | B < \theta < C) = (\lambda z)^\alpha E(\theta|B < \theta < C) = (\lambda z)^\alpha E(\theta) \frac{\Phi(\sigma - b) - \Phi(\sigma - c)}{\Phi(c) - \Phi(b)}$

2. $E(q^E|\Delta < 0, \theta \leq B) = E((\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} \theta^{\frac{1}{1-\alpha}} | \theta \leq \min(A, B)) = (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} E(\theta^{\frac{1}{1-\alpha}}) \frac{\Phi(\min(a,b) - \sigma/(1-\alpha))}{\Phi(\min(a,b))}$,
   where $E(\theta^{\frac{1}{1-\alpha}}) = \exp(\frac{\bar\theta}{1-\alpha} + \frac{\sigma^2}{2(1-\alpha)^2})$.

3. $E(q^E|\Delta \geq 0, \theta > B) = E((\lambda z)^\alpha \theta | \theta \geq \max(B, C)) = (\lambda z)^\alpha E(\theta) \frac{\Phi(\sigma - \max(b,c))}{\Phi(-\max(b,c))}$.

4. $E(q^E | \triangle \geq 0, \theta \leq B) = E((\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} \theta^{\frac{1}{1-\alpha}} | A \leq \theta \leq B) = (\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} E(\theta^{\frac{1}{1-\alpha}}) \frac{\Phi(\sigma/(1-\alpha)-a) - \Phi(\sigma/(1-\alpha)-b)}{\Phi(b) - \Phi(a)}.$

Finally, since the unobserved component of talent ε is assumed independent of x, expected gross income conditional on non-entrepreneurship is simply

$$E(q^A | \mathbf{1}_E = 0) = \mu(1+x)^\gamma$$

Appendix C – Cost of entry into the labor market

Denoting by $\hat{\triangle}(z, \theta, x)$ the expression $y^E(\theta, z) - y^A(z, x) + c$, we obtain,

$$\hat{\triangle}(z, \theta, x) \geq 0 \Leftrightarrow \begin{cases} (1-\alpha)\theta^{\frac{1}{1-\alpha}}(\frac{\alpha}{r})^{\frac{\alpha}{1-\alpha}} - \mu(1+x)^\gamma + c \geq 0 & \text{if } \theta \leq B(z) \\ \theta(\lambda z)^\alpha - \lambda r z - \mu(1+x)^\gamma + c \geq 0 & \text{if } \theta > B(z) \end{cases}$$

which, in terms of the agent's entrepreneurial ability θ, is equivalent to,

$$\hat{\triangle}(z, \theta, x) \geq 0 \Leftrightarrow \begin{cases} \theta \geq \left(\frac{\mu(1+x)^\gamma - c}{1-\alpha}\right)^{1-\alpha}(\frac{r}{\alpha})^\alpha & \text{if } \theta \leq B(z) \\ \theta \geq (\lambda z)^{-\alpha}[\mu(1+x)^\gamma + r\lambda z - c] & \text{if } \theta > B(z) \end{cases}$$

Call $\hat{A}(x) \equiv \left(\frac{\mu(1+x)^\gamma - c}{1-\alpha}\right)^{1-\alpha}(\frac{r}{\alpha})^\alpha$ and $\hat{C}(z, x) \equiv (\lambda z)^{-\alpha}[\mu(1+x)^\gamma + r\lambda z - c]$.

The *probability of entrepreneurship* for a household with observables z and x is then,

$$\hat{P}_E(z, x) \equiv P(\mathbf{1}_E = 1) = P(\hat{\triangle}(z, \theta, x) \geq 0)$$

where the latter probability can be computed as in Lemma 1 by replacing A and C with $\hat{A}$ and $\hat{C}$ respectively.

Note that for $c = 0$ we have $\hat{\triangle}(z, \theta, x) = \triangle(z, \theta, x)$. The *probability of involuntary entrepreneurship*, $\hat{P}_I(z, x)$ is defined as the difference between the probability of entrepreneurship for given $c > 0$ and that for $c = 0$ (no entry cost), or

$$\hat{P}_I(z, x) \equiv P(\hat{\triangle}(z, \theta, x) \geq 0) - P(\triangle(z, \theta, x) \geq 0).$$

The *expected gross income* conditional on entrepreneurship can be derived in the same way as in Appendix B using that

$$E(q^E | \mathbf{1}_E = 1) = E(q^E | \hat{\triangle} \geq 0)$$

and replacing $\triangle$, A and C with $\hat{\triangle}$, $\hat{A}$ and $\hat{C}$ respectively in the relevant expressions.