Learning about Comparative Advantage in Entrepreneurship: Evidence from Thailand

JOB MARKET PAPER

Anant Nyshadham*†

January 2012

Abstract

Entrepreneurial activity has been argued to be an important stimulus of growth, especially in less developed countries. However, measuring the returns to entrepreneurship is made difficult due to potential selection on the basis of unobservable abilities and the agents' imperfect information about their own comparative advantage in entrepreneurship. I develop a novel extension to projection-based panel data methods to estimate a model of returns to entrepreneurship that accounts for heterogeneous abilities, learning, and financial constraints. My approach has two main advantages: 1) it is robust to arbitrary relationships between latent heterogeneity and sector and input choices; and 2) using this method, I can test between the full model and nested models featuring homogeneous returns and/or perfect information, allowing the data to determine the need for the additional complexity. I estimate the model using panel data from Thailand, and find strong evidence of selection into entrepreneurship on the basis of comparative advantage. The results show that the households with low ability in the default sector have high productivity gains from switching to entrepreneurship, and suggest that these households learn about their comparative advantage through lower-than-expected productivity in the default sector and/or positive productivity realizations in entrepreneurship. Importantly, the structural results do not suggest a salient role for savings or credit constraints in entrepreneurship decisions. I conduct additional analysis using agricultural output prices and their interactions with the household's farm acreage as instruments for savings and self-reported credit constraints and find no significant effects of these on entry into entrepreneurship.

*Yale University, Department of Economics; anant.nyshadham@yale.edu; http://nyshadham.squarespace.com/

†I would like to especially thank Mike Boozer for his invaluable guidance and support. I would also like to thank Ach Adhvaryu, Joe Altonji, Prashant Bharadwaj, A.V. Chari, Penny Goldberg, Fabian Lange, Costas Meghir, Chris Udry, and participants of the Yale Labor and Development Prospectus Workshops for their helpful comments. All errors are my own.

1 Introduction

The development literature has proposed several important drivers of growth: agricultural technology adoption (e.g. high-yielding varieties, fertilizer, mechanization, irrigation), human capital investment (e.g. schooling, healthcare, nutrition), financial deepening (e.g. access to credit, savings, insurance), etc. In addition, studies have argued that entrepreneurship is, potentially, an important stimulus of growth, particularly in contexts where agents engage in subsistence agricultural production by default (see, for example, Foster & Rosenzweig (2004)). Some studies have suggested that the poor participate in low-growth sectors (unskilled wage labor and subsistence agriculture instead of entrepreneurship and skilled labor) and invest less in productivity enhancing technologies (physical and human capital) due to low actual or perceived net returns.

Specifically, under this hypothesis, agents that choose not to switch to higher productivity technologies or switch into higher growth sectors fall into two general categories: low gross return and high cost. A low gross return could be the result of low ability in or preference for the sector or technology in question. On the other hand, high cost could be due to low access to the sector or technology or low access to the financial resources necessary for participation in said sector or adoption of said technology.

In particular, we might suspect that some agents have a comparative advantage in the technology or sector in question while others have a comparative advantage in the alternative. This heterogeneity in relative abilities will generate heterogeneous gross returns on which agents will base their decision of whether or not to adopt. The returns to schooling literature (e.g. Card (1995)), for example, has emphasized the role of heterogeneous ability in schooling choice. Some studies have shown that agents with low educational attainment might actually have low returns to education, and, therefore, their decisions to accumulate less schooling are optimal, perhaps both from an individual and societal welfare perspective.[1]

However, a model of economic decision-making with comparative advantage must also take

[1] See, for example, Nyshadham (2011).

a stance on whether agents have perfect information about their relative abilities. The literature on the adoption of agricultural technologies (e.g. Foster & Rosenzweig (1995), Conley & Udry (2010)) has emphasized learning about returns as an important determinant of the rate of adoption. Though the specific learning mechanism varies across these studies, they share the proposition that agents have imperfect initial beliefs about the returns to adoption, generating dynamics in the decision process.

Other literatures have focused on financial constraints as an obstacle to switching sectors or technologies. In particular, the entrepreneurship literature has emphasized the role of credit constraints in the entry decision, proposing that these constraints preclude many households, perhaps even the highest return households, from starting businesses. Several theoretical studies have explored the entrepreneurship decision under financial constraints. Banerjee and Newman (1993, 1994) develop a model of occupational choice, financial constraints and long-term growth which predicts that entrepreneurial decisions under financial constraints can lead to a poverty trap, in that, given sufficient wealth inequality, the poor will choose subsistence wage work due to a lack of access to the capital necessary for entrepreneurship.

Buera (2009) develops a dynamic model of entrepreneurship under financial constraints which suggests that the relationship between wealth and entrepreneurship is non-monotonic; that is, it is positive for low levels of wealth, but negative for high levels of wealth. A simulation of this model, calibrated to US data, suggests that credit constraints have larger effects on the intensive margin of scale than on the extensive margin of entry. Buera, Kaboski, and Shin (2011) develop a model which predicts that industries with higher capital intensities will suffer larger productivity losses due to capital misallocation than will industries with lower capital requirements. Taken together, the results of previous work suggest that a study of the decision to enter the entrepreneurial sector in developing contexts (where entrepreneurial activities tend to have relatively low capital requirements) ought to explore determinants of sectoral choice other than financial constraints.

Empirical studies in developed contexts have found mixed evidence of the relationship between wealth and entrepreneurship. Hurst and Lusardi (2004) find empirical evidence in the

US for a positive relationship at the highest levels of wealth and no significant relationship otherwise. Using data from the NLSY, Dunn and Holtz-Eakin (2000) find that own wealth has little effect on entrepreneurship, and parental wealth has, at most, weak effects. On the other hand, parental entrepreneurial ability and experience have stronger effects on the entrepreneurial activity of children.

The empirical evidence from the related development literature on the returns to improved access to financial resources amongst microenterprises has been mixed as well. In a field experiment in Sri Lanka, de Mel, McKenzie, and Woodruff (2008) find that the return to exogenously provided capital amongst microenterprises is higher than the market interest rate and that these returns are even larger for high ability entrepreneurs. Similarly, Burgess and Pande (2005) find that improved access to banking institutions (which, in theory, provide both credit and savings accounts) in rural India increased output in non-agricultural production. However, Dupas and Robinson (2011) provide experimental evidence from Kenya that access to institutional savings accounts alone can increase productive investments, independent of access to credit. Finally, Karlan and Zinman (2010) conduct an experiment in Manila which finds weak average effects of access to microcredit on profits and scale of production of microenterprises. In sum, these studies suggest that improved access to credit might be no more important a determinant of scale of entrepreneurial production than access to institutional savings accounts, and provide only limited information on the relative importance of entrepreneurial ability as a determinant of scale. Additionally, these studies do not provide evidence of effects along the extensive margin of entrepreneurial entry.

Recent experimental studies in the microenterprise literature have found mixed evidence of the effects of improved access to microcredit on entrepreneurial entry as well as effects on consumption of non-durable goods and investment in durable goods by entrepreneurial status. Crepon, Devoto, Duflo, and Pariente (2011) find in Morocco that improved access to credit increased the scale of existing farm and livestock enterprises, but had no effect on the creation of new enterprises. Additionally, amongst non-entrepreneurial households, improved access to credit increased food consumption and durable expenditures. Similarly, Banerjee, Duflo, Glennerster, and Kinnan (2010) find positive effects of access to credit on durable expenditure of entrepreneurial households and non-durable expenditure of non-entrepreneurial households. However, in this context, the results indicate large effects on entrepreneurial entry in the short-run. Nevertheless, on the same sample of households on which I conduct my analysis in this study, Kaboski and Townsend (2012), in a non-experimental evaluation of one of the largest scale government microfinance initiatives in the world ("Million Baht Village Fund" program), find no impact of improved access to credit on business starts.

Midrigan and Xu (2011) provide a possible explanation for these heterogeneous effects of credit on entrepreneurship. They propose a dynamic model of firm entry under capital constraints and present simulation results, alternately calibrating the model to plant-level data from Korea and Columbia. They show that welfare losses due to credit constraints are quite small because high ability firms can quickly save themselves out of these constraints. These results suggest that, perhaps, as in models of schooling choice and technology adoption, ability or comparative advantage might be the predominant determinant of the entrepreneurship decision.

None of these existing studies address the relative importance of credit and ability as drivers of movement along the extensive margin of entrepreneurial entry. The schooling literature (e.g. Carneiro and Heckman (2002)), for example, has shown that a model of heterogeneous ability can just as easily explain behavior once attributed to credit constraints, while the entrepreneurship literature is lacking such a comparison. Pauslon, Townsend, and Karaivanov (2006) present a model of entrepreneurship with, alternately, limited liability for borrowers and moral hazard, and structurally estimate this model using cross-sectional data from the Townsend Thai Project. Though the model that Paulson *et al.* present allows for heterogeneous entrepreneurial ability along with these financial market imperfections, their estimation results do not address the relative importance of ability and credit constraints in entrepreneurial choice.

In the empirical portion of this study, I also use data from the Townsend Thai Project, but exploit the panel structure of the survey. Preliminary graphical analysis of data from 2000 to 2009 shows little change in the percentage of households engaged in entrepreneurship amidst a strong upward trend in savings and downward trend in self-reported financial constraints. Though there is some evidence of a relationship between financial constraints and entrepreneurship earlier in the decade, the implementation from 2001 to 2002 of the "Million Baht Village Fund" program–which allocated a lump sum of a million baht to each of 77,000 villages across Thailand to be distributed amongst households in the form of microfinance loans–appears to have effectively relaxed these constraints by the second half of the decade. Additionally, as mentioned above, Kaboski and Townsend (2012) find no effects on business starts at the time of the program's implementation.

I conduct additional preliminary analysis of the role of financial constraints in the entrepreneurship decision. Using agricultural output prices and their interactions with the household's farm acreage as instruments for savings and self-reported credit constraints, I find no significant effects of these on entrepreneurial entry. This evidence leads me to question the importance of credit constraints in the entrepreneurial entry decision in this empirical context and to explore alternate determinants of entrepreneurship.

In this paper, I present a model which includes comparative advantage in entrepreneurship over the default sector (mostly, subsistence agriculture) and learning, as well as financial constraints. I then develop an estimation strategy, building on projection-based panel data methods (e.g. Chamberlain (1982, 1984), Islam (1995), Suri (2011)), that recovers consistent estimates of the average return to entrepreneurship in the presence of heterogeneity and learning. I can also estimate parameters that characterize the degree of heterogeneity and the relationship between entrepreneurial and non-entrepreneurial earnings, as well as the degree and direction of learning.

The empirical results provide strong evidence that households select into entrepreneurship on the basis of comparative advantage. Specifically, households with high returns to entrepreneurship (and hence, those households that ultimately select into entrepreneurship) have low counterfactual earnings in the default sector. The estimates of the learning parameters, though imprecise, provide evidence that households do, in fact, learn about their comparative advantage in entrepreneurship over time. In particular, the results suggest that households learn from lower-than-expected productivity in the default sector and/or high productivity realizations in entrepreneurship that they have a comparative advantage in entrepreneurship and accordingly choose to switch into or stay in the entrepreneurial sector.

Nested models that restrict returns to be homogeneous, both with and without learning, are rejected easily in this empirical context. Neither the static nor the dynamic model with heterogeneous returns can be rejected; however, the estimate of the parameter measuring the degree of heterogeneity is only statistically significant in the full model with learning. I interpret these results as validating the need for the additional complexity in the preferred model in order to fully explain the observed patterns of sectoral choices and incomes in the data.

In contrast, the structural results do not support an important role for savings or credit constraints in the entrepreneurship decision process, as suggested by the preliminary analysis. That is, a model in which households endogenously ease their heterogeneous financial constraints to entrepreneurship through savings also predicts dynamic, heterogeneous returns to switching sectors, but is distinguishable from the preferred model of learning about comparative advantage in its predictions of switching trajectories following particular productivity shocks. Specifically, under the model of heterogeneous financial constraints and endogenous savings, constrained households should be *more* likely to switch into the entrepreneurial sector following the arrival of a positive productivity shock. On the other hand, under the preferred model of learning about comparative advantage in entrepreneurship, households should be *less* likely to switch sectors following a positive productivity shock, having updated their beliefs about their relative ability in the farm sector. The latter is represented more heavily in the data.

This paper makes two main contributions to the literature. This is the first paper, to my knowledge, to explore the entrepreneurship decision in a model which allows for learning about comparative advantage. Second, I make a methodological contribution with an extension to projection-based panel methods which allows for the estimation of a dynamic correlated random coefficients (DCRC) model. This approach has two main advantages: 1) it is robust to arbitrary relationships between latent heterogeneity and sector and input choices; and 2) using this method, I can test between the full model and nested models featuring homogeneous returns and/or perfect information, allowing the data to determine the need for the additional complexity.

I provide evidence that households on the margin are more likely constrained by low ability rather than high costs. That is, it seems that the scarce resource hindering entrepreneurial entry in this setting is not credit, but rather entrepreneurial talent. This is particularly informative for policy-makers in developing contexts like the one investigated in this study. Specifically, this study has two important policy implications: 1) investing in additional financial resources in contexts like the one studied here will not likely have the strong effects on entrepreneurial activity proposed by previous studies, nor will such an endeavor be necessarily welfare enhancing, and 2) allocating resources towards training programs to improve skills necessary for successful entrepreneurship (e.g. bookkeeping, communications and marketing, supply and inventory management, etc.) could have a large impact on sectoral choice.

The remainder of this paper is organized as follows: section 2 presents the model, section 3 develops the estimation strategy, section 4 presents the nested models, section 5 discusses the data and descriptive evidence in support of the approach, section 6 reports and discusses the results, and section 7 concludes.

2 Model

2.1 Production Functions

Let us consider a model of household production with two possible technologies. Following Evans and Jovanovic (1989) and Paulson, Townsend, and Karaivanov (2006), one of the technologies or sectors will represent an entrepreneurial endeavor taking capital as an input, while the other will represent default production. In Thailand, the context in which the empirical analysis in this study is conducted, the default sector is mostly subsistence agriculture (and to some degree wage labor as well in urban regions). Accordingly, default production will also take capital input.

Gross output of a household operating in the default sector is given by the following production function:

$$Y_{it}^F = e^{\beta_t^F} K_{iFt}^{\rho^F} e^{\eta_i^F}, \tag{1}$$

where β_t^F is the average productivity on the farm (or in wage labor), K_{iFt} is capital input in farm production, and η_i^F is the heterogeneous component of farm-specific productivity. On the other hand, gross output of a household operating in the entrepreneurial sector is given by:

$$Y_{it}^E = e^{\beta_t^E} K_{iEt}^{\rho^E} e^{\eta_i^E}, \tag{2}$$

where β_t^E is the average productivity in entrepreneurial activities, K_{iEt} is capital input under entrepreneurship, and η_i^E is the heterogeneous component of productivity in entrepreneurial activities.

2.2 Sector and Input Decisions

I will assume that the price of capital is r and that households face no cost of capital adjustment. I will first consider the case in which households are unconstrained with respect to capital. That is, households can acquire as much capital as desired at the given price r. Then, the household's input allocation decision in each sector can be represented as the solution to the following maximization problem:

$$\max_{K_{ijt}} \Big[e^{\beta_t^j} K_{ijt}^{\rho^j} e^{\eta_i^j} - rK_{ijt} \Big], \quad j \in \{E, F\} \tag{3}$$

The household's optimal capital input level in sector j, as a function of of η_i^j, is

$$K_{ijt}^* = \left(\frac{\rho^j}{r} e^{\beta_t^j + \eta_i^j} \right)^{\frac{1}{1-\rho^j}} \tag{4}$$

Then, household i will choose to produce in the entrepreneurial sector in period t (i.e. $D_{it} = 1$) if $[y_{it}^E(K_{iEt}^*) - rK_{iEt}^*] - [y_{it}^F(K_{iFt}^*) - rK_{iFt}^*] > 0$, and in the default farm sector otherwise. For the sake of analytical and expositional simplicity, I will assume that $\rho^E \approx \rho^F \equiv \rho$.[2] Using (1) and (2) and substituting in for optimal capital input in each sector using (4), I derive a cutoff rule for entrepreneurship. Household i will choose to produce in the entrepreneurial sector if and only if

$$e^{(\eta_i^E - \eta_i^F)} > e^{(\beta_t^F - \beta_t^E)} \tag{5}$$

Note that sectoral choice depends on the relationship between β_t^E and β_t^F, but more importantly, on the relationship between η_i^E and η_i^F as well. That is, the household's relative ability in entrepreneurship, $(\eta_i^E - \eta_i^F)$, will drive the entry decision.

2.3 Comparative Advantage

Since with 2 sectors only the relative magnitude of η_i^F and η_i^E can be identified, I will define, following Lemieux (1993, 1998) and Suri (2011)[3], η_i^F and η_i^E in terms of the household's relative productivity in entrepreneurship over default farm activity $(\eta_i^E - \eta_i^F)$ using the following projections:

$$\eta_i^F = b_F(\eta_i^E - \eta_i^F) + \tau_i \tag{6}$$

$$\eta_i^E = b_E(\eta_i^E - \eta_i^F) + \tau_i \quad , \tag{7}$$

where $b_E = (\sigma_E^2 - \sigma_{EF})/(\sigma_E^2 + \sigma_F^2 - 2\sigma_{EF})$, $b_F = (\sigma_{EF} - \sigma_F^2)/(\sigma_E^2 + \sigma_F^2 - 2\sigma_{EF})$, with $\sigma_{EF} \equiv Cov(\eta_i^E, \eta_i^F)$, $\sigma_E^2 \equiv Var(\eta_i^E)$, and $\sigma_F^2 \equiv Var(\eta_i^F)$. The household's *absolute advantage* is represented by τ_i; that is, τ_i has the same effect on the household's productivity in both sectors and, accordingly, does not affect the sectoral choice.

The household-specific output gain in entrepreneurship over default production can be re-

[2]Relaxing this assumption will not substantively change the interpretation of the estimation results below.
[3]The original exposition of this model of self-selection on comparative advantage can be found in Roy (1951).

defined to be entrepreneurial *comparative advantage*, η_i, as

$$\eta_i \equiv b_F(\eta_i^E - \eta_i^F). \tag{8}$$

Defining $\phi \equiv b_E/b_F - 1$ and using equations (6) and (7), I can express the heterogeneous components of sector-specific productivities in terms of absolute advantage and entrepreneurial comparative advantage :

$$\begin{aligned} \eta_i^F &= \eta_i + \tau_i & (9) \\ \eta_i^E &= (1+\phi)\eta_i + \tau_i \quad . & (10) \end{aligned}$$

Taking logs of production functions (1) and (2) and substituting in (9) and (10), I get

$$\begin{aligned} y_{it}^F &= \beta_t^F + \rho k_{it}^F + \eta_i + \tau_i & (11) \\ y_{it}^E &= \beta_t^E + \rho k_{it}^E + (1+\phi)\eta_i + \tau_i. & (12) \end{aligned}$$

Defining D_{it} as a dummy for entrepreneurship which takes value $D_{it} = 1$ if household i owns a business in period t and $D_{it} = 0$ otherwise, I can write a generalized, log-linear gross output equation:

$$\begin{aligned} y_{it} &= D_{it} \;\; \beta_t^E + \rho k_{it}^E + (1+\phi)\eta_i + \tau_i \;\; + (1 - D_{it}) \;\; \beta_t^F + \rho k_{it}^F + \eta_i + \tau_i \\ &= \beta_t^F + (\beta_t^E - \beta_t^F)D_{it} + \rho[k_{it}^F + (k_{it}^E - k_{it}^F)D_{it}] + \eta_i(1 + \phi D_{it}) + \tau_i & (13) \end{aligned}$$

2.4 Learning

I will now amend the model to allow for imperfect information. In particular, I will assume that households know β_t^F, β_t^E, ρ, τ_i and ϕ, but have imperfect information about η_i. In particular, I will introduce an additive productivity shock, ε_{it}, to η_i in equation (13) and assume that $\varepsilon_{it} \sim$

$N(0, \sigma_\varepsilon^2 = 1/h_\varepsilon)$. The generalized log-linear production function then becomes:

$$y_{it} \;=\; \beta_t^F + (\beta_t^E - \beta_t^F)D_{it} + \rho[k_{it}^F + (k_{it}^E - k_{it}^F)D_{it}] + (\eta_i + \varepsilon_{it})(1 + \phi D_{it}) + \tau_i \tag{14}$$

Households hold the initial belief that $\eta_i \sim N(m_{i0}, \sigma^2 = 1/h)$; and this belief is refined each period using output observations, y_{it}. That is, from y_{it}, households can compute

$$l_{it} = \frac{y_{it} - \beta_t^F - (\beta_t^E - \beta_t^F)D_{it} - \rho^F k_{it}^F + (\rho^E k_{it}^E - \rho^F k_{it}^F)D_{it} - \tau_i}{(1 + \phi D_{it})} = \eta_i + \varepsilon_{it}, \tag{15}$$

a noisy signal of their entrepreneurial comparative advantage, η_i, which is independent of the their period t sectoral choice.

Let $l_i^t = (l_{i1}, ..., l_{it})$ denote the history of household i's normalized entrepreneurial comparative advantage observations through period t. Then, the posterior distribution of η_i given history l_i^t is distributed $N(m_t(l_i^t), 1/h_t)$, where

$$m_t(l_i^t) = \frac{hm_{i0} + h_\varepsilon(l_{i1} + ... + l_{it})}{h + th_\varepsilon}, \quad \text{and} \quad h_t = h + th_\varepsilon \tag{16}$$

Note that the specific learning mechanism proposed here allows households to learn about returns to entrepreneurship each period, irrespective of the sector in which the household is producing in that period.[4] The intuition behind this proposed mechanism is that comparative advantage, η_i, is an index of fundamental skills which affect productivity in both sectors, but are valued differentially across the two sectors (e.g. marketing, supply chain management, etc.). Assuming that the household knows ϕ but not η_i corresponds to assuming the household how much each sector values these skills but not their own skill stock. Accordingly, households can learn about their stock through production in either sector.

For example, suppose that η_i represents the household's ratio of physical strength to marketing ability. The agricultural output market is heavily commoditized, and, therefore, the sale

[4] This learning structure is borrowed from Gibbons, Katz, Lemieux, Parent (2005) who use it to study learning about comparative advantage in a model of occupational choice. They, in turn, borrow heavily from the classic development in DeGroot (1970). Please see these previous works for more in depth discussion of this framework.

of agricultural goods requires less marketing and salesmanship. Of course, in contrast, physical strength might be an important determinant of output in agricultural production. On the other hand, the entrepreneurial sector corresponds to the household's running a noodle shop.[5] Accordingly, entrepreneurial earnings are increasing in the household's ability to draw customers, but do not depend heavily on the household's physical strength.

The assumptions of the model imply that the household recognizes that entrepreneurship rewards marketing ability more than does agricultural production and that agricultural production strongly rewards physical strength; however, the household is unsure of its specific ratio of physical strength to marketing ability. Of course, an excellent salesman might still be able to earn more in the agricultural sector than someone who is bad at sales. Therefore, a household that initially believes it is bad at marketing will operate in the agricultural sector to start; however, should this household find this period that it is better able to market its agricultural goods than it expected, it will decide to open a restaurant next period, knowing that the restaurant business is very lucrative for a household with strong marketing ability. The mechanism, of course, works in the opposite direction as well.

In particular, this deviates from an experimentation or learning by doing framework. In such a framework, the household knows its marketing ability, but is unsure of how much the entrepreneurial sector rewards this ability. Given that the technologies employed in the entrepreneurial sector (e.g. restaurant, mechanic shop, barber shop, trading, etc.) are not especially new to the region, I find the learning by doing framework to be less appropriate in this context. The technology adoption literature that has emphasized the learning by doing mechanism has generally focused on the introduction of new technologies. In these contexts, it is less reasonable to assume that households know which skills the new technology or sector will favor. Nevertheless, though I will not discuss alternatives in this study, the empirical strategy developed below will be generally robust to other learning mechanisms.[6]

Lastly, I should note that, to the degree that both sectors reward some skills (e.g. work ethic)

[5]Indeed, restaurants make up a large fraction of household businesses in the sample.

[6]See Bharadwaj, Nyshadham, & Suri (2011) for a comparison across various learning mechanisms using a similar empirical approach to the one developed here.

equally, these skills are represented by τ_i and will affect the levels of earnings for the household in both sectors, but will not affect the return to switching sectors.

2.4.1 Learning: Sector and Input Decisions

The household's period t sectoral choice and corresponding input decision will be different in the model with learning about comparative advantage. This will be of particular importance for the empirical strategy below. I discuss these differences here.

The timing of decisions is as follows:

1. household i chooses its production technology and the corresponding optimal level of capital input at the beginning of period t using its current expectation of its comparative advantage in entrepreneurship $m_{i,t-1} \equiv m_{t-1}(l_i^{t-1})$
2. household i engages in production during period t and observes y_{it}
3. at the end of period t, household i calculates l_{it} as in (15) and updates its expectation of η_i according to (16).

Then, the household's input allocation decision in each sector can now be represented as the solution to the following amended maximization problem:

$$\max_{K_{ijt}} E_t \; e^{\beta_t^j} K_{ijt}^{\rho} e^{\eta_i(1+\phi D_{it})+\tau_i} - rK_{ijt} \qquad j \in \{E, F\} \tag{17}$$

where the expectation is with respect to the agent's information at the beginning of period t. The household's optimal capital input level in each sector will be a function of $E_t[\eta_i] = m_{i,t-1}$:[7]

$$K_{iEt}^* = \kappa\Big(m_{i,t-1}, \sigma_t^2; \phi\Big) \tag{18}$$

$$K_{iFt}^* = \kappa\Big(m_{i,t-1}, \sigma_t^2\Big), \tag{19}$$

[7] I omit the full equations representing these choices for simplicity of exposition. Though the full equations are reported in Appendix B, it should be noted that the exact functional form of the choices will not be used in the estimation strategy below. Rather the estimation strategy is only concerned with the dependence of choices on $m_{i,t-1}$ and the evolution of $m_{i,t-1}$ over time.

where $\sigma_t^2 = (h + th_\varepsilon)/\{h_\varepsilon[h + (t-1)h_\varepsilon\}$ is the variance of the prior distribution at the beginning of period t.[8]

Remembering that $(\eta_i^E - \eta_i^F) \equiv \phi\eta_i$ and that $E_t[\exp\{\phi\eta_i\}] = \exp\{\phi m_{i,t-1} + (1/2)\phi^2\sigma_t^2\}$, the cutoff rule determining sectoral choice from equation (5) becomes: $D_{it} = 1$ if and only if

$$\exp\{\phi m_{i,t-1}\} > \exp\{-(\beta_t^E - \beta_t^F) - (1/2)\phi^2\sigma_t^2\} \tag{20}$$

The sectoral choice now depends on $m_{i,t-1}$ and ϕ. Though the discussion of the sectoral choice here regards a comparison of levels of profits, the empirical strategy developed below will estimate a generalized log production function in order to recover estimates of the structural parameters of interest. If I, accordingly, take logs of both sides of equation (20), I can comment on how the model predicts sectoral choices and incomes will change with the evolution of $m_{i,t-1}$:

$$\begin{aligned} m_{i,t-1} &> \frac{-(\beta_t^E - \beta_t^F) - (1/2)\phi^2\sigma_t^2}{\phi}, \quad \text{if} \quad \phi > 0 \\ m_{i,t-1} &< \frac{-(\beta_t^E - \beta_t^F) - (1/2)\phi^2\sigma_t^2}{\phi}, \quad \text{if} \quad \phi < 0 \end{aligned} \tag{21}$$

Note that the sign of ϕ will determine which direction of evolution in $m_{i,t-1}$ will drive switching in and out of entrepreneurship and the effect of this evolution of $m_{i,t-1}$ on income. In particular, if $\phi < 0$ I should expect that a *downward* evolution in $m_{i,t-1}$ will decrease non-entrepreneurial earnings, increase returns to entrepreneurship, and, accordingly drive households to switch in or stay in the entrepreneurship sector, while an *upward* evolution will have the opposite effects on non-entrepreneurial earnings and the returns to entrepreneurship and will drive households to switch out or stay out of entrepreneurship.

[8] Note that if $\ln\nu$ is normally distributed with mean μ and variance σ^2, then $E(e^\nu) = e^{\mu+(1/2)\sigma^2}$.

2.5 Limited Liability and Capital Constraints

To address the emphasis placed on credit constraints in the existing entrepreneurship literature, I will now introduce one form of implied capital constraints through limited liability borrowing and discuss the implications for input and sector decisions. Of course, other forms of financial constraints (e.g. moral hazard, as in Paulson, Townsend, and Karaivanov (2006)) could be at play in this context. Nevertheless, the point of this section is mostly to illustrate the ability of the empirical strategy proposed below to deal with input restrictions more generally. I will reserve the discussion of robustness to alternate forms of financial constraints for the empirical strategy section below.

Following Paulson, Townsend, and Karaivanov (2006), suppose now that when a household borrows capital, it has the opportunity to default. That is, a household that has chosen to participate in sector j allocates $(A_{it} + K_{ijt})$ as capital input into the selected production technology, where A_{it} is the household's available savings and K_{ijt} is additional capital that is borrowed (or lent). I will at first assume A_{it} is exogenously given, and later discuss what happens to sector and input choices when A_{it} is endogenized. If the household chooses to repay the loan, it receives

$$\begin{aligned} D_{it} &= 1: \qquad e^{\beta_t^E} K_{iEt}^{\rho} e^{(1+\phi)(\eta_i+\varepsilon_{it})+\tau_i} + r(A_{it} - K_{iEt}) \\ D_{it} &= 0: \qquad e^{\beta_t^F} K_{iFt}^{\rho} e^{\eta_i+\varepsilon_{it}+\tau_i} + r(A_{it} - K_{iFt}) \end{aligned} \tag{22}$$

If the household chooses to default, it receives

$$\begin{aligned} D_{it} &= 1: \qquad e^{\beta_t^E} K_{iEt}^{\rho} e^{(1+\phi)(\eta_i+\varepsilon_{it})+\tau_i} + \pi A_{it} \\ D_{it} &= 0: \qquad e^{\beta_t^F} K_{iFt}^{\rho} e^{\eta_i+\varepsilon_{it}+\tau_i} + \pi A_{it} \end{aligned} \tag{23}$$

where π is the fraction of assets A_{it} that the household must forfeit as collateral for the defaulted loan.[9]

[9]Note that because the shock, ε_{it}, affects payoffs in both repayment and default states symmetrically, the default decision will not depend on this period's realization of ε_{it}. Therefore, there will be no default in equilibrium.

Then, in equilibrium, a household can only borrow

$$K_{ijt} \leq \quad 1 + \frac{\pi}{r} \quad A_{it}, \tag{24}$$

where $j \in \{E, F\}$. Then, we have that K^*_{iEt} and K^*_{iFt} are given by (18) or (19), respectively, when the credit constraint is not binding. Note that K^*_{ijt} does not depend on assets, A_{it}, in (18) and (19). On the other hand, if

$$m_{i,t-1} > \quad \ln \; (\lambda A_{it})^{1-\rho}\frac{r}{\rho} \quad - \beta^F_t - (\beta^E_t - \beta^F_t)D_{it} - \tau_i - (1/2)(1+\phi D_{it})^2\sigma^2_t \quad \frac{1}{1+\phi D_{it}} \tag{25}$$

where $\lambda \equiv \quad 1 + \frac{\pi}{r} \quad$, then the constraint binds and $K^*_{ijt} = \lambda A_{it}$. That is, the lender will only lend up to λA_{it} in equilibrium due to the risk of default.

Now, with limited liability borrowing, the optimal capital choice, and likely the sector choice as well, will depend on assets A_{it}, the current expectation of comparative advantage, $m_{i,t-1}$, and whether or not the household's credit constraint binds, which itself depends on A_{it} and $m_{i,t-1}$. I will discuss the implications of credit constraints, with alternately exogenous and endogenous assets, for the estimation in section 3.2 below, but will begin by developing the method for the unconstrained case.

3 Estimation

Redefining coefficients in equation (14), I arrive at the estimating equation:

$$y_{it} = \alpha_t + \beta_t D_{it} + \rho k_{it} + (\eta_i + \varepsilon_{it})(1 + \phi D_{it}) + \tau_i + \zeta_{it} \tag{26}$$

where $\alpha_t \equiv \beta^F_t$, $\beta_t \equiv (\beta^E_t - \beta^F_t)$, $k_{it} \equiv k^F_{it} + (k^E_{it} - k^F_{it})D_{it}$, and measurement error ζ_{it} is assumed mean independent of sector and input decisions conditional on η_i and τ_i. That is, in particular, I will assume $E(D_{it}|\zeta_{it}, \eta_i, \tau_i) = E(D_{it}|\eta_i, \tau_i)$ and $E(k_{it}|\zeta_{it}, \eta_i, \tau_i) = E(k_{it}|\eta_i, \tau_i)$. Also, for the sake of parsimony, I will also assume $\beta_t = \beta \; \forall t$.[10]

As discussed above, both D_{it} and k_{it} will depend on the mean of the household's prior distribution on η_i coming into period t, $m_{i,t-1}$, which I cannot observe. Accordingly, OLS estimates of β and ρ will be biased. I now develop a strategy, building on Chamberlain (1982, 1984), Islam (1995) and Suri (2011), which allows me to consistently estimate β and ρ, recover ϕ, and validate the importance of learning in this empirical context.

In particular, in order to recover consistent estimates of β and ρ, I must purge the composite unobserved term, $(\eta_i + \varepsilon_{it})(1 + \phi D_{it}) + \tau_i + \zeta_{it}$, of its correlation with D_{it} and k_{it}. For ease of exposition, I will ignore capital choices for the time being and deal only with the endogeneity in sectoral choice. The method will be extended to allow for endogenous capital choices as well in section 3.2 below. I know from section 2.4 that the portion of $(\eta_i + \varepsilon_{it})$ which correlates with sectoral choices is $m_{i,t-1}$. I will begin by decomposing $m_{i,t-1}$ into two components which have distinct effects on the household's history of sectoral choices. Note that the Bayesian updating of beliefs implies that the mean of the prior distribution is a martingale. That is, the law of motion for $m_{i,t}$ is

$$m_{i,t} = m_{i,t-1} + \xi_{it} \quad \Rightarrow \quad m_{i,t-1} = m_{i0} + \sum_{k=1}^{t-1} \xi_{ik}, \tag{27}$$

where ξ_{it} is a noise term orthogonal to $m_{i,t-1}$. Then, denoting $m_i^{t-1} \equiv \sum_{k=1}^{t-1} \xi_{ik}$ as the sum of the signals received up to period $t-1$, I have

$$y_{it} = \alpha_t + \beta_t D_{it} + (m_{i0} + m_i^{t-1} + \varphi_{it})(1 + \phi D_{it}) + v_{it}, \tag{28}$$

where $v_{it} \equiv \tau_i + \zeta_{it}$ is orthogonal to sectoral choice in period t, D_{it}, by construction and $\varphi_{it} \equiv \eta_i + \varepsilon_{it} - (m_{i0} + m_i^{t-1})$ is orthogonal to D_{it} by nature of the martingale structure of $m_{i,t-1}$.

Extending the approaches developed in Chamberlain (1982, 1984), Islam (1995), and Suri (2011), we can overcome the endogeneity of D_{it} by projecting m_{i0} and m_i^{t-1} onto the history

[10] Relaxing this assumption does not significantly change the empirical results.

of sectoral choices. In particular, the law of motion of the prior, as expressed in equation (27), suggests that the initial belief, m_{i0}, will affect sectoral choices in all periods. On the other hand, the cumulative update, m_i^{t-1}, will only affect sectoral choices in period t onwards.

I will set $T = 2$ in the estimation below.[11] In the 2 period case, I have a projection of the initial belief which appears in the estimating equation for both periods and a belief update projection which appears only in the period 2 estimating equation: [12]

$$
\begin{aligned}
m_{i0} &= \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \psi_{i0} && (29) \\
m_i^1 &= \theta_0 + \theta_2 D_{i2} + \psi_{i1} && (30)
\end{aligned}
$$

Note that the martingale structure of the prior on η_i implies that learning is *efficient*; that is, all information the household will use to make its decision at time t is fully summarized in the initial condition m_{i0} and the sum of the orthogonal updates to period $t-1$, m_i^{t-1}. In other words, the path by which the prior reaches $m_{i,t-1}$ will not, conditional on $m_{i,t-1}$ itself, affect sectoral choice in period t, D_{it}. Most importantly, the path by which the sum of the updates reaches m_i^{t-1} will not, conditional on both the initial belief m_{i0} and m_i^{t-1} itself, affect D_{it}. Therefore, I need not include past sectoral choices in the update projection in equation (30) nor the interactions of future sectoral choices (though in a 2 period estimation, this is irrelevant).

Plugging projections (29) and (30) into equation (28), and grouping terms, I get the following log gross output equations:

$$
\begin{aligned}
y_{i1} &= \alpha_1 + \lambda_0 + D_{i1} \; \beta + (1+\phi)\lambda_1 + \phi\lambda_0 \; + D_{i2} \; \lambda_2 \; + D_{i1} D_{i2} \; (1+\phi)\lambda_{12} + \phi\lambda_2 \\
&\quad + (1+\phi D_{i1})(\varphi_{i1} + \psi_{i0}) + \upsilon_{i1} && (31)
\end{aligned}
$$

$$
y_{i2} = \alpha_2 + \lambda_0 + \theta_0 + D_{i1} \; \lambda_1 \; + D_{i2} \; \beta + (1+\phi)(\lambda_2 + \theta_2) + \phi(\lambda_0 + \theta_0)
$$

[11] In the Appendix I explore an estimation in 3 periods because the learning structure is better defined than in the 2 period case, but must adopt a more restrictive treatment of capital than the one shown below due to the analytical complexity.

[12] Note that beliefs at the start of period 1 consist only of the initial condition m_{i0} and, therefore, sectoral choice in period 1 will be made only on the basis of this initial belief

$$+ D_{i1}D_{i2}\ \ (1+\phi)\lambda_{12}+\phi\lambda_1\ \ + (1+\phi D_{i2})(\varphi_{i2}+\psi_{i0}+\psi_{i1})+\upsilon_{i2} \tag{32}$$

where ψ_{i0} and ψ_{i1} are the portions of m_{i0} and m_i^{t-1}, respectively, that are orthogonal to sectoral choices in all periods by construction of the projections. The important point here is that I must properly specify projections (29) and (30) (that is, I must include all necessary elements of the history of productive decisions) in order to ensure that the projection errors ψ_{i0} and ψ_{i1} are, indeed, orthogonal to current choices. Note that the estimating equation is a generalized, log linear structural production function and, accordingly, once I express the unobservable components in terms of all observable choices which depend on these unobservable components, I have estimable equations expressing income in each period entirely in terms of observables.

Specifically, we have the following corresponding reduced form regressions:

$$\ln w_{i1} = \delta_1+\gamma_1 D_{i1}+\gamma_2 D_{i2}+\gamma_3 D_{i1}D_{i2}+\nu_{i1} \tag{33}$$

$$\ln w_{i2} = \delta_2+\gamma_4 D_{i1}+\gamma_5 D_{i2}+\gamma_6 D_{i1}D_{i2}+\nu_{i2} \tag{34}$$

Following Chamberlain (1982, 1984) and Suri (2011), my empirical strategy will be to first estimate the reduced form coefficients $\{\gamma_j : j \in [1,...,6]\}$ by seemingly unrelated regressions (SUR) and then to estimate from these coefficients the structural parameters of the model using minimum distance. There are 6 structural parameters of the model, $\{\lambda_1,\lambda_2,\lambda_{12};\theta_2;\beta;\phi\}$, to be identified from the 6 reduced form coefficients using the minimum distance restrictions implied by the model. The minimum distance restrictions are

$$\begin{aligned}
\gamma_1 &= \beta+(1+\phi)\lambda_1+\phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_{12}+\phi\lambda_2 \\
\gamma_4 &= \lambda_1 \\
\gamma_5 &= \beta+(1+\phi)(\lambda_2+\theta_2)+\phi(\lambda_0+\theta_0)
\end{aligned}$$

$$\gamma_6 \quad = \quad (1+\phi)\lambda_{12} + \phi\lambda_1 \tag{35}$$

It appears from (35) that there are 8 structural parameters to be estimated. However, I will impose the following normalizations:

$$\lambda_0 \quad = \quad -\lambda_1\overline{D_{i1}} - \lambda_2\overline{D_{i2}} - \lambda_{12}\overline{D_{i1}D_{i2}} \tag{36}$$

$$\theta_0 \quad = \quad -\theta_2\overline{D_{i2}} \quad , \tag{37}$$

where $\overline{D_{ij}}$ is the average entrepreneurship decision in period j and $\overline{D_{i1}D_{i2}}$ is the average of the interaction between the entrepreneurship decisions in periods 1 and 2.

Because this model is just-identified, I cannot jointly test the restrictions imposed by this model using an over-identification test. However, in the extension discussed below, which incorporates endogenous capital choices along with the endogenous sectoral choices, the model is over-identified and can, accordingly, be tested.

Note that I have not included any exogenous covariates here. In theory, ν_{it} could include, along with v_{it}, any exogenous covariates from (28). Though the inclusion of exogenous covariates will affect reduced form expressions (33) and (34), it will not affect the relationships between the reduced form coefficients on the choices and the structural parameters of interest. Nevertheless, as I am estimating a log-linearized production function, I do not believe any additional covariates are appropriate with the exception of inputs, which are endogenous as shown above. I reserve the discussion of the treatment of endogenous inputs for section 3.2.

3.1 Structural Interpretation of Projection Coefficients

I observe in the data the conditional sample mean of log gross output for each entrepreneurship history in each period (i.e. $E(y_{it}|D_{i1}, D_{i2})$). I can express these conditional moments in two ways: 1) in terms of the estimated parameters $\{\lambda_1, \lambda_2, \lambda_{12}; \theta_2; \beta; \phi\}$, and 2) in terms of the underlying components of the model $E(m_{i0}|D_{i1}, D_{i2})$, $E(m_i^1|D_{i1}, D_{i2})$, and, of course, β and ϕ. Comparing these two sets of expressions, I can derive structural interpretations for the estimated projection coefficients.

The interpretations for the coefficients from the initial belief projection are given by:

$$\lambda_1 = E[m_{i0}|D_{i1}=1, D_{i2}=0] - E[m_{i0}|D_{i1}=0, D_{i2}=0]; \quad (38)$$

$$\lambda_2 = E[m_{i0}|D_{i1}=0, D_{i2}=1] - E[m_{i0}|D_{i1}=0, D_{i2}=0]; \quad (39)$$

$$\begin{aligned} \lambda_{12} &= \Big\{E[m_{i0}|D_{i1}=1, D_{i2}=1] - E[m_{i0}|D_{i1}=1, D_{i2}=0]\Big\} \\ &\quad - \Big\{E[m_{i0}|D_{i1}=0, D_{i2}=1] - E[m_{i0}|D_{i1}=0, D_{i2}=0]\Big\} \end{aligned} \quad (40)$$

The interpretation for the coefficient from the belief update projection is given by:

$$\begin{aligned} \theta_2 &= E[m_i^1|D_{i1}=1, D_{i2}=1] - E[m_i^1|D_{i1}=1, D_{i2}=0] \\ &= E[m_i^1|D_{i1}=0, D_{i2}=1] - E[m_i^1|D_{i1}=0, D_{i2}=0] \end{aligned} \quad (41)$$

These expressions suggest that if $\theta_2 < 0$, then households that switch into entrepreneurship, or do not switch out, experience relatively lower earnings in the non-entrepreneurial sector than those that switch out or stay out. If I also have that $\phi < 0$, then those households that experience negative shocks in the non-entrepreneurial sector and, subsequently, switch into entrepreneurship have *larger* returns to entrepreneurship than those that do not receive these negative updates and, therefore, choose to stay in the non-entrepreneurial sector. That is, entrepreneurial households select into entrepreneurship on the basis of their comparative advantage in entrepreneurship, and households with the high returns to entrepreneurship have low non-entrepreneurial earnings.

3.2 Endogenous Inputs

As shown in section (2.5), a household's optimal capital allocation and sectoral choice in the presence of credit constraints will depend on its level of savings, A_{it}, and its current expectation of its comparative advantage, $m_{i,t-1}$. Indeed, even in the unconstrained case, optimal input and sectoral choices depend on $m_{i,t-1}$.

Note that, as the estimating equation (26) corresponds to a generalized production function, A_{it} has no place in this equation. That is, A_{it} has no effect on gross earnings except through its effect on inputs and subsequent sectoral choices when the credit constraint binds. Therefore, I am not concerned with any portion of A_{it} that is not captured in the observed k_{it} and D_{it}.

Certainly, an endogenous determination of A_{it} will indeed generate a dependence between A_{it} and the unobservable $m_{i,t-1}$, and therefore, alter the functional form of the relationship between k_{it} and $m_{i,t-1}$. However, because I do not rely on the specific functional form of this relationship, but rather simply the notion that k_{it} depends on $m_{i,t-1}$ and that $m_{i,t-1}$ evolves in a particular way, the estimation strategy will be unaffected by the endogenous accumulation of savings.

While A_{it} has no effect on gross earnings except through its effect on input and sector choices, $m_{i,t-1}$ has a direct effect on y_{it} by definition. Reintroducing capital into equation (28), I get

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + (m_{i0} + m_i^{t-1})(1 + \phi D_{it}) + \varphi_{it}(1 + \phi D_{it}) + v_{it}, \tag{42}$$

where v_{it} and φ_{it} are orthogonal to input decision k_{it} in period t, along with D_{it}.

Therefore, I must concern myself with the correlation between k_{it} (and, of course, D_{it}) and $m_{i0}+m_i^{t-1}$. Now, following the approach presented above, in order to purge the composite error of its correlation with both D_{it} and k_{it}, I must include in the projections of m_{i0} and m_i^{t-1} not only the history of sectoral choices and, when appropriate, the interactions of sectoral choices across time, but also the history of input choices and its interaction with the history of sectoral choices. The new projections are

$$\begin{aligned} m_{i0} &= \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \lambda_{k1-1} k_{i1} D_{i1} + \lambda_{k1-2} k_{i1} D_{i2} \\ &\quad + \lambda_{k1-12} k_{i1} D_{i1} D_{i2} + \lambda_{k2-1} k_{i2} D_{i1} + \lambda_{k2-2} k_{i2} D_{i2} + \lambda_{k2-12} k_{i2} D_{i1} D_{i2} + \psi_{i0} \end{aligned} \tag{43}$$

$$m_i^1 = \theta_0 + \theta_2 D_{i2} + \theta_{k2} k_{i2} + \theta_{k2-2} k_{i2} D_{i2} + \psi_{i1} \tag{44}$$

I then proceed as above by substituting these new projections into equation (42) to get reduced form estimating equations similar to equations (33) and (34), but now including capital from each year and their interactions with the history of sectoral choices. The corresponding reduced form regressions are

$$\begin{aligned} \ln w_{i1} &= \delta_1 + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i1} D_{i2} + \gamma_4 k_{i1} + \gamma_5 k_{i2} + \gamma_6 k_{i1} D_{i1} + \gamma_7 k_{i1} D_{i2} \\ &\quad + \gamma_8 k_{i1} D_{i1} D_{i2} + \gamma_9 k_{i2} D_{i1} + \gamma_{10} k_{i2} D_{i2} + \gamma_{11} k_{i2} D_{i1} D_{i2} + \nu_{i1} \qquad (45) \\ \ln w_{i2} &= \delta_2 + \gamma_{12} D_{i1} + \gamma_{13} D_{i2} + \gamma_{14} D_{i1} D_{i2} + \gamma_{15} k_{i1} + \gamma_{16} k_{i2} + \gamma_{17} k_{i1} D_{i1} + \gamma_{18} k_{i1} D_{i2} \\ &\quad + \gamma_{19} k_{i1} D_{i1} D_{i2} + \gamma_{20} k_{i2} D_{i1} + \gamma_{21} k_{i2} D_{i2} + \gamma_{22} k_{i2} D_{i1} D_{i2} + \nu_{i2} \qquad (46) \end{aligned}$$

As above, I estimate the reduced form coefficients $\{\gamma_j : j \in [1, ..., 22]\}$ by SUR and then estimate from these coefficients the structural parameters of the model. There are 17 structural parameters of the model,

$$\{\lambda_1, \lambda_2, \lambda_{12}, \lambda_{k1}, \lambda_{k2}, \lambda_{k1-1}, \lambda_{k1-2}, \lambda_{k1-12}, \lambda_{k2-1}, \lambda_{k2-2}, \lambda_{k2-12}; \theta_2, \theta_{k2}, \theta_{k2-2}; \rho, \beta; \phi\},$$

to be identified from the 22 reduced form coefficients using MD estimation with the restrictions implied by the model. The minimum distance restrictions from this model are presented in the Appendix. This model is, therefore, well over-identified and the restrictions implied by the model can be jointly tested. The over-identification test statistic under optimal minimum distance estimation (OMD) equals the minimized value of the objective function and is distributed χ^2 with 5 degrees of freedom.

3.3 Threats to Identification

In this section, I reiterate the identifying assumptions set forth above and discuss circumstances under which they might be violated.

1. Sequential Exogeneity - the current period's shock to productivity is mean zero, conditional on the prior at the beginning of period. If households can predict future productiv-

ity shocks and respond to them in their sector and input decisions, the update projection, as specified, will not fully account for the endogeneity in these choices.

2. Properly Specified Production Function - all of the household's production decisions which depend on the unobservable are included in the structural production function and appropriately represented in the projections. If the household makes additional production decisions which are not observed, the projections, as specified, will not be complete and the resulting projection errors will not necessarily be orthogonal to the regressors of interest. The most obvious example of such a scenario is if the household allocates labor hours across the two sectors and leisure each period. Labor is discussed in section 5.3 below.

4 Nested Models

In this section, I show how the basic framework presented in section 3 above nests restricted models of heterogeneous returns to entrepreneurship with perfect information, homogeneous returns with imperfect information, and a simple fixed effects model with homogeneous returns and perfect information. For each of the nested models, I will start by amending the estimating equation (26) to reflect the particular set of restrictions imposed and, then, redefine the belief projections, estimating equations, and implied minimum distance restrictions, accordingly.

4.1 Heterogeneous Returns with Perfect Information: Correlated Random Coefficients

In the static correlated random coefficients (CRC) model, the estimating equation is nearly the same as in the full model:

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i(1 + \phi D_{it}) + v_{it} \tag{47}$$

However, now the household is assumed to have perfect information about its entrepreneurial comparative advantage, η_i; hence, there is no longer an additive productivity shock, ε_{it}. Therefore, the relationship between η_i and the history of sectoral choices is static. Note, however, that v_{it} could still include exogenous, transitory shocks that shift households from period to period above and below the cutoff for entrepreneurial entry. That is, households will sort into a particular entrepreneurship history on the basis of η_i and their expectations of y_{it}^F and y_{it}^E; however, these expectations will not evolve over time as they do in the imperfect information case. Accordingly, I need only a single projection in which I project η_i onto the entrepreneurship decisions in both periods, their interaction, the capital choices in both periods, and the interaction of capital in both periods with entrepreneurship decisions and their interaction:

$$\begin{aligned} \eta_i \quad = \quad & \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \lambda_{k1-1} k_{i1} D_{i1} + \lambda_{k1-2} k_{i1} D_{i2} \\ & + \lambda_{k1-12} k_{i1} D_{i1} D_{i2} + \lambda_{k2-1} k_{i2} D_{i1} + \lambda_{k2-2} k_{i2} D_{i2} + \lambda_{k2-12} k_{i2} D_{i1} D_{i2} + \psi_{i0} \end{aligned} \tag{48}$$

Substituting (48) into (47), I get equations for log gross output in terms of entrepreneurship in all periods and the interactions of these choices. Then, the corresponding reduced form equations are identical to those from the full model presented in equations (33) and (34); however, the minimum distance restrictions imposed by this model are different than those imposed by the full model. Under this model, I will estimate only 14 structural parameters

$$\{\lambda_1, \lambda_2, \lambda_{12}, \lambda_{k1}, \lambda_{k2}, \lambda_{k1-1}, \lambda_{k1-2}, \lambda_{k1-12}, \lambda_{k2-1}, \lambda_{k2-2}, \lambda_{k2-12}; \rho, \beta; \phi\}$$

from the 22 reduced form coefficients.

This nested model imposes 3 additional restrictions on the full model, namely

$$\theta_2 = \theta_{k2} = \theta_{k2-2} = 0 \tag{49}$$

The over-identification test statistic for this model corresponds to a joint test of the same restrictions imposed in the full model along with the additional restrictions in (49). That is, if I find that I can reject the full set of restrictions imposed by this static CRC model, but cannot reject a joint test of the restrictions imposed in the preferred dynamic CRC model, I can conclude that the additional restrictions in (49) are violated. As mentioned above, the test statistic is equal to the minimized value of the criterion function, but is now distributed χ^2 with 8 degrees of freedom.

4.2 Homogeneous Returns with Imperfect Information: Dynamic Correlated Random Effects

In a dynamic correlated random effects model (DCRE), the household is assumed, as in the preferred model, to have imperfect information about η_i; however, now η_i has the same effect on earnings in both sectors. In particular, the estimating equation becomes

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i + \varepsilon_{it} + v_{it}, \tag{50}$$

where η_i is now the household's fixed effect, which is known by the household (though still unobserved by the econometrician). Note that β is in essence the population mean of the distribution of η_i. Accordingly, this model could alternately be interpreted as one in which household's learn about the average return to entrepreneurship, β.

The household's current expectation of η_i can, once again, be split into two parts: the initial belief, m_{i0}, and the sum of the innovations to date, m_i^{t-1}. I can proceed, as above, by projecting m_{i0} onto entrepreneurship and input choices in all periods, and m_i^{t-1} onto choices in period t and all future choices:

$$m_{i0} = \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \psi_{i0} \tag{51}$$

$$m_i^1 = \theta_0 + \theta_2 D_{i2} + \theta_{k2} k_{i2} + \psi_{i1} \tag{52}$$

Notice now that even in the initial belief projection (51), I have not included the interactions of entrepreneurship decisions across periods nor have I included interactions between sector and input choices. This is because, once I assume that η_i has no effect on the return to entrepreneurship, the changes in choices over time will no longer depend on the initial belief, though the choice in each period still will.

Therefore, the projections imply the following simplified reduced form equations:

$$y_{i1} = \delta_1 + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 k_{i1} + \gamma_4 k_{i2} + \nu_{i1} \tag{53}$$

$$y_{i2} = \delta_2 + \gamma_5 D_{i1} + \gamma_6 D_{i2} + \gamma_7 k_{i1} + \gamma_8 k_{i2} + \nu_{i2} \tag{54}$$

However, in the spirit of econometrically testing between the nested models, I will use the full reduced form equations implied by the most general model and test the restrictions that the reduced form coefficients which appear in equations (45) and (46) from the full model, but not in equations (53) and (54) are zero. Therefore, from the 22 reduced form coefficients, I will estimate 8 structural parameters. That is, this model imposes 9 additional restrictions on the preferred model:

$$\lambda_{12} = \lambda_{k1-1} = \lambda_{k1-2} = \lambda_{k1-12} = \lambda_{k2-1} = \lambda_{k2-2} = \lambda_{k2-12} = \theta_{k2-2} = \phi = 0 \tag{55}$$

Accordingly, I need only estimate $\{\lambda_1, \lambda_2, \lambda_{k1}, \lambda_{k2}; \theta_2, \theta_{k2}; \rho, \beta\}$.

Once again, the over-identification test statistic for this model corresponds to a joint test of the same restrictions imposed in the full model along with the additional restrictions in (55). The test statistic for this model is distributed χ^2 with 14 degrees of freedom.

4.3 Homogeneous Returns with Perfect Information: Correlated Random Effects

The most restricted model imposes both that returns to entrepreneurship are homogeneous and that households have perfect information about their earnings in both sectors. That is, the only source of heterogeneity is additive and fixed over time. This amounts to assuming that the data generating process is a simple household fixed effects model. Under these assumptions, the estimating equation becomes

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i + \upsilon_{it} \tag{56}$$

I now need only a single projection of η_i on the entrepreneurship decisions and input choices from all periods:

$$\eta_i = \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \psi_{i0} \tag{57}$$

As in the DCRE case above, I need not include the interactions of these decisions with each other nor across periods.

This model imposes 11 additional restrictions on my preferred model:

$$\lambda_{12} = \lambda_{k1-1} = \lambda_{k1-2} = \lambda_{k1-12} = \lambda_{k2-1} = \lambda_{k2-2} = \lambda_{k2-12} = \theta_2 = \theta_{k2} = \theta_{k2-2} = \phi = 0 \tag{58}$$

Notice that the set of additional restrictions in (58) includes the additional restrictions from both the static CRC model, (49), and the DCRE model, (55). I will estimate 6 structural parameters $\{\lambda_1, \lambda_2, \lambda_{k1}, \lambda_{k2}; \rho, \beta\}$ from the 22 reduced form coefficients. The over-identification test from this estimation is distributed χ^2 with 16 degrees of freedom.

Using the over-identification tests on all of the nested models, I can explore the degree to which the added complexity in the preferred model (non-additive heterogeneity in returns and a relaxation of strict exogeneity to sequential exogeneity) is important in describing the relationship between income and entrepreneurship in the data. This is a major advantage to the theoretical and, particularly, the empirical approach I employ in this study.

5 Data

The data set used in the analysis is taken from the annual panel of the Townsend Thai Project. In 1997, the original survey was conducted on households from 4 provinces of Thailand. Two provinces were chosen from each of two distinct regions: the more developed Central Region and the more rural Northeast. Within each of the four provinces, 12 sub-regions (tambons) were randomly selected. Within each tambon, 4 villages were randomly selected for the sample.

From each of the 4 provinces, 4 of the original 12 tambons were randomly selected for annual resurvey. Consequently, of the original 48 tambons, 16 (4 from each province) are included in the 12 year annual household and business panel from which I will extract the data to be used in the empirical analysis. From 1999 onwards, questions regarding household businesses were added to the household survey instrument. I will construct a balanced panel using data from the 2005 and 2008 waves. In particular, I will use all households for which income and entrepreneurship information is available in both years. The sample I use consists of 1103 households.

The 3 year gap between survey waves ensures that households have sufficient time to adjust entrepreneurial activity, should they want to. Among the 1103 households in my sample, over 25% change their entrepreneurship status between 2005 and 2008. However, the proportion of households participating in the entrepreneurial sector is roughly stable across waves: 44% in 2005 and 47% in 2008.

The survey instrument includes questions regarding income over the 12 months prior to survey from farm and livestock activities, wage or salary work, household businesses, and other income such as rent and interest, as well as questions regarding input expenditure in farm and business enterprises. Information on savings, borrowing and lending, and participation in financial institutions was also collected. Finally, households were asked if they believed their farms and/or businesses would be more profitable if they were expanded, a measure of their being credit constrained.

5.1 Summary Statistics

In Tables Ia-c, I report means and standard deviations for variables of interest in the data. Table Ia presents summary statistics for the entire sample of log of gross income, entrepreneurship, input expenditure, household demographics, savings, self-reported credit constraints, and borrowing. I find that income grows only slightly in the sample from 2005 to 2008. However, the percentage of households with savings (a positive balance in an institutional savings account) grows considerably and the percentage of households that report being credit constrained drops to nearly 0. On the other hand, the probability that a household borrowed money in the past year and the probability that the household owns at least one business remain fairly stable.

In Tables Ib and Ic, I report summary statistics for the variables of interest by entrepreneurship history. Specifically, I split up the sample into households that engage in entrepreneurship in both years, in neither of the years, those that switch into entrepreneurship in 2008, and those that switch out in 2008. Note that these categories are strictly mutually exclusive. I will note first that, though it appears that the percentage of households that engage in entrepreneurship remains roughly the same each year, there is quite a bit of switching in and out of entrepreneurship. As mentioned above, roughly 25% of the sample switches their entrepreneurial status. In this sample, approximately 11% switch out and 14% switch in.

Table Ib also shows that households that run businesses tend to have similar gross incomes to those that don't; although households that never own a business have slightly lower incomes and households that own a business in both periods have slightly higher incomes. I also find that expenditure is higher among entrepreneurial households. Households that engage in entrepreneurship tend to be larger than those that do not; however, no perceivable differences exist between specific entrepreneurship histories. No significant difference exists in gender composition of households across entrepreneurship histories. Entrepreneurial households appear to be slightly younger on average and better educated than non-entrepreneurial households.

In Table Ic, I find that households that engage in entrepreneurship are more likely to have savings than those who do not. However, if we look over time at households that switch in, for example, it would appear that, if in fact there is a relationship, savings accrue contemporaneously with entrepreneurship, or even following it, rather than savings driving the entrepreneurship decision. Entrepreneurial households are actually more likely to report feeling financially constrained than non-entrepreneurial households in 2005, but the probability of reporting constraints goes to roughly 0 in 2008 for all entrepreneurship histories. Finally, entrepreneurial households are more likely to borrow than non-entrepreneurial households, but within entrepreneurship comparisons provide mixed evidence on the relationship.

5.2 Preliminary Analysis

As mentioned above in the discussion of Table Ia, the summary statistics for savings, self-reported financial constraints, and entrepreneurship show a pattern inconsistent with the notion that credit constraints are an important determinant of entrepreneurial entry. Using data from all waves between 2000 and 2009, I plot these variables to verify the trends suggested by the summary statistics. In particular, in Figure I, I plot the means for savings, self-reported credit constraints, and entrepreneurship in each of the 10 waves as well as a fitted curve over time for each of the variables. Focusing on the years between 2005 and 2008 (marked by red vertical lines), I find that the trends suggested by the summary statistics in Table Ia are present in this higher-frequency panel as well. In particular, savings increases and credit constraints decrease, but entrepreneurship stays fairly flat over these years.

Note that Figure I shows some evidence of a relationship between these variables earlier in the decade. In particular, there is a contemporaneous rise in savings and entrepreneurship and a decline in financial constraints in 2001 (also marked by a red vertical line). This change corresponds to the initial implementation of Thailand's "Million Baht Village Fund" program which allocated a lump sum of a million baht to each of 77,000 villages to be distributed amongst households in the form of microfinance loans. The program was rolled out rapidly from 2001 to 2002. Note that savings and entrepreneurship continue to rise through 2002, 2003 and 2004, while self-reported constraints fall dramatically. By 2005, less than 20% of households report financial constraints. From 2005 onwards, participation in the entrepreneurial sector remains stable at roughly 45 %; the percentage of household with positive savings plateaus at above 80%; and the percentage of households reporting financial constraints continues to fall to nearly 0 in 2009.

As mentioned above, I use data from the 2005 and 2008 waves of the survey in the structural estimation below. This data was collected several years after the implementation of Thailand's "Million Baht Village Fund" program. Though, Kaboski and Townsend (2012) find significant short-term effects of the program on consumption, investment, savings, and income growth, they find no significant effects on entrepreneurial entry. Furthermore, the dramatic decrease in self-reported financial constraints and increase in savings following program implementation, along with the lack of change in entrepreneurship later in the decade, as depicted in Figure I, suggest that the "Million Baht Village Fund" program significantly diminished any role of financial constraints in the entrepreneurship decision for the years studied in this paper.

To explore this notion a bit further, I estimate the effects of variation in the global price of rice, the predominant agricultural output of Thailand, on savings, self-reported constraints, and entrepreneurship.[13] For this analysis, I once again use data from all waves between 2000 and 2009 in order to allow for greater variation in the price of rice. These regressions are run using household fixed effects specifications and the results are reported in Table II. In columns 4-6 of Table II, I report results from the regression of savings, self-reported constraints, and the household business dummies, respectively, on the global price of rice and household fixed effects. In columns 1-3, I report results from specifications which also include the household's farm acreage (in Thai rai units [14]) and its interaction with the global price of rice. Across both sets of regressions, I find that output price shocks increase savings and decrease financial constraints, but do not significantly affect entrepreneurship.

Finally, I use the price of rice and its interaction with household farm rai to instrument for the savings and constrained dummies in a household fixed effects instrumental variables regression of entrepreneurship on savings and constraints, alternately. The results from these regressions are reported in Table III. Once again, I find no evidence of an effect of savings and/or financial constraints on entrepreneurship. Taken together, Figure I and the results shown in Tables II and III suggests that perhaps access to credit is not an important determinant of entrepreneurship decisions, and provides motivation for the exploration of alternate drivers of entrepreneurial entry such as latent ability.

The model presented above proposes that evolution in households' beliefs about their entrepreneurial comparative advantage drives them to switch in and out of entrepreneurship.

[13]Price data is taken from the IMF monthly agricultural commodity prices and averaged over the year.
[14]1 acre equals roughly 2.5 rai

Specifically, the model imposes that intertemporal correlations in choices are due to households learning about their static, but unknown comparative advantage. Nevertheless, a model with persistent, sector-specific i.i.d. shocks, but perfect information could also explain switching in the absence of any learning mechanism. In order, to motivate the imposition of a learning structure on the dynamic nature of latent heterogeneity, I explore trends in switching.

In particular, a learning mechanism would predict a downward trend in switching as the beliefs of a cohort of households converge to the true values of their comparative advantages, while persistent sector-specific shocks should generate a consistent level of switching in all periods, so long as the distribution from which these shocks are drawn is stationary. In Figure II, I plot the percentage of households that switch, either into or out of entrepreneurship, over time. I also reproduce the plot of entrepreneurship percentages over time for the sake of comparison. Indeed, the percentage of households that switch their entrepreneurial status from the last period is decreasing over time. Given the evidence from this preliminary analysis, I suspect that the model proposed above and estimated below is appropriate for this context.

5.3 Labor

Note that the model presented in section 2 does not explicitly address the role of labor in the production technologies. This is because the data used in the estimation do not include information on time use or allocations of labor across sectors. In particular, though the survey asks the primary occupation of each member of the household, it does not collect labor hours. Therefore, I cannot observe the inframarginal allocations of hours across household farm activities, unpaid labor in household business, wage or salary work, and leisure. Nevertheless, under some assumptions, the omission of the household's labor decisions does not affect the empirical analysis.

In particular, if there is no market for entrepreneurial labor, leisure is not valued, and the demographic composition of the household is either fixed over time or subject to only unpredictable, exogenous shocks, then labor supply is given by the number and demographic characteristics of members of the household and is supplied inelastically across the two sectors in a fixed ratio to capital. In this case, the productivity of the household's labor endowment will represent portions of the household's η_i and τ_i. Specifically, to the degree that labor is equally valued across sectors, the labor endowment of the household will represent one aspect of the household's absolute advantage, τ_i, while any dimension of the labor endowment that is differentially valued across sectors will contribute to the household's comparative advantage, η_i.

I first explore the appropriateness of the labor market assumptions in this context. In Table IV, I present summary statistics of the percentage of households with business owners, unpaid family workers, and wage employees in each sector as members. I find, in the top panel of Table IV, that participation in the industries which make up the entrepreneurial sector (fish and shrimp farming, raising livestock, shopkeeping, and trading) is made up largely of business ownership and unpaid family labor, with limited wage employment. On the other hand, participation in default sector industries (farming, construction, factory work, janitorial service, and high skilled labor such as medicine, teaching, and financial services) is more balanced, with farm participation favoring ownership and household labor and both low and high skilled labor naturally favoring market employment. This evidence supports the notion that, though some labor markets exist in this context, there is, at best, an imperfect market for entrepreneurial labor.

Table V explores whether household labor endowments are fixed over time. Specifically, I report summary statistics of binary variables for whether the household's size, number of males, number of primary educated members, and number of members without a primary occupation change over time. I find that both the household's size and demographic composition change over time. The pressing question then becomes whether these changes are unpredictable, exogenous shocks to the household's labor endowment or endogenous decisions of the household to improve productivity.

To explore this notion, I present in Table VI OLS and household FE regressions of entrepreneurship on household size, and number of males, primary educated members, and non-working members in the household. In column 1, results from the OLS regression of entrepreneurship on household size and demographics suggest that the demographic composition of the household does, in fact, affect entrepreneurship decisions in the cross-section. This evidence supports the notion that the number of primary educated members of the household and the number of non-working members make up a portion of the household's comparative advantage η_i. On the other hand, household size does not effect the entrepreneurship decision, suggesting that the size of the labor endowment of the household is, perhaps, equally valued across sectors and, therefore, reflected in τ_i.

In column 2 of Table VI, I present results from the household FE regression of entrepreneurship on household size and demographics. The coefficients in this specification are identified off of changes in the regressors of interest within a household. I find no evidence of a strong partial correlation between household size or demographic composition and the entrepreneurship decision. Point estimates are small with tight standard errors. These results provide strong evidence in support of the notion that changes in size and composition of the labor endowment of the household do *not* reflect endogenous decisions on the part of the household. That is, if the household were endogenously changing the size or composition of its household in order to improve its productivity in one of the sectors or if sectoral choices were responding to predictable shocks to household composition, these changes in the size and composition of the household ought to correlate with entrepreneurship decisions.

The evidence discussed here alleviates to some degree concerns about the omission of labor allocation decisions of the household from the estimation. Nevertheless, this omission is certainly a short-coming of the empirical analysis in this study. Furthermore, I have, unfortunately, no way of testing, using this data, for the valuation of leisure in this context. The inelastic supply of the household's entire labor endowment is a necessary assumption for the exogeneity of labor.

6 Results

In this section, I present results from the empirical analysis discussed in section 3. However, for the sake of comparison, I begin by presenting ordinary least squares and household fixed effects estimates of the average return to entrepreneurship.

6.1 OLS and FE

In Table VII, I regress the log of total gross income of the household over the 12 months prior to survey on a binary for whether the household owned at least one business during that year. The results reported in column 3 of Table VII are from the specification with no additional covariates. The point estimate is quite large, positive, and significant at the 1 percent level. A unit change in the probability of a household owning a business is associated with a 64.6 percent increase in the household's income. In column 2, I include log input expenditure as a control and rerun the analysis. The inclusion of inputs significantly attenuates the estimate. The point estimate of the effect of entrepreneurship on log gross income is now 24.5 percent, but is still significant at the 1 percent level. In column 3, I also include village by time dummies to control for variations in input and output prices over time. That is, assuming that all households within a village face the same prices in each period, including these dummies accounts for the effects of these input and output prices on the household's choices and incomes. With these additional covariates, the point estimate rises slightly to 30.7 percentage points and is still significant at the 1 percent level.

In columns 4-6 of Table VII, I present results from specifications identical to those in columns 1-3, respectively, but with the addition of household fixed effects. The coefficients across all specifications are smaller in magnitude than the corresponding OLS estimates. In these FE specifications, I find that that village x time price controls have little effect on the coefficient of interest as compared to that from the specification including only inputs and the household fixed effects. However, as in the OLS specifications, the inclusion of log input expenditure decreases the magnitude of the effect of entrepreneurship on log gross income. In columns 4 and 5, I find that owning a household business is associated with a 17.8 and 19.4 percent increase in income, respectively, and these estimates are significant at the 5 percent level.

Of course, as discussed above, to the degree that households choose to engage in entrepreneurship on the basis of their perceived comparative advantage in it over farm production or wage work, the estimate of the coefficient of interest in OLS and FE specifications will be biased for the average return to entrepreneurship. The estimation strategy proposed and discussed above will allow me to recover a consistent estimate of the average return to entrepreneurship in the presence of both heterogeneous entrepreneurial abilities and learning as well as account for financial constraints, and will also allow me to quantify the degree of heterogeneity and learning in the data.

6.2 Reduced Form Coefficients

In Tables VIII and IX, I present the reduced form coefficients from which I will estimate the structural parameters of the econometric models set forth above using minimum distance. In the reduced form specifications reported in Table VIII, I regress the log of total gross income from each period on the entrepreneurship dummies for each period and their interactions. In Table IX, I report reduced form coefficients corresponding to the models with endogenous capital. As shown above, these specifications include, in addition to the history of entrepreneurship decisions, log input expenditure in both periods and their interactions with the history of entrepreneurship decisions. The reduced form coefficients are not particularly informative; accordingly, I will not provide a discussion of their interpretation here. Also, for the sake of brevity, I do not report reduced form coefficients corresponding to the specifications which include price controls.[15]

6.3 Structural Minimum Distance Estimates

6.3.1 No Covariates

In Table X, I present the optimal minimum distance estimates from the full CRC model with learning and the three nested, restricted models with no additional covariates. I present results from the CRE model in column 1. As mentioned above, the CRE model corresponds to a household fixed effects data generating process, that is, a model with homogeneous returns to entrepreneurship and perfect information. In particular, under this model latent ability has no effect on returns to entrepreneurship and the household's perception of this ability does not

[15]Reduced form results for other specifications are available upon request.

change over time.

Therefore, λ_j measures the correlation of the household's entrepreneurial choice in period j with the fixed effect. We need only one such parameter per period. The estimates of the λ's are both positive and precisely estimated. I will reserve, for the sake of brevity, the discussion of the interpretation of the projection coefficients in the context of the model for later, when I present results from the preferred model. The estimate of the average return to entrepreneurship, β, is also positive and very precisely estimated. The point estimate is .3044, which is quite similar to the results from the FE regression reported in column 6 of Table VII. The restrictions implied by this model (namely, no heterogeneity in returns and no learning) cannot be rejected. However, this test has limited power due to the minimal degrees of freedom. Furthermore, the analogous test on the full model cannot be applied because that model is just-identified in the case of no covariates. Accordingly, I will not place much stock in these statistics, and will reserve the discussion of a comparison of the models for the case of endogenous capital below.

Column 2 of Table X reports results from the dynamic CRE model which, once again, restricts returns to be homogeneous, but now allows for households to have imperfect information about this return. In the context of this model, the λ's characterize the heterogeneity in initial beliefs of households by entrepreneurship decisions, whereas the θ characterizes the degree and direction of learning (i.e. the heterogeneity in the update to beliefs between periods 1 and 2). The estimate of β is nearly identical to that in the static CRE model. The λ's are also nearly identical to those in the CRE model, while the θ is small and insignificant. The learning structure does not seem to improve the fit of the model, though this is likely due to the limited scope for learning in a 2 period model without endogenous capital.

In column 3 of Table X, I present estimates from the static CRC model which allows for heterogeneous returns but again restricts information on entrepreneurial comparative advantage to be perfect. This model implies that latent heterogeneity will not only affect entrepreneurship decisions in each period, but also the specific history of choices across periods. Therefore, I have now 3 λ's corresponding to 4 possible entrepreneurship histories over the two periods, with the omitted history being never owning a business. Once again, I find that the λ's are precisely estimated. The estimate of β is once again positive and slightly larger in magnitude than in the homogeneous returns models, though less precisely estimated. The estimate of ϕ, which measures the degree to which households base their entrepreneurial decisions on their comparative advantage in entrepreneurship, is negative but insignificant. A negative estimate of ϕ implies that households with low non-entrepreneurial earnings have high returns to entrepreneurship; however, the coefficient is not statistically significant from 0 so I will not dwell on its interpretation here.

Finally, in column 4 of Table X, I present estimates of the full model which allows for both selection on entrepreneurial comparative advantage and imperfect information. The estimates closely resemble those from the static CRC model, with the addition of a small and insignificant estimate of the θ_2. Of course, as shown in sections 2.4.1 and 2.5, the estimation should account for endogeneity in capital allocations as well as in sectoral choices. I discuss next results from the estimation of models with endogenous capital choices.

6.3.2 Endogenous Capital

In Table XI, I present results from all four models with the addition of endogenous capital as discussed in section 3.2. Once again, column 1 displays results from the minimum distance estimation of the static CRE model. There are now 4 λ's (i.e. 1 additional for each of the input choices). The estimates of λ_1 and λ_{k2} are positive and significant at the 1 percent level, while the estimates of λ_2 and λ_{k1} are small and insignificant. The estimate of the average return to capital, ρ, is positive and significant at the 1 percent level with a point estimate of nearly .06. The estimate of β, though still positive and precisely estimated, drops in magnitude from the no covariates case to a point estimate .1858. The estimates of ρ and β are quite similar to the results from the household FE regressions presented in column 5 of Table VII, as expected. This model, unlike in the no covariates case, is well over-identified. The χ^2 test statistic corresponding to a joint test of the restrictions imposed by this simplest model is just over 85 with a p-value of less than 0.0001. I can easily reject this model in this empirical context.

In column 2, I present results from the dynamic CRE model with endogenous capital. Specifically, this model introduces two new parameters θ_2 and θ_{k2}, corresponding to the entrepreneurial decision and capital choice in period 2, respectively. The estimates of these parameters are small and insignificant as in the case of no covariates. The estimates of the λ's are quite similar to those from the static CRE model. Though the estimates of ρ and β are qualitatively similar to those in column 1, the point estimate of ρ is slightly larger (0.0638) and that of β is smaller (0.1633). This model is also easily rejected with a χ^2 test statistic of roughly 84 and a corresponding p-value of less than 0.0001.

Column 3 displays results from the static CRC model. This model includes a total of 11 λ's corresponding to the history of entrepreneurial choices, the history of capital inputs, and their interactions. The estimate of ρ is qualitatively similar to those from the homogeneous returns model, with a slightly larger point estimate (0.0671) that is still significant at the 1 percent level. The point estimate of β is larger in this model, with a point estimate of 0.2191, and is significant at the 1 percent level. The estimate of ϕ is, as in the analogous no covariates model, negative but insignificant at conventional levels. The χ^2 test statistic of this model is just under 15 with a corresponding p-value of 0.061. This model, though still weakly rejected at the 10 percent level, appears to the explain the data much better than do the homogeneous returns models presented in columns 1 and 2.

Finally, in column 4, I present results from the estimation of the most general model allowing for both heterogeneous returns and learning. The estimates of the λ's are qualitatively similar to those in column 3, as are the estimates of ρ and β; however, the magnitudes of the estimates of both ρ and β are larger than those from the static CRC model with point estimates of 0.0726 and 0.2408, respectively. Both estimates are still significant at the 1 percent level. The point estimate of ϕ in column 4 is large, negative and significant at the 5 percent level with a point estimate of -0.4614. As mentioned above a significant ϕ is evidence of selection into entrepreneurship on the basis of heterogeneous returns. In particular, a negative ϕ corresponds to selection on comparative advantage such that households with low non-entrepreneurial earnings have high returns to entrepreneurship and, accordingly, are the households that choose to engage in entrepreneurship.

The estimate of the θ_2, though insignificant at conventional levels, is larger than in the no covariates case and still negative. The estimate of θ_{k2} is also negative and insignificant. I cautiously interpret these negative θ's as suggestive evidence of learning about comparative advantage through negative productivity shock realizations in the default sector and/or positive shock realizations in the entrepreneurial sector, as discussed in section 3.1. Given that the estimate of the ϕ is only significant with inclusion of these learning parameters, I believe that the learning structure is, indeed, important for explaining household behavior in the data. The imprecision in the estimates of the θ's is likely, at least in part, due to the limited scope afforded the learning structure in a two period estimation. Nevertheless, the full model is also rejected with a test statistic of 13 and a p-value of 0.022.

In the Appendix, I explore an extension of the estimation to a 3 period model. Due to the analytical complexity of fully endogenizing both entrepreneurship decisions and capital allocations in 3 periods, I employ a more restrictive treatment of capital in order to estimate these models. The results from the 3 period estimation is qualitatively similar to the results from the 2 period estimation discussed here; however, the magnitudes of the estimates are generally much larger and the estimates of the learning parameters are negative, large and significant.

6.3.3 Endogenous Capital with Price Controls

Lastly, in Table XII, I present results from the estimation of all four models with endogenous capital, as in Table XI, but now with the inclusion of village by time dummies as exogenous covariates. To the degree that input and output prices vary at the village level, the inclusion of village by time dummies in the first stage reduced form equations will purge the structural estimates of the effects of general, non-linear trends in these prices. Across all four models, the results are quite similar to those in Table XI. Controlling for price variation has little effect on the results. However, one notable difference is that the static CRC model, presented in column 3, can no longer be rejected at conventional levels and the CRC model with learning is only weakly rejected at the 10 percent level. The homogeneous returns models, presented in columns 1 and 2, are still overwhelmingly rejected. Additionally, the estimates of the θ's from the full model in column 4 are larger in this specification, though still insignificant.

Figure III presents graphically the degree of heterogeneity and learning in the estimated perceived returns to entrepreneurship from the full model with both endogenous capital and price controls (i.e. the dynamic CRC model with learning from column 4 of Table XII). That is, I can calculate from the estimated structural parameters the expected productivity gains from engaging in entrepreneurship that households uses in their entry decision in each period (i.e. $\beta + \phi(m_{i,t-1})$). Note that the estimates will recover the average perceived return for a given entrepreneurship history (i.e., they will only differentiate households by their entrepreneurship history); accordingly, there will be 4 different perceived returns in each time period. Figure III shows that households that switch into entrepreneurship and those that choose to stay in entrepreneurship, indeed, expect higher productivity gains in period two, whereas households that choose to stay out or switch out of entrepreneurship do not perceive such increases in the productivity gains. Additionally, the average perceived productivity gain over time varies across these different types of households, verifying that there is heterogeneity even in the initial beliefs. The differences between productivity gains within history across time in Figure III are not statistically significant, as mentioned above, but support a learning interpretation for the dynamics observed in the data.

Figure IV repeats this exercise for the static CRC model with both endogenous capital and price controls corresponding to column 3 in Table XII. Notice in this model perceived productivity gains will vary by entrepreneurship history, but not within entrepreneurship history over time. That is, the formula for perceived productivity gains is $\beta + \phi(\eta)$ in this model, which does not vary over time. Once again, I find that the perceived productivity gains vary by entrepreneurship history, and, as shown in Table XII, this variation across histories is statistically significant.

6.4 Credit Constraints vs. Comparative Advantage

Notice both of these competing models of endogenous easing of financial constraints and learning about comparative advantage produce dynamic, heterogeneous returns to switching sectors as presented in the preferred econometric framework of the dynamic correlated random coefficients (DCRC) model. However, using the differential predictions for switching trajectories given by the competing models of endogenous easing of financial constraints and learning about comparative advantage, I can distinguish between the two models.

To the degree that latent heterogeneity reflects predominantly financial constraints rather than relative entrepreneurial abilities, the estimate of ϕ in both the static and dynamic CRC models should be positive. That is, if, as predicted by previous theoretical work, the highest ability households are most constrained, positive productivity shocks last period should drive these high ability households into entrepreneurship and lead to a positive relationship between earnings in the two sectors.

Similarly, in the dynamic model, the estimates of the θ's ought to be positive as well. That is, if households are endogenously easing credit constraints through savings, a positive productivity shock in the default sector this period should make households more likely to switch into the entrepreneurial sector next period. Furthermore, because under this scenario, the entrepreneurial sector is the more productive sector for these households, the effect of the positive productivity shock in the default sector last period on earnings will translate into higher earnings in the current period. The θ's will capture this relationship between current productivity shocks and future choices.

The negative estimates of the θ's and the ϕ validate the interpretation of latent dynamic heterogeneity as learning about comparative advantage. That is, the negative θ's imply that a positive productivity shock in the current period *reduces* the probability of switching sectors next period, as predicted by the preferred model of learning about comparative advantage. Taken together with the negative estimate of ϕ, the results imply that this decision to stay in the sector in which the household received a positive productivity shock leads to higher earnings in the next period. Overall, the results show that the learning about comparative advantage mechanism dominates the sorting decision.

7 Conclusion

Previous studies have argued that entrepreneurship is, potentially, an important stimulus of growth, especially in developing contexts. However, despite the perceived importance of household enterprise, a majority of households do not engage in entrepreneurial activities. Previous studies have proposed that credit constraints preclude many households, perhaps even the highest return households, from starting businesses. On the other hand, heterogeneous abilities or, specifically, comparative advantage in entrepreneurship would also justify less than full participation in the entrepreneurial sector. However, heterogeneity in costs and/or returns are difficult to observe and account for in the estimation of returns to entrepreneurship, in particular when heterogeneity is non-additive and dynamic.

In this study, I present a model which includes both comparative advantage in entrepreneurship and learning, along with constraints on capital. I then develop an econometric approach, following projection-based panel data methods, to estimate this model. Furthermore, I test between the full model and nested models which restrict returns to be homogenous and/or information to be perfect. I estimate these models on data from an annual panel survey in Thailand and find strong evidence of selection on comparative advantage in entrepreneurship. Specifically, the results suggest that households with the lowest earnings in non-entrepreneurial production have the highest returns to entrepreneurship. The estimates of the learning parameters, though imprecise, provide evidence that households learn about their high entrepreneurial comparative advantage from lower-than-expected productivity in the default sector and/or high productivity realizations in entrepreneurship.

These results are informative for policy-makers in developing contexts like the one investigated in this study in two ways. First, the results suggest that investing in additional financial resources for this region will not likely have strong effects on entrepreneurial activity. Second, allocating resources toward training programs to improve skills necessary for successful entrepreneurship among non-entrants could have a large impact on sectoral choice.

Some recent experimental interventions have begun to explore the effects of training programs on the growth and success of microenterprises. For example, Karlan and Valdivia (2011) explore the incremental benefits to adding business training to microfinance interventions in Peru. They find no evidence of effects on business outcomes. However, if my results are correct, this lack of an effect of training on existing enterprises is not surprising. That is, under my model, households sort into the entrepreneurial sector because they feel that their business skills are already relatively sharp; rather it is the marginal non-entrant household that might benefit most from training programs. The testing of this hypothesis is left to future study.

References

[1] Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2010. "The miracle of microfinance? Evidence from a randomized evaluation." Working Paper.

[2] Banerjee, Abhijit V. and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy,* vol. 101, no. 2: 274-298.

[3] Banerjee, Abhijit V. and Andrew F. Newman. 1994. "Poverty, Incentives, and Development." *American Economic Review,* vol. 84, no. 2: 211-215.

[4] Buera, Francisco. "A dynamic model of entrepreneurship with borrowing constraints: theory and evidence." *Annals of Finance,* vol. 5, no. 3-4: 443-464.

[5] Buera, Francisco, Joseph Kaboski and Yongseok Shin. "Finance and Development: A Tale of Two Sectors." *American Economic Review,* vol. 101, no. 5: 1964-2002.

[6] Burgess, Robin and Rohini Pande. 2005. "Do rural banks matter? Evidence from the Indian Social Banking Experiment." *American Economic Review,* vol. 95, no. 3: 780-795.

[7] Card, David. 1995. "Earnings, Schooling, and Ability Revisited." *Research in Labor Economics,* vol. 14.

[8] Carneiro, Pedro and James J. Heckman. 2002. "The Evidence on Credit Constraints in Post-Secondary Schooling." *The Economic Journal,* vol. 112, no. 482: 705-734.

[9] Chamberlain, Gary. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics,* vol. 18: 5-46.

[10] Chamberlain, Gary. 1984. "Panel Data." *Handbook of Econometrics,* ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland.

[11] Conley, Timothy G. and Christopher R. Udry. 2010. "Learning about a new technology: pineapple in Ghana." *American Economic Review,* vol. 100, no.1: 35-69.

[12] Crepon, Bruno, Florencia Devoto, Esther Duflo, and William Pariente. 2011. "Impact of microcredit in rural areas of Morocco: Evidence from a Randomized Evaluation." Working Paper.

[13] DeGroot, Morris. 1970. *Optimal statistical decisions*. New York: McGraw Hill.

[14] de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics,* vol. 123, no. 4: 1329-1372.

[15] Dunn, Thomas and Douglas Holtz-Eakin. "Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links." *Journal of Labor Economics,* vol. 18, no. 2: 282-305.

[16] Dupas, Pascaline and Jonathan Robinson. 2011. "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya." Working Paper URL: `http://www.stanford.edu/~pdupas/SavingsConstraints.pdf`

[17] Evans, David S. and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy,* vol. 97, no. 4: 808-827.

[18] Foster, Andrew and Mark Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy,* vol. 103, no. 6: 1176-1209.

[19] Foster, Andrew and Mark Rosenzweig. 2004. "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000." *Economic Development and Cultural Change,* vol. 52, no. 3: 509-542.

[20] Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. 2005. "Comparative Advantage, Learning, and Sectoral Wage Determination." *Journal of Labor Economics,* vol. 23, no. 4: 681-723.

[21] Hurst, Erik and Annamaria Lusardi. "Liquidity Constraints, Household Wealth, and Entrepreneurship." *Journal of Political Economy*, vol. 112, no. 2: 319-347.

[22] Islam, Nazrul. 1995. "Growth Empirics: A Panel Data Approach." *The Quarterly Journal of Economics*, vol. 110, no. 4: 1127-1170.

[23] Kaboski, Joseph and Robert Townsend. 2012. "The Impact of Credit on Village Economies." *American Economic Journal: Applied Economics.* forthcoming

[24] Karlan, Dean and Martin Valdivia. 2011. "Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions." *Review of Economics and Statistics*, vol. 93, no. 2: 510-527.

[25] Karlan, Dean and Jonathan Zinman. 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies*, vol. 23, no. 1: 433-46.

[26] Lemieux, Thomas. 1993. "Estimating the Effects of Unions on Wage Inequality in a Two-Sector Model With Comparative Advantage and Non-Random Selection." Working Paper, CRDE and Departement de Sciences Economiques, Universite de Montreal.

[27] Lemieux, Thomas. 1998. "Estimating the Effects of Unions on Wage Inequality in a Two-Sector Model With Comparative Advantage and Non-Random Selection." *Journal of Labor Economics*, vol. 16, no. 2: 261-291.

[28] Midgrigan, Virgiliu and Daniel Xu. "Finance and Misallocation: Evidence from Plant-level Data." Working paper, New York University.

[29] Nyshadham, Anant. 2011. "Describing the Marginal Enrollee: Merit-Based Tuition Subsidies Revisited." Working Paper URL: `http://nyshadham.squarespace.com/storage/AN_scholarship_sep2011.pdf`

[30] Paulson, Anna L., Robert M. Townsend, and Alexander Karaivanov. 2006. "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship." *Journal of Political Economy*

[31] Suri, Tavneet. 2011. "Selection and Comparative Advantage in Technology Adoption." *Econometrica,* vol. 79, no. 1: 159-209.

A Extension to 3 Periods

Please see the online Appendix at:

`http://nyshadham.squarespace.com/storage/Nyshadham_JMP_Appendix_Nov2011.pdf`

B Omitted Equations

The capital choices from the model with learning about comparative advantage are:

$$K^*_{iEt} = \left[\frac{\rho}{r} e^{\beta^E_t + (1+\phi)m_{i,t-1} + 1/2(1+\phi)^2\sigma^2_t + \tau_i} \right]^{\frac{1}{1-\rho}} \tag{59}$$

In the case of $D_{it} = 0$, the household's optimal capital level is

$$K^*_{iFt} = \left[\frac{\rho}{r} e^{\beta^F_t + m_{i,t-1} + 1/2\sigma^2_t + \tau_i} \right]^{\frac{1}{1-\rho}} \tag{60}$$

The minimum distance restrictions implied by the 2 period dynamic CRC model with endogenous capital are:

$$\begin{aligned}
\gamma_1 &= \beta + (1+\phi)\lambda_1 + \phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_{12} + \phi\lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_6 &= (1+\phi)\lambda_{k1-1} + \phi\lambda_{k1} \\
\gamma_7 &= \lambda_{k1-2} \\
\gamma_8 &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-2} \\
\gamma_9 &= (1+\phi)\lambda_{k2-1} + \phi\lambda_{k2} \\
\gamma_{10} &= \lambda_{k2-2}
\end{aligned}$$

$$
\begin{aligned}
\gamma_{11} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + (1+\phi)(\lambda_2+\theta_2) + \phi(\lambda_0+\theta_0) \\
\gamma_{14} &= (1+\phi)\lambda_{12} + \phi\lambda_1 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} + \theta_{k2} \\
\gamma_{17} &= \lambda_{k1-1} \\
\gamma_{18} &= (1+\phi)\lambda_{k1-2} + \phi\lambda_{k1} \\
\gamma_{19} &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-1} \\
\gamma_{20} &= \lambda_{k2-1} \\
\gamma_{21} &= (1+\phi)(\lambda_{k2-2} + \theta k2 - 2) + \phi(\lambda_{k2} + \theta k2) \\
\gamma_{22} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-1} \qquad (61)
\end{aligned}
$$

The corresponding normalizations of λ_0 and θ_0 are

$$
\begin{aligned}
\lambda_0 &= -\lambda_1\overline{D_{i1}} - \lambda_2\overline{D_{i2}} - \lambda_{12}\overline{D_{i1}D_{i2}} - \lambda_{k1}\overline{k_{i1}} - \lambda_{k2}\overline{k_{i2}} - \lambda_{k1-1}\overline{k_{i1}D_{i1}} - \lambda_{k1-2}\overline{k_{i1}D_{i2}} \\
&\quad + \lambda_{k1-12}\overline{k_{i1}D_{i1}D_{i2}} + \lambda_{k2-1}\overline{k_{i2}D_{i1}} + \lambda_{k2-2}\overline{k_{i2}D_{i2}} + \lambda_{k2-12}\overline{k_{i2}D_{i1}D_{i2}} + \psi_{i0} \qquad (62) \\
m_i^1 &= \theta_0 + \theta_2\overline{D_{i2}} + \theta_{k2}\overline{k_{i2}} + \theta_{k2-2}\overline{k_{i2}D_{i2}} + \psi_{i1}
\end{aligned}
$$

The minimum distance restrictions implied by the 2 period static CRC model with endogenous capital are:

$$
\begin{aligned}
\gamma_1 &= \beta + (1+\phi)\lambda_1 + \phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_{12} + \phi\lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1}
\end{aligned}
$$

$$
\begin{aligned}
\gamma_5 &= \lambda_{k2} \\
\gamma_6 &= (1+\phi)\lambda_{k1-1} + \phi\lambda_{k1} \\
\gamma_7 &= \lambda_{k1-2} \\
\gamma_8 &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-2} \\
\gamma_9 &= (1+\phi)\lambda_{k2-1} + \phi\lambda_{k2} \\
\gamma_{10} &= \lambda_{k2-2} \\
\gamma_{11} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + (1+\phi)\lambda_2 + \phi\lambda_0 \\
\gamma_{14} &= (1+\phi)\lambda_{12} + \phi\lambda_1 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} \\
\gamma_{17} &= \lambda_{k1-1} \\
\gamma_{18} &= (1+\phi)\lambda_{k1-2} + \phi\lambda_{k1} \\
\gamma_{19} &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-1} \\
\gamma_{20} &= \lambda_{k2-1} \\
\gamma_{21} &= (1+\phi)\lambda_{k2-2} + \phi\lambda_{k2} \\
\gamma_{22} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-1} \qquad (63)
\end{aligned}
$$

The normalization of λ_0 is identical to that from the DCRC model above.

The minimum distance restrictions implied by the 2 period dynamic CRE model with endogenous capital are:

$$
\begin{aligned}
\gamma_1 &= \beta + \lambda_1 \\
\gamma_2 &= \lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1}
\end{aligned}
$$

$$
\begin{aligned}
\gamma_5 &= \lambda_{k2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + \lambda_2 + \theta_2 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} + \theta_{k2} \\
\gamma_3 &= \gamma_6 = \gamma_7 = \gamma_8 = \gamma_9 = \gamma_{10} = \gamma_{11} = \gamma_{14} = \gamma_{17} = \gamma_{18} = \gamma_{19} = \gamma_{20} = \gamma_{21} = \gamma_{22} = 0 \quad (64)
\end{aligned}
$$

Finally, the minimum distance restrictions implied by the 2 period static CRE model with endogenous capital are:

$$
\begin{aligned}
\gamma_1 &= \beta + \lambda_1 \\
\gamma_2 &= \lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + \lambda_2 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} \\
\gamma_3 &= \gamma_6 = \gamma_7 = \gamma_8 = \gamma_9 = \gamma_{10} = \gamma_{11} = \gamma_{14} = \gamma_{17} = \gamma_{18} = \gamma_{19} = \gamma_{20} = \gamma_{21} = \gamma_{22} = 0 \quad (65)
\end{aligned}
$$

Figure I

Trends in Savings, Self-reported Constraints, and Entrepreneurship

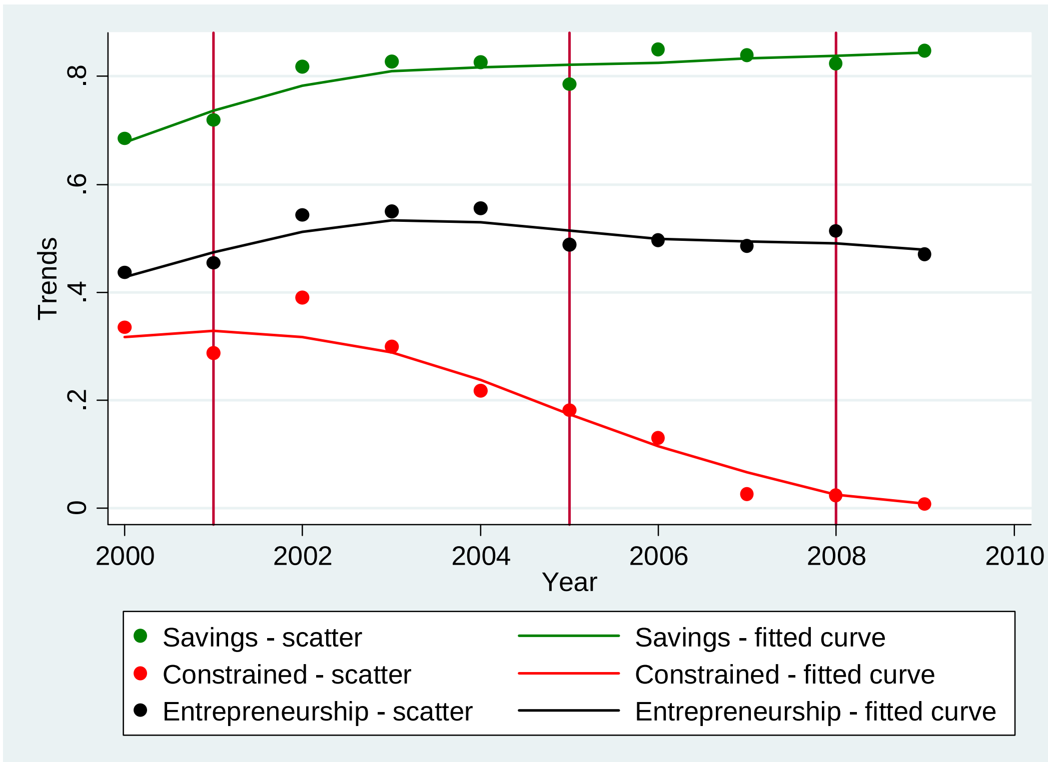

Figure II

Trends in Entrepreneurship and Switching

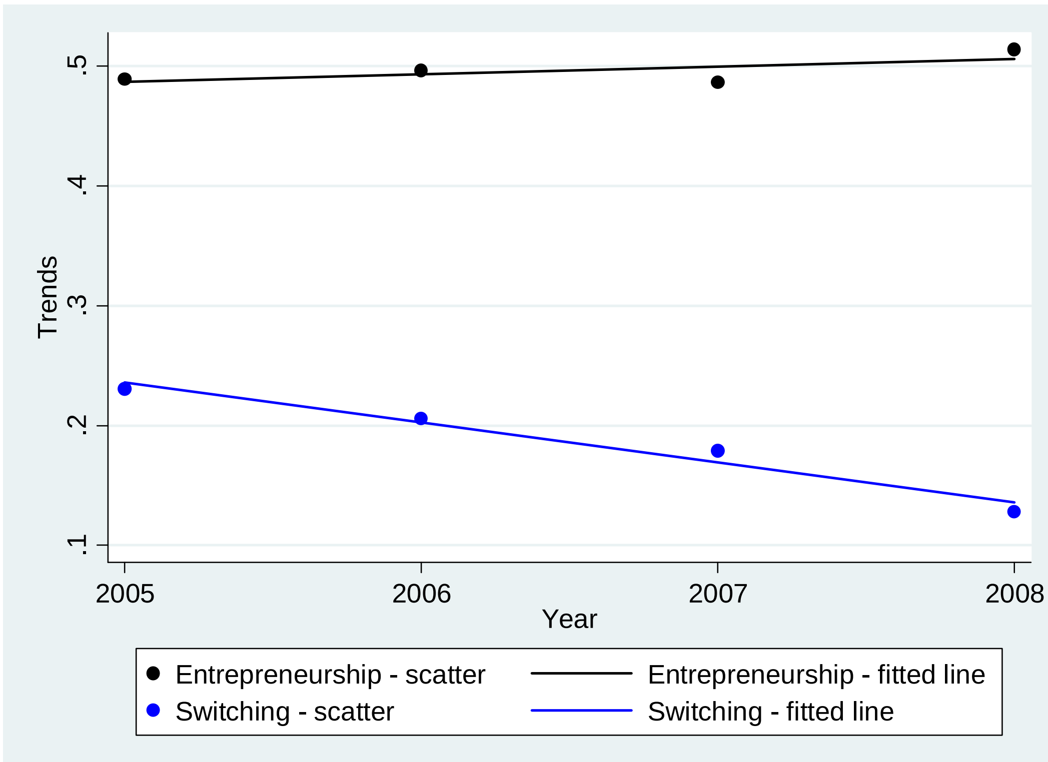

Figure III

Dynamic CRC: Perceived Productivity Gains ($\beta + \phi\ m_{i,t-1}$)

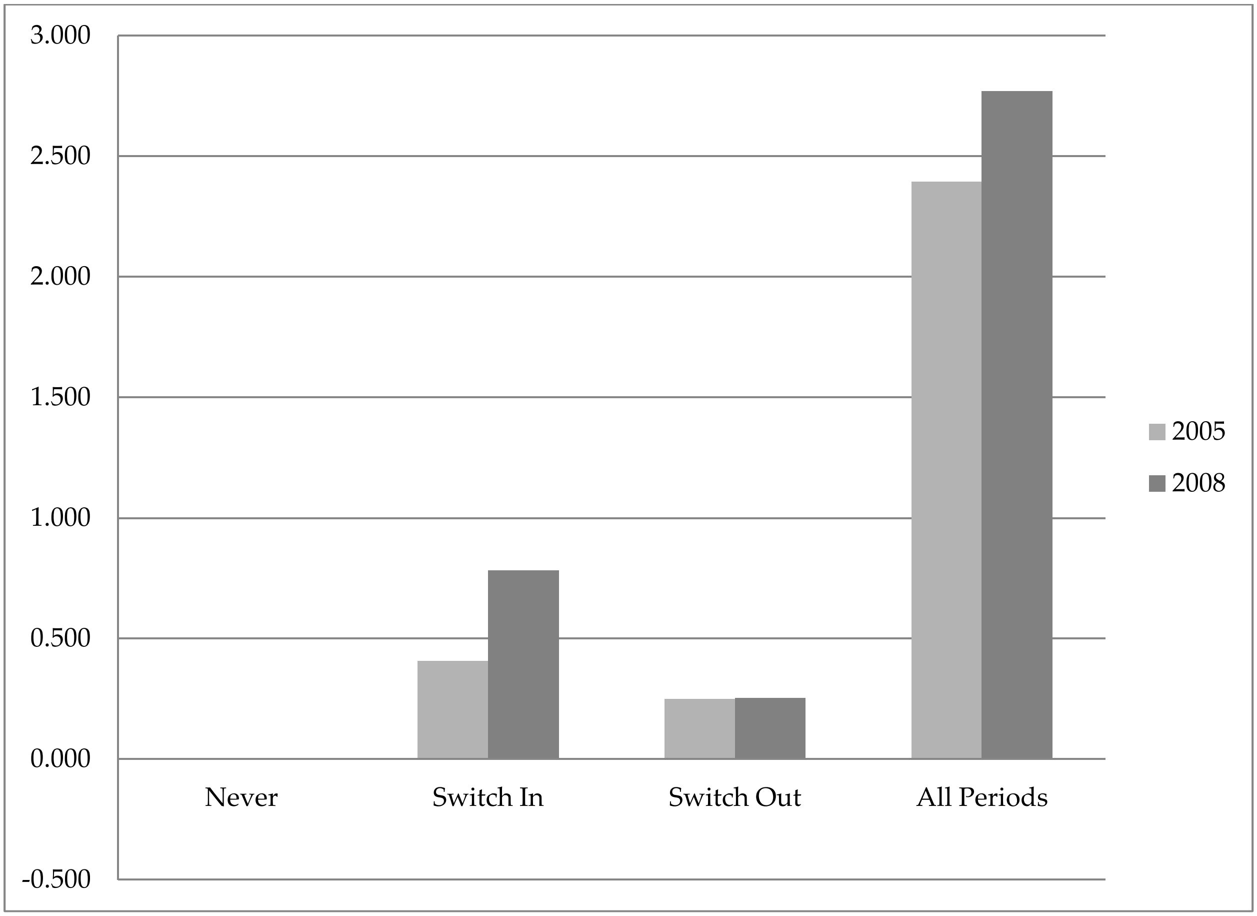

Figure IV

Static CRC: Perceived Productivity Gains ($\beta + \phi\ \eta$)

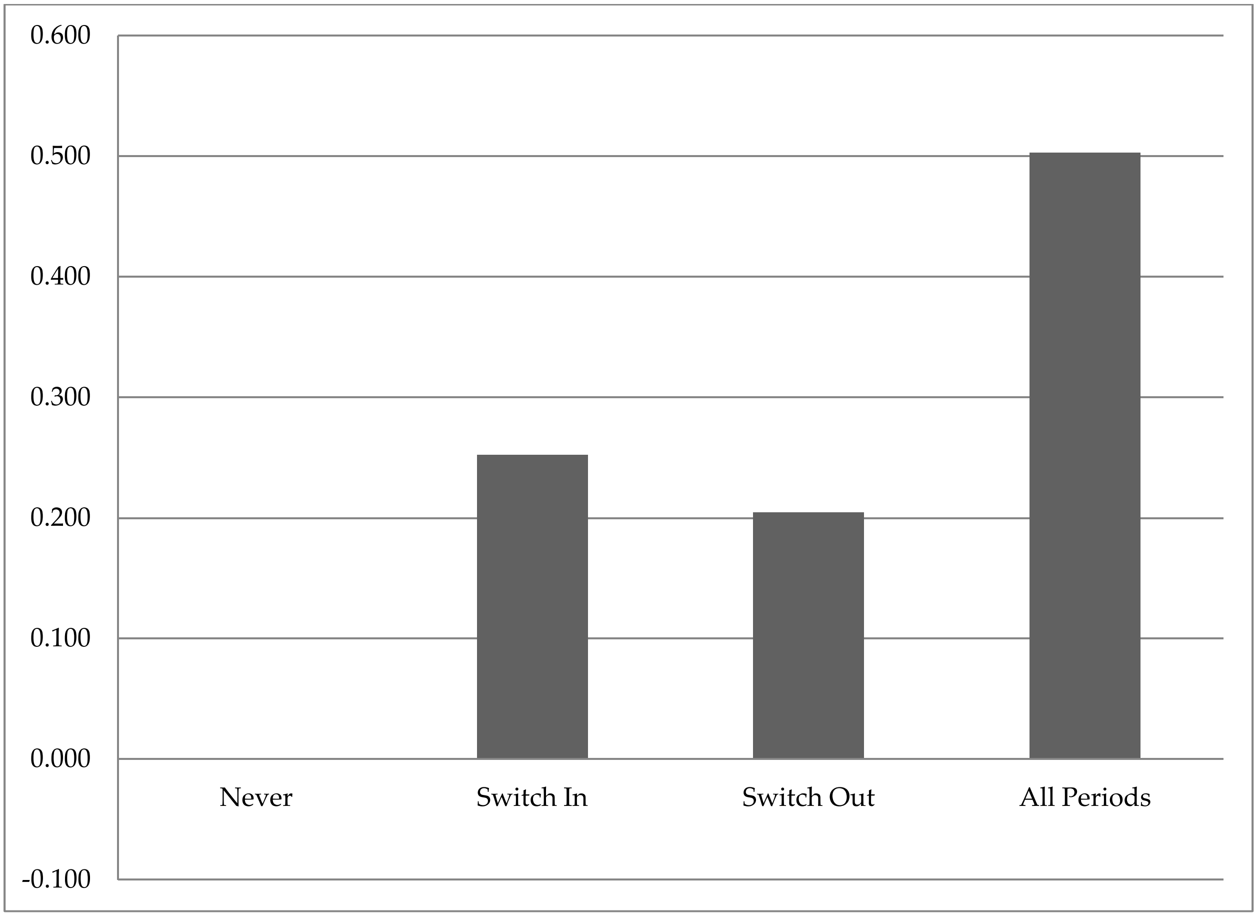

Table Ia: Summary Statistics

Count	1103	
	Mean	**SD**
Income		
ln(gross income), 2005	11.58	1.04
ln(gross income), 2008	11.84	1.03
Entrepreneurship		
Household Business, 2005	0.44	0.50
Household Business, 2008	0.47	0.50
Inputs		
ln(Total Expenditure), 2005	8.23	4.09
ln(Total Expenditure), 2008	8.16	4.50
Household Demographics, 2005		
Household Size	4.23	1.74
Average Age	37.64	13.20
Proportion Male	0.47	0.20
Proportion Completed Primary School	0.27	0.26
Savings		
Household Has Savings, 2005	0.77	0.42
Household Has Savings, 2008	0.83	0.37
Credit Constrained		
Expansion would be profitable, 2005	0.18	0.38
Expansion would be profitable, 2008	0.03	0.16
Borrowing		
Any Loans, 2005	0.80	0.40
Any Loans, 2008	0.77	0.42

Table Ib: Summary Statistics by Entrepreneurship History (Income, Expenditure, and Demographics)

	Business in Both Years		Switch In		Switch Out		Never Own Business	
Count	364		156		123		460	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Income								
ln(gross income), 2005	11.99	1.08	11.42	0.90	11.85	0.91	11.23	0.96
ln(gross income), 2008	12.24	1.18	11.99	0.79	11.82	0.90	11.49	0.89
Inputs								
ln(Total Expenditure), 2005	10.44	2.19	8.09	3.74	9.97	2.38	6.07	4.57
ln(Total Expenditure), 2008	10.59	2.64	9.59	2.98	7.01	4.90	6.07	4.86
Household Demographics, 2005								
Household Size	4.36	1.60	4.49	1.70	4.30	1.72	4.02	1.85
Average Age	35.89	11.35	35.25	11.61	38.35	13.05	39.64	14.73
Proportion Male	0.48	0.18	0.49	0.18	0.47	0.20	0.46	0.23
Proportion Completed Primary School	0.32	0.26	0.28	0.25	0.27	0.25	0.23	0.25

Table Ic: Summary Statistics by Entrepreneurship History (Financial Constraints)

	Business in Both Years		Switch In		Switch Out		Never Own Business	
Count	364		156		123		460	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Savings								
Household Has Savings, 2005	0.87	0.34	0.74	0.44	0.86	0.35	0.68	0.47
Household Has Savings, 2008	0.90	0.30	0.86	0.35	0.88	0.33	0.76	0.43
Credit Constrained								
Expansion would be profitable, 2005	0.26	0.44	0.10	0.30	0.28	0.45	0.10	0.31
Expansion would be profitable, 2008	0.03	0.18	0.05	0.22	0.01	0.09	0.02	0.13
Borrowing								
Any Loans, 2005	0.90	0.31	0.82	0.38	0.83	0.38	0.71	0.45
Any Loans, 2008	0.87	0.34	0.83	0.37	0.78	0.42	0.67	0.47

Table II: Agricultural Price and Savings

Household FE Estimates of Effects Global Price of Rice on Savings, Constraints, and Entrepreneurship

	Price x Farm Intensity			Price		
	Savings Account	Self-reported Constraints	Household Business	Savings Account	Self-reported Constraints	Household Business
Price x Farm Acreage	0.000532	-0.0418***	0.000582			
	(0.00782)	(0.00855)	(0.00910)			
Price	0.0169***	-0.0586***	0.00155	0.0165***	-0.0674***	0.000744
	(0.00259)	(0.00283)	(0.00301)	(0.00209)	(0.00225)	(0.00244)
Farm Acreage	0.0484	0.213***	0.0813**			
	(0.0315)	(0.0344)	(0.0366)			
Observations	11,039	11,039	11,039	11,040	11,323	11,040
R-squared	0.007	0.088	0.001	0.007	0.084	0.000

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table III: Entrepreneurship Decision

Household FE IV Estimates of Effects Savings and Constraints on Entrepreneurship

Saving	0.0973	
	(0.144)	
Constrained		-0.0246
		(0.0365)
Farm Rai	0.0780***	0.0854***
	(0.0280)	(0.0287)
First Stage - F Stat: Saving	22.09	
First Stage - p-value: Saving	< 0.0001	
First Stage - F Stat: Constrained		305.71
First Stage - p-value: Constrained		< 0.0001
Observations	11,039	11,039
R-squared	0.0221	0.0056

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table IV: Labor Market

Percentage of Households with Business Owners, Unpaid Family Workers, and Wage Employees as Members

	Business Owner		Unpaid Family Worker		Wage Employee	
	Mean	SD	Mean	SD	Mean	SD
All Entrepreneurial Industries	0.258	0.438	0.144	0.351	0.043	0.203
Fish or Shrimp Farming	0.033	0.178	0.032	0.175	0.029	0.169
Raising Livestock	0.149	0.356	0.086	0.280	0.033	0.178
Shop / Mechanic	0.076	0.265	0.054	0.226	0.037	0.188
Trade	0.098	0.297	0.063	0.242	0.033	0.178
All Default Industries	0.457	0.498	0.388	0.487	0.419	0.494
Farm	0.456	0.498	0.334	0.472	0.214	0.411
Construction	0.030	0.172	0.029	0.169	0.076	0.265
Low Skilled (Factory, Janitorial, etc.)	0.030	0.170	0.087	0.282	0.144	0.351
High Skilled (Nurse, Teacher, Accountant, etc.)	0.030	0.170	0.030	0.170	0.118	0.323

Table V: Changes in Labor Endowments

Changes in Household Demographics	Mean	SD
1(Change in Household Size)	0.551	0.498
1(Change in Number of Males)	0.430	0.495
1(Change in Number of Primary Educated)	0.514	0.500
1(Change in Number of Unemployed, Inactive, In School)	0.503	0.500

Table VI: Labor Endowments

Partial Correlations of Household Demographics with Entrepreneurship

	Household Business	
	OLS	FE
Household Size	0.0170	0.00672
	(0.0109)	(0.0188)
Number of Males	-0.0145	-0.0180
	(0.0149)	(0.0276)
Number of Primary Educated	0.0616***	0.0138
	(0.0112)	(0.0184)
Number of Unemployed, Inactive, In School	-0.0526***	-0.0207
	(0.0120)	(0.0167)
Observations	2,206	2,206
R-squared	0.0482	0.0324

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table VII: OLS and FE Estimates of Returns to Entrepreneurship

OLS and FE Estimates of Effects of Entrepreneurship on ln(Gross Earnings)

	OLS			FE		
	Prices & Inputs	Inputs	No Covariates	Prices & Inputs	Inputs	No Covariates
Household Business	0.307***	0.245***	0.646***	0.178**	0.194**	0.332***
	(0.0452)	(0.0467)	(0.0516)	(0.0797)	(0.0812)	(0.0804)
ln(Input Expenditure)	0.106***	0.103***		0.0675***	0.0646***	
	(0.00640)	(0.00653)		(0.0130)	(0.0130)	
Observations	2,206	2,206	2,206	2,206	2,206	2,206
R-squared	0.432	0.239	0.095	0.860	0.828	0.815

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Price controls consist of village x time dummies.

Table VIII: Reduced Forms (No Covariates)

3 Period Reduced Form Estimates

	ln(gross income), 2005	ln(gross income), 2008
Household Business 2005	0.622125***	0.330006***
	(0.093139)	(0.090673)
Household Business 2008	0.189557**	0.499113***
	(0.084756)	(0.075249)
Household Business 2005 x 2008	-0.05661	-0.08094
	(0.130628)	(0.126392)
Observations	1103	1103

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table IX: Reduced Forms (Endogenous Capital)

3 Period Reduced Form Estimates

	ln(gross income), 2005	ln(gross income), 2008
Household Business 2005	-0.41665	-0.0900717
	(0.556897)	(0.3444885)
Household Business 2008	-0.51667*	-0.1824934
	(0.26424)	(0.2472835)
Household Business 2005 x 2008	-1.52826**	-2.143647***
	(0.744836)	(0.5956344)
ln(Total Input Expenditure) 2005	0.054438***	0.0017576
	(0.012373)	(0.0108551)
ln(Total Input Expenditure) 2008	0.018068	0.0683273***
	(0.011494)	(0.0099818)
ln(Total Input Expenditure) 2005 x Household Business 2005	0.095389*	0.0375822
	(0.056787)	(0.0343597)
ln(Total Input Expenditure) 2005 x Household Business 2008	-0.03761	-0.033545*
	(0.023784)	(0.0181576)
ln(Total Input Expenditure) 2005 x [Household Business 2005 x 2008]	0.127256*	0.061872
	(0.076373)	(0.0538664)
ln(Total Input Expenditure) 2008 x Household Business 2005	-0.02025	-0.0036433
	(0.020782)	(0.0195239)
ln(Total Input Expenditure) 2008 x Household Business 2008	0.087205***	0.0738686***
	(0.025812)	(0.0245127)
ln(Total Input Expenditure) 2008 x [Household Business 2005 x 2008]	0.024283	0.1338589***
	(0.044773)	(0.0497242)
Observations	1103	1103

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table X: OMD Structural Estimates (No Covariates)

3 Period Minimum Distance Estimates

	CRE	DCRE	CRC	DCRC
λ_1	0.2889***	0.2888***	0.3313***	0.3300***
	(0.0608)	(0.0630)	(0.0900)	(0.0907)
λ_2	0.1668***	0.1677***	0.1877**	0.1896**
	(0.0628)	(0.0644)	(0.0833)	(0.0848)
λ_{12}			-0.0301	-0.0288
			(0.2606)	(0.2685)
θ_2		-0.0038		-0.0083
		(0.0590)		(0.0729)
β	0.3044***	0.3064***	0.3436	0.3493
	(0.0546)	(0.0624)	(0.2050)	(0.2146)
ϕ			-0.1647	-0.1732
			(0.8056)	(0.8235)
χ^2	0.428	0.4238	0.0135	
df	3	2	1	0
observations	1103	1103	1103	1103
p-value	0.9344	0.809	0.9075	

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table XI: OMD Structural Estimates (Endogenous Capital)

3 Period Minimum Distance Estimates

	CRE	DCRE	CRC	DCRC
λ_1	0.2830***	0.2915***	-0.0057	0.0179
	(0.0541)	(0.0562)	(0.3378)	(0.3295)
λ_2	0.0393	0.0310	-0.5282**	-0.4863*
	(0.0560)	(0.0580)	(0.2569)	(0.2639)
λ_{12}			-2.7344**	-3.6703**
			(1.2236)	(1.8511)
λ_{k1}	-0.0063	-0.0074	-0.0042	-0.0062
	(0.0078)	(0.0079)	(0.0102)	(0.0105)
λ_{k2}	0.0299***	0.0310***	0.0079	0.0098
	(0.0081)	(0.0082)	(0.0105)	(0.0109)
$\lambda_{k1\text{-}1}$			0.0361	0.0358
			(0.0323)	(0.0322)
$\lambda_{k1\text{-}2}$			-0.0446**	-0.0505**
			(0.0211)	(0.0228)
$\lambda_{k1\text{-}12}$			0.1518*	0.1841*
			(0.0793)	(0.1004)
$\lambda_{k2\text{-}1}$			-0.0095	-0.0104
			(0.0179)	(0.0187)
$\lambda_{k2\text{-}2}$			0.0970***	0.0962***
			(0.0246)	(0.0250)
$\lambda_{k2\text{-}12}$			0.1144	0.1755
			(0.0749)	(0.1205)
θ_2		0.0392		-0.3772
		(0.0620)		(0.4516)
θ_{k2}		-0.0067		-0.0082
		(0.0071)		(0.0082)
$\theta_{k2\text{-}2}$				0.0342
				(0.0376)
ρ	0.0595***	0.0638***	0.0671***	0.0726***
	(0.0087)	(0.0098)	(0.0102)	(0.0119)
β	0.1858***	0.1633***	0.2191***	0.2408***
	(0.0510)	(0.0607)	(0.0647)	(0.0878)
ϕ			-0.3052	-0.4614**
			(0.2113)	(0.2149)
χ^2	85.1951	84.2665	14.9055	13.149
df	16	14	8	5
observations	1103	1103	1103	1103
p-value	<0.0001	<0.0001	0.061	0.022

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table XII: OMD Structural Estimates (Endogenous Capital and Price Controls)

3 Period Minimum Distance Estimates

	CRE	DCRE	CRC	DCRC
λ_1	0.2099***	0.2133***	0.1465	0.1627
	(0.0484)	(0.0510)	(0.2425)	(0.2464)
λ_2	0.1396	0.1356**	-0.2109	-0.1345
	(0.0518)	(0.0545)	(.2433)	(0.2606)
λ_{12}			-2.1101**	-4.1961
			(1.0329)	(2.9516)
λ_{k1}	0.0056	0.0055	0.0068	0.0050
	(0.0071)	(0.0072)	(0.0091)	(0.0096)
λ_{k2}	0.0231***	0.0231***	0.0133	0.0139
	(0.0077)	(0.0079)	(0.0096)	(0.0105)
λ_{k1-1}			0.0143	0.0137
			(0.0235)	(0.0247)
λ_{k1-2}			-0.0346**	-0.0453**
			(0.0168)	(0.0204)
λ_{k1-12}			0.1512**	0.2543
			(0.0739)	(0.1627)
λ_{k2-1}			-0.0130	-0.0123
			(0.0153)	(0.0171)
λ_{k2-2}			0.0603**	0.0601**
			(0.0253)	(0.0259)
λ_{k2-12}			0.0603	0.1715
			(0.0508)	(0.1585)
θ_2		0.0149		-0.7488
		(0.0683)		(0.7710)
θ_{k2}		-0.0002		-0.0050
		(0.0079)		(0.0090)
θ_{k2-2}				0.0709
				(0.0677)
ρ	0.0608***	0.0610***	0.0641***	0.0686***
	(0.0084)	(0.0095)	(0.0095)	(0.0119)
β	0.1764***	0.1688***	0.2287**	0.3512***
	(0.0519)	(0.0631)	(0.1138)	(0.1166)
ϕ			-0.1432	-0.5512*
			(0.3476)	(0.2947)
χ^2	67.2846	67.2263	12.8105	9.2845
df	16	14	8	5
observations	1103	1103	1103	1103
p-value	<0.0001	<0.0001	0.1185	0.0982

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Price controls are village x time dummies.