Learning about Comparative Advantage in Entrepreneurship: Evidence from Thailand

Anant Nyshadham*†

February 2014

Abstract

I estimate returns to entrepreneurship in Thailand using a novel extension to panel data methods and recover estimates of structural parameters that reveal whether learning about ability drives the entrepreneurship decision in additional to the role financial constraints might play. I find large, positive average returns, but low marginal returns. Structural estimates suggest that learning about comparative advantage in entrepreneurship plays an important role in the sorting decision. Self-reported data on expected incomes support the importance of learning. This study helps reconcile confounding results from the literature. Heterogeneous ability and churning suggest that some households benefit from consulting, training, and financing interventions, while others do not and are, therefore, unlikely to adopt.

JEL Classification Codes: O12, L26, J24, C33

Keywords: microenterprise, learning, ability, credit constraints, nonseperable unobservable heterogeneity, dynamic correlated random coefficients

*Asst. Prof., Univ of Southern California, Economics; nyshadha@usc.edu; http://www.anantnyshadham.com/

†I would like to especially thank Mike Boozer, Penny Goldberg, and Chris Udry for their invaluable guidance and support in the development of this dissertation. I would also like to thank Ach Adhvaryu and Tavneet Suri for numerous helpful discussions and resources; and Joe Altonji, A.V. Chari, Joe Kaboski, Fabian Lange, Costas Meghir, and participants in workshops and seminars at Yale, Stanford, UC Berkeley, Warwick, USC, UIUC, UC Riverside, Johns Hopkins, and the Joint Federal Reserve Bank of Minneapolis/BREAD Conference on Finance and Development for their helpful comments. All errors are my own.

1 Introduction

Despite a great deal of recent work on drivers of entrepreneurial growth and entry in developing countries, still relatively little is known about which types of households choose to start enterprises, which of these household enterprises survive and grow, and of utmost importance to policy makers, why some households enter and potentially succeed while others fail and exit.[1] This paper proposes that households have heterogeneous returns to entrepreneurship due to heterogeneous relative abilities in enterprise over default production sectors (largely, agriculture) and also have imperfect information about their relative abilities. I develop a model of learning about comparative advantage in entrepreneurship which generates dynamics in household entrepreneurial entry and exit decisions as well as capital allocations across sectors. Using this model and a novel extension to a class of panel data models, I reconcile mixed results from recent experimental interventions and explain a set of stylized empirical facts among entrepreneurs in developing countries.

When surveying the data on enterprise activity in various developing contexts, the following empirical regularities emerge: 1.) a large percentage of households engage in some non-agricultural enterprise at any time, and 2.) very few household enterprises grow to represent the primary source of income of the household or to employ workers from outside of the household. In attempting to explain these stylized facts, the literature to date on enterprise in developing countries has focused on constraints to growth. That is, previous studies have presumed that entrepreneurs in developing countries intend to grow their businesses in employment, scale, and profitability, but cannot do so due to market frictions and/or skill constraints.

Earlier work explored the role of access to financial resources in enterprise growth and found mixed results. These studies are comprehensively reviewed in McKenzie (2010).[2] In addition, more recent experimental studies have measured the effects of improved access to microcredit on consumption, durable expenditures, and enterprise activity. In general, the literature has found positive effects on consumption for non-entrepreneurial households and on expenditure on durables for entrepreneurial households (e.g. Banerjee et al. (2010); Crepon et al.

[1] Schoar (2009) and de Mel, McKenzie, and Woodruff (2010) provide interesting descriptive evidence and cross-sectional relationships between enterprise performance and characteristics, but causal evidence and rigorous panel analysis is still lacking in the literature.

[2] See, for example, Burgess and Pande (2005); de Mel, McKenzie, and Woodruff (2008); Dupas and Robinson (2013).

(2011); Kaboski and Townsend (2011)). However, this literature has reported mixed results on entrepreneurial performance (e.g. de Mel, McKenzie, and Woodruff (2008); Dupas and Robinson(2013)) and entry (e.g. Banerjee et al. (2010); Crepon et al. (2011); Kaboski and Townsend (2012)).

A recent, complementary literature has begun to explore entrepreneurial ability or skill as another important determinant of enterprise performance and growth. Some studies have found positive effects on performance of existing enterprises (e.g. Karlan and Valdivia (2011); Bruhn, Karlan, and Schoar (2012); Calderon, Cunha, De Giorgi (2013); Bloom et al (2013)), while others have found insignificant effects or even negative effects (e.g. de Mel, McKenzie, and Woodruff (2012); Karlan, Knight, and Udry (2012)). The completed studies to date in this literature are reviewed in McKenzie and Woodruff (2012). While there is no clear consensus yet on the impacts of skills training and managerial consulting on enterprise growth, the studies in this literature have revealed some common issues. Namely, as noted in the McKenzie and Woodruff (2012) review; enterprise samples are prone to attrition; adoption of training and consulting interventions is often low; and even among adopters implementation of enhanced practices is sometimes limited. Furthermore, this literature is lacking in evidence of the impacts of improved entrepreneurial skill on enterprise entry and exit.[3]

This paper investigates both heterogeneity in access to financial resources and entrepreneurial skill, as well as uncertainty regarding these parameters, as drivers of entrepreneurial entry, exit and performance. A model of economic decision-making which incorporates heterogeneity and uncertainty, though not previously explored in the development entrepreneurship literature, is well-motivated by studies from related literatures. Specifically, along with entrepreneurship, economists have proposed several other important determinants of growth: agricultural technology adoption (e.g. high-yielding varieties, fertilizer, mechanization, irrigation), human capital investment (e.g. schooling, healthcare, nutrition), etc. These literatures have studied models incorporating, alternately, heterogeneity in costs to adopting or investing (e.g. savings, liquidity, or credit constraints), heterogeneous gross returns (e.g. ability in, or preference for, the new technology or sector), and uncertainty regarding returns.[4]

[3] de Mel, McKenzie, and Woodruff (2012) is one notable exception.

[4] Card (1995); Carneiro and Heckman (2002)), for example, have emphasized the role of heterogeneous ability in

I draw on this theoretical motivation from related literatures to propose and test a model of entrepreneurial entry, exit and performance which reconciles the empirical regularities among entrepreneurs in various developing contexts mentioned above with additional empirical facts from the Thai context. Namely, I address an under-emphasized feature of entrepreneurship from the developing world–a high degree of switching or "churning" both in and out of enterprise, particularly among young entrepreneurs. I propose a model of learning about relative ability in entrepreneurship over the default sector of production (often farming) in the presence of simple financial constraints. I then estimate returns to entrepreneurship using this model, the Townsend Thai Project data, and a novel econometric approach. The econometric strategy developed below uses the optimized history of sectoral choices and input levels to reveal the household's heterogeneous type (comparative advantage) as well as evolutions in the household's beliefs regarding their type.[5] Using this strategy, I also recover structural parameters which inform the degree to which the dynamics of entrepreneurship in developing contexts are driven by learning about ability.[6]

The results show a large positive return to entrepreneurship on average, but a great deal of heterogeneity in returns as well. Specifically, the return faced by the household on the margin of switching into the entrepreneurial sector appears to be quite low. Furthermore, the structural estimates suggest that, after accounting for financial constraints in the both the choice of empirical context and the estimating equation, evolutions in beliefs about comparative advantage in entrepreneurship indeed contribute to observed dynamics in the sorting decision. Specifically, farming households are more likely to switch into enterprise after *negative* productivity shocks,

addition to financial constraints. 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 (e.g. Nyshadham (2012)). The literatures on the adoption of agricultural technologies (e.g. Foster and Rosenzweig (1995), Conley and Udry (2010)) and health technologies (e.g. Adhvaryu (2013)) have emphasized learning about returns as an important determinant of the rate of adoption. Ashraf, Berry, and Shapiro (2010) show that even in static adoption decisions heterogeneous returns can lead to selection.

[5]Note that the approach in this paper nests and formalizes the argument made by Schoar (2009) for two extreme, discrete types of entrepreneurial households in developing countries. The continua of relative abilities and financial constraints proposed here will allow easily for both "transformative" and "subsistence" entrepreneurs and allows for entrepreneurial type to be determined endogenously by both ability and cost realizations.

[6]The structural estimation of a dynamic discrete choice model under unobservable heterogeneity is, to the best of my knowledge, novel in the development literature and uniquely applied here to the entrepreneurship decision of developing country households. However, there is a rich literature in labor economics that explores the degree to which heterogeneous skills or ability, particularly unobserved factors, predict differential occupational paths and responses to shocks and information using structural estimation of dynamic models. Keane and Wolpin (1997) is one classic example; Adda et al (2013) is just one recent addition to this literature.

rather than after *positive* shocks as would be the the case if endogenous easing of financial constraints were the sole or primary driver of entrepreneurship dynamics. These structural results are consistent with graphical evidence comparing the evolution in income expectations across households with different enterprise histories. [7]

Taken together, the theoretical framework and structural results in this paper serve to reconcile the mixed findings of previous studies on enterprise responses to financial, managerial, and human capital interventions as well as to explain additional, less emphasized stylized facts on entrepreneurial "churning" or switching. First, returns to enterprise are quite heterogeneous, suggesting that some households would benefit greatly from consulting, training, and even financing interventions, while others would benefit little and be less likely to adopt or comply. Second, households are uncertain about their returns to enterprise and therefore switch in and out of enterprise a great deal (a very common empirical fact in developing contexts), especially early in the productive life cycle. This churning would further reduce the incentives for the subset of households with not-yet-converged expected returns to invest newly available credit into enterprises and/or additional effort into improving business practices. While this study does not purport to explain the entirety of entrepreneurial switching observed in many developing contexts, it argues that a large portion of switching and a reduction in switching over time, particularly among older households that have observed more production signals and have spent a longer time in a particular sector, is explained by a model of learning about comparative advantage in entrepreneurship.[8]

This study makes two main contributions to the literature on enterprise in developing countries. It is the first paper, to my knowledge, to model entrepreneurship as a sorting decision across sectors in the presence of learning about comparative advantage. This model helps to interpret

[7] On the other hand, a comparison of evolutions in savings, borrowing and self-reported financial constraints across households with different enterprise histories provides little evidence of a strong role for financial constraints in the entrepreneurship decision in this empirical context.

[8] Note that learning may not be the only driver of sectoral churning. Adhvaryu and Nyshadham (2013) show households switch in and out of enterprise to weather transitory health shocks among productive members of the household. Adhvaryu, Kala and Nyshadham (2013) show that coffee-growing households engage in enterprise during years in which the global price of coffee is low. This paper differentiates this type of entrepreneurial smoothing endeavor from learning about comparative advantage in entrepreneurship using coincident evolutions in entrepreneurial incomes and capital allocations to entrepreneurship. That is, smoothers will return to agriculture when farm labor productivity and output prices recover and are unlikely to find high returns or even positive returns to entrepreneurship, but households converging to enterprise as the optimal sector of production will realize large, positive returns to entrepreneurship and, accordingly, will choose to stay in the entrepreneurial sector.

not only the results presented below in this study, but also the large set of mixed results from other recent studies in the microenterprise literature. It is also the first paper, to my knowledge, to estimate returns to entrepreneurship in the presence of *both* learning about ability and dynamic financial constraints. Accordingly, it is also, therefore, the first paper to provide structural evidence of the degree to which heterogeneous ability contributes to entrepreneurship decisions in developing settings, after accounting for the presence of heterogeneous financial constraints. Results of this sort are of great importance for informing future policy and research interventions regarding entrepreneurial entry and growth in developing countries.[9]

Finally, this paper also makes a methodological contribution to the literature on estimating models with dynamic, non-separable heterogeneity in unobservables. Specifically, I develop an extension to the class of Chamberlain panel data methods which allows for the estimation of dynamic correlated random coefficients models with discrete regressors of interest. I also discuss how this class of models nests and allows for empirical testing of additional restrictions imposed by simpler models (e.g. without dynamics and/or with additively separable unobservable heterogeneity). Suri (2011) is the most recent study to contribute to this literature and develops the nested correlated random coefficients model. I extend the model from Suri (2011) by relaxing the strict exogeneity assumption to allow for the estimation of time-varying, heterogeneous returns such as those corresponding to models of private learning.[10]

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

[9] As noted below, the theoretical approach in this paper borrows from Gibbons et al (2005). The contributions of this study over Gibbons et al (2005) are the application to entrepreneurship in developing contexts as opposed to job sorting in wage labor markets in developed contexts; the addition of capital and financial constraints to adapt to a comparison of production technologies instead of wage equations; and the structural estimation of this model's parameters as opposed to the IV approach used by Gibbons et al (2005).

[10] I contribute to a rich literature on estimation of nonlinear models in panel data. The literature to date is comprehensively reviewed in Arellano and Bonhmme (2011). The empirical approach in this paper offers a candidate method for estimating dynamic correlated random coefficient panel data models with discrete regressors of interest. The method is fully parametric and restricts the regressor with the correlated random coefficient to be discrete, while allowing for other endogenous regressors to be continuous so long as they enter linearly.

2 Data and Motivation

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.

For the structural analysis undertaken below, 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 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, as well as what they expected their net incomes to be next year.

Despite having 11 waves of data available from 1999-2009, I use only 2 waves in the structural estimation due to the analytical complexity of the strategy undertaken here. As discussed further below, the number of incidental structural parameters to be estimated increases nonlinearly in the number of periods. This issue is exacerbated by the number of production decisions that are treated as fully endogenous (i.e. in what follows I will fully endogenize both sectoral choices as well as corresponding decisions regarding input expenditures). The restriction to a two period estimation also helps to minimize attrition from the balanced panel.

Notably, the sample in waves 2005 onwards were freshened with new households, making them nearly 25% larger than waves from earlier in the panel study. The 3 year gap between survey waves in 2005 and 2008, while short enough to reduce attrition from the balanced panel, is long enough to ensure 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 similar across waves: 44% in 2005 and 47% in 2008. These summary statistics are remarkably stable across all waves, as shown in the preliminary graphical evidence below, and the empirical patterns identified and explained in this study are qualitatively consistent across wave selections.

I focus my analysis on the second half of the past decade in order to avoid confounding the estimation of drivers of equilibrium sorting across sectors with systemic shocks to the Thai economy as a whole like the financial crisis and subsequent political regime change in the late 1990s and early 2000s. That is, the model developed and estimated in this paper is one of dynamic equilibrium sorting of households across sectors on the basis of idiosyncratic productivity shocks. While systemic shocks such as financial infrastructure development or global commodity price shocks are no doubt important for sectoral choice of households in developing contexts, they are outside of the scope of this study and seemingly, as shown below, less informative as drivers of sorting in this empirical context.[11]

2.1 Summary Statistics

In Tables I through IV, I report means and standard deviations for variables of interest in the data. Table I presents for the entire sample summary statistics of gross income, entrepreneurship overall and by business type, input expenditure in total and by category, household demographics, savings, borrowing and self-reported credit constraints. I find that income grows by roughly 25% in the sample from 2005 to 2008, while total input expenditures grow by roughly 30%. The percentage of households with savings (a positive balance in an institutional savings account) and the percentage of households with any outstanding loans are both high in 2005 and remain high

[11] Though the precision of the estimates are sensitive to the gap between waves, the proximity of the waves to the financial crisis and regime change in the previous decade in Thailand, and to size of the balanced panel sample and attrition, the pattern of results is robust to the selection of waves. Results using other are available upon request.

TABLE I: SUMMARY STATISTICS

Count	1103	
	Mean	SD
Income (Thai Baht)		
Gross Income, 2005	193,984.50	370,023.50
Gross Income, 2008	254,883.30	459,292.80
Entrepreneurship		
Household Business, 2005	0.44	0.50
Household Business, 2008	0.47	0.50
Entrepreneurship by Business Type		
Shop/Mechanic/Salon Business, 2005	0.07	0.26
Shop/Mechanic/Salon Business, 2008	0.08	0.26
Fish/Shrimp Business, 2005	0.02	0.13
Fish/Shrimp Business, 2008	0.01	0.11
Trade Business, 2005	0.05	0.21
Trade Business, 2008	0.06	0.25
Other Business, 2005	0.36	0.48
Other Business, 2008	0.38	0.49
Inputs (Thai Baht)		
Total Input Expenditure, 2005	72,732.35	269,254.60
Total Input Expenditure, 2008	96,130.81	304,586.80
Business Inputs by Category (Thai Baht), 2005		
Land and Equipment Rent & Maintenance	4,580.68	24,781.30
Labor	3,355.91	31,382.49
Raw Inputs (e.g. Inventory)	30,674.77	220,698.20
Other	2,899.38	25,581.80
Farm Inputs by Category (Thai Baht), 2005		
Land and Equipment Rent & Maintenance	6,067.54	15,367.16
Labor	5,172.49	17,133.37
Raw Inputs (e.g. Seeds, Fertilizer, Pesticides)	19,413.61	105,066.30
Other	537.07	5,331.64
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
Borrowing		
Any Loans, 2005	0.80	0.40
Any Loans, 2008	0.77	0.42
Credit Constrained		
Expansion would be profitable, 2005	0.18	0.38
Expansion would be profitable, 2008	0.03	0.16

in 2008. These patterns correspond to a large reduction to nearly 0 in the percentage of households reporting that an expansion in their farm and/or non-farm enterprises would be profitable. On the other hand, the probability that a household owns at least one business remains fairly stable. These patterns indicate that high levels, and even growth in, access to and utilization of financial resources do not seem to generate large scale entrepreneurial entry in this context. Below, I present graphical evidence in Figure I that these patterns are preserved in the longer panel as well.

I find that, while the data on types of enterprises in the sample is only informative to a limited extent with many households specifying often unintelligible "other" types of business, the largest categories of enterprise reported are shops and trade/merchant businesses. We might expect these businesses to require less in the way of durable assets or large physical capital investments and more in the way of inventory and other raw inputs. Indeed, we see that both business and farm input expenditures are largely concentrated in raw inputs such as inventory for enterprises and seeds, fertilizer and pesticide for farms. Expenditure on these items is an order of magnitude larger than expenditure on labor or rental and maintenance of land and physical capital. While this might predict a limited role for capital expenditure and durables in entry and exit decisions, as modeled and supported empirically below, it should be noted that these empirical facts regarding types of enterprise activity and categories of input expenditure are not unique to the context explored in this paper, but are quite common across developing contexts.[12]

In Tables II through IV, 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. These histories 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 II also shows that, amongst switchers, households that run businesses tend to have similar gross incomes to those that don't. That is, although households that never own a business

[12]See Adhvaryu, Nyshadham (2013) and Adhvaryu, Kala, and Nyshadham (2013).

TABLE II: INCOME, EXPENDITURE, AND DEMOGRAPHICS BY ENTREPRENEURSHIP HISTORY

	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 (Thai Baht)								
Gross Income, 2005	311,862.20	566,941.90	132,809.60	124,460.40	218,649.00	369,201.40	114,858.70	119,722.80
Gross Income, 2008	428,398.90	729,547.90	224,122.70	218,879.30	204,048.70	221,706.30	141,604.10	136,192.10
Inputs (Thai Baht)								
Total Expenditure, 2005	152,018.30	405,798.20	29,055.44	54,440.49	92,519.07	339,890.00	19,514.40	59,530.20
Total Expenditure, 2008	219,777.90	495,955.50	72,198.94	139,382.50	38,008.27	66,681.57	21,945.86	41,564.95
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 III: FINANCIAL CONSTRAINTS BY ENTREPRENEURSHIP HISTORY

	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
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
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

have lower incomes and households that own a business in both periods have measurably higher incomes, comparing income gains and losses across households that switch into enterprise and households that switch out reveals more modest and intriguing changes. Specifically, households that switch into enterprise earn relatively little before switching and quite a bit more after switching into enterprise; while households that switch out earn more in 2005 than those that stay out or switch in but measurably less than those that stay in enterprise. Furthermore, households that switch out of enterprise earn roughly the same incomes both in and out of the enterprise sector.

On the other hand, I find that input expenditure is systematically higher among entrepreneurial households during their enterprise periods. 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 III, I find that households that engage in entrepreneurship are more likely to have savings and outstanding loans 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 and loans accrue contemporaneously with entrepreneurship, or even following it, rather than savings driving the entrepreneurship decision. Furthermore, the fact that both savings and borrowing activities appear quite high across all entrepreneurial histories seems to suggest the absence of strong financial frictions in this sample. Entrepreneurial households are 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.

Lastly, Table IV shows that not only are shops and trading businesses most likely among entrepreneurial households they are also most likely among those households who stay in enterprise or switch into enterprise. Households that switch out are a bit more likely to have fish/shrimp raising businesses, though shops and trading business are still common among these households. Specifically, I do not find any strong suggestive evidence that any specific type of enterprise appears to be more successful; on the contrary there is some evidence that fish/shrimp businesses are less likely to succeed than shops and trade businesses, though fish/shrimp raising

TABLE IV: BUSINESS TYPES BY ENTREPRENEURSHIP HISTORY

	Business in Both Years		Switch In		Switch Out	
Count	364		156		123	
	Mean	SD	Mean	SD	Mean	SD
Shop/Mechanic Business						
2005	0.18	0.39			0.09	0.29
2008	0.18	0.39	0.11	0.31		
Fish/Shrimp Business						
2005	0.03	0.18			0.07	0.25
2008	0.03	0.18	0.01	0.11		
Trade Business						
2005	0.11	0.31			0.10	0.30
2008	0.14	0.35	0.13	0.34		
Other Business						
2005	0.82	0.38			0.79	0.41
2008	0.81	0.39	0.80	0.40		

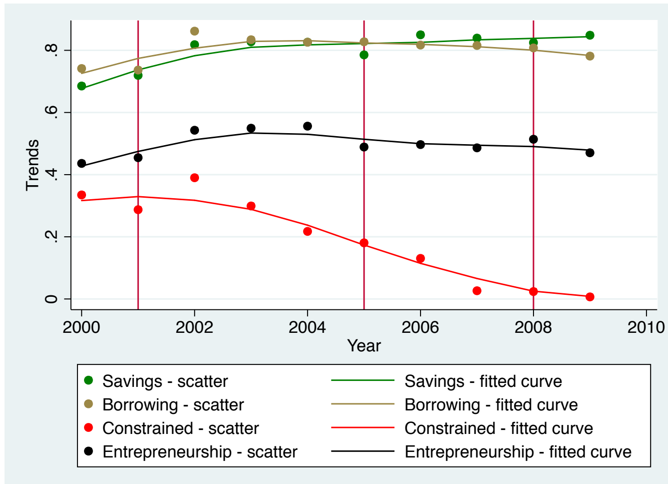
is not overall that common a business in the sample.[13]

2.2 Preliminary Evidence

I begin my analysis of the drivers of entrepreneurship by providing descriptive evidence of the importance of financial resources and constraints for the households' enterprise participation decisions in the empirical context explored in this study. Using data from all 10 waves in the 2000's, I plot in Figure I trends in 4 variables: the proportion of households in each year that 1) own at least one non-agricultural enterprise (black); 2) have a positive balance in an institutional savings account (green); 3) have any outstanding loans (brown); and 4) report feeling financially constrained in expanding their farm or non-farm enterprise to increase profitability (red). It is clear that, despite an upward trend in and high level of savings and borrowing and a steep downward trend in self-reported constraints, aggregate entrepreneurship in the sample stays roughly stable year to year. This evidence suggests that perhaps, at least at an aggregate level, financial constraints do not seem to be the primary determinant of entrepreneurship in this context, particularly later in the decade.

[13]Note that some households engage in multiple enterprise activities, so the sums of the enterprise type percentages will add to larger than 1.

FIGURE I: TRENDS IN SAVINGS, SELF-REPORTED CONSTRAINTS, AND ENTREPRENEURSHIP



Note that there appears to be some, albeit weak, evidence of a relationship between financial resources and entrepreneurship earlier in the decade. In particular, there is a contemporaneous rise in savings, borrowing and entrepreneurship and a decline in financial constraints in 2001 (marked by the left most 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, corresponding to a spike in borrowing in 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.

However, though Kaboski and Townsend (2012) find significant short-term effects of this national-level microcredit initiative on consumption, investment, savings, and income growth, they find no significant effects on entrepreneurial entry. Indeed, Figure I shows that the aggregate participation falls back toward its long-run average level by 2005. From 2005 onwards, participation in the entrepreneurial sector remains stable at roughly 45 %; the percentages of households with positive savings and active loans both plateau 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. The dramatic decrease in self-reported financial constraints and high level of saving and borrowing following program implementation, along with the lack of change in entrepreneurship later in the decade, suggest that the "Million Baht Village Fund" program diminished the role of financial constraints in the entrepreneurship decision for the years studied in this paper, though this role appears limited even earlier in the decade.

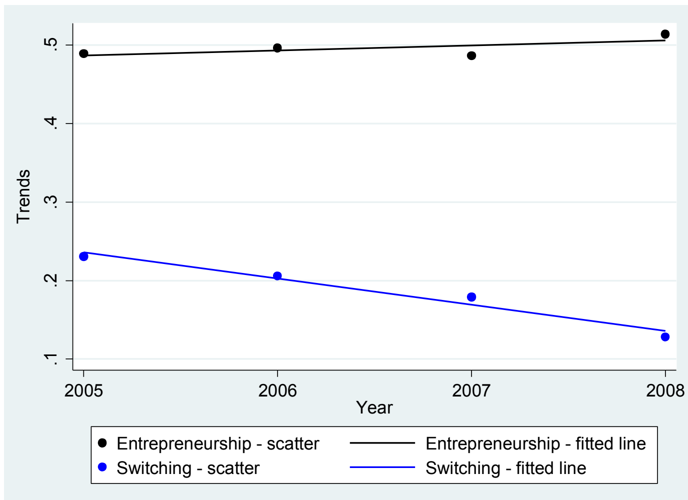
The next pressing question, given the remarkably stable aggregate entrepreneurial participation rate in the sample, is whether the same households engage in enterprise each year. In Figure II, I plot the proportion of households switching their entrepreneurial status from the previous year against the same annual aggregate participation for the time period used in the empirical analysis below (2005-2008). Figure II Panel A depicts a high rate of "churning" or entrepreneurial switching (roughly 25% in 2005). This is a relatively unemphasized and less studied empirical regularity among developing households (see, for example, Adhvaryu, Kala, and Nyshadham (2013)) and corresponds to the issues observed in a great deal of the experimental training interventions discussed above. Additionally, I find a downward trend in switching over time. This reduction in bilateral switching occurs despite a fairly flat aggregate participation rate indicating balanced entrepreneurial entry and exit among the sample.

The high level of (bilateral) switching along with the reduction in switching over time suggest a (symmetric) convergence to an optimal sector as predicted by a model of learning. However, a model of learning would predict that households who have observed more production signals (i.e. older households) would be less likely to switch than would be younger households with less precise priors. More specifically, as households age they should be more likely to stick to a sector and therefore reduce their entrepreneurial switching. As a last piece of motivational evidence, I investigate variation in switching over the productive life cycle of a households in two ways.

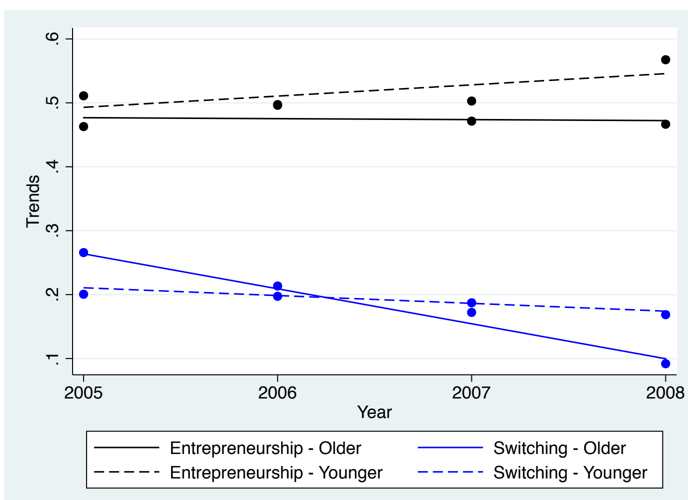
First, I split the sample cohort of households by maximum age of the household (above and below median) and again plot annual entrepreneurial participation and switching. Figure II Panel B shows that the reduction in switching over time is indeed concentrated among house-

FIGURE II: ENTREPRENEURIAL SWITCHING

Panel A: Trends in Entrepreneurship and Switching



Panel B: Trends in Entrepreneurship and Switching by Max Age of HH



holds with older members, while younger households seem to have a persistently high level of switching.[14] Second, I investigate the probability of switching out of a sector as a function of how long a household has operated in that sector. In Figure III, I plot the percentage of households at each level of tenure in a given sector that switch out of that sector in the following year. Panel A shows that 30% of households who have been out of enterprise for a year switch into enterprise the following year; while only 20% of households who have been out for 2 years switch in the following year. The probability continues to decline with nearly 0% of households who have spent 9 years out of the enterprise sector switching into enterprise.

Interestingly, Panel B shows a nearly identical pattern in switching out of enterprise. That is, over 25% of households who have owned an enterprise for only 1 year switch out of enterprise the following year; while only 10 to 18% of households who own an enterprise for between 2 and 7 years switch out. The probability of switching out declines to nearly 0 at 8 or more years of ownership. These patterns in switching by sectoral tenure, while not conclusive evidence of learning, are consistent with a model in which households learn about their relative abilities in enterprise. On the other hand, a model in which dynamics in entrepreneurial entry and exit are determined by endogenous easing of financial constraints over time would predict that households would be more likely to switch into enterprise after *more* years in farming, rather than less, due to the time it takes to accumulate savings or collateral against which to borrow. Similarly, financial constraints cannot explain the symmetric pattern in the probability of switching *out* of enterprise over time.

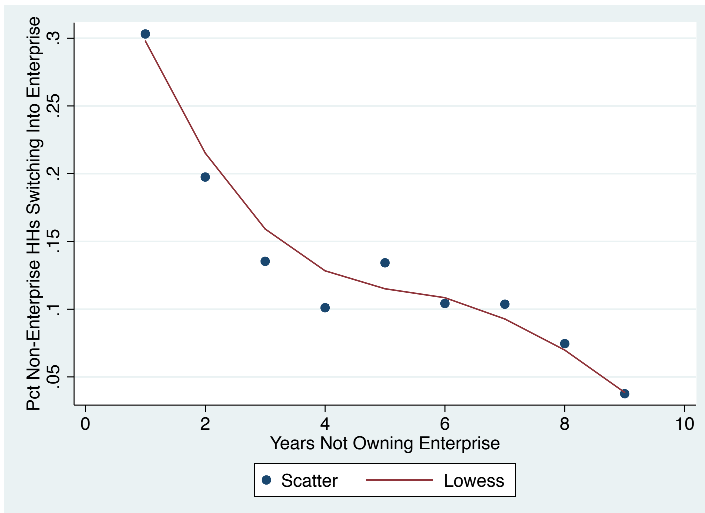
Before developing a model which attributes entrepreneurial dynamics, at least in part, to learning about abilities rather than exclusively to financial constraints, I investigate the role of financial constraints in entrepreneurial choices in one additional way. I estimate the effects of variation in the global price of rice, the predominant agricultural output of Thailand, on savings, borrowing, self-reported constraints, and entrepreneurship.[15] These regressions are run using household fixed effects specifications and the results are reported in Tables D.1 and D.2 of section

[14] I have used the maximum age of the household under the assumption that the number of productivity signals observed by the household is defined by the oldest member of the household. Using mean age of the household produces qualitatively similar figures.

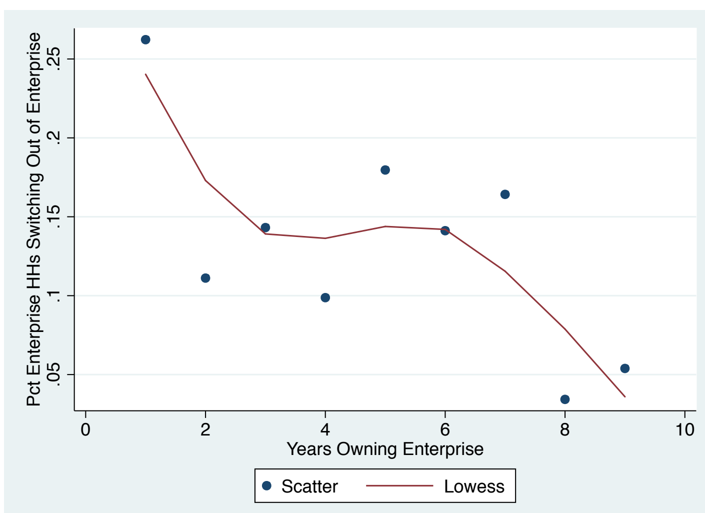
[15] Price data is taken from the IMF monthly agricultural commodity prices and averaged over the year. For this analysis, I once again use data from all waves between 2000 and 2009 in order to allow for greater variation in the price of rice.

Figure III: Sectoral Duration

Panel A: Switching Into Enterprise Against Years Out of Enterprise



Panel B: Switching Out of Enterprise Against Years In Enterprise



D.1 in the Appendix. The results show that indeed the global price of rice and its interaction with farm acreage of the household impact savings, borrowing, and self-reported financial constraints, significantly. However, there is no evidence of a significant impact of rice prices and the their interaction with household farm acreage on entrepreneurship, neither directly in a reduced form specification nor structurally through financial resources and constraints in instrumental variables specifications.

Taken together, Figures I through III and the results shown in Tables D.1 and D.2 suggest that perhaps access to credit is not an important determinant of entrepreneurship decisions. This preliminary evidence provides motivation for the exploration of alternate drivers of entrepreneurship such as latent ability. The model, structural estimation and subsequent interpretation of coefficients below will formalize this comparison across alternative drivers of entrepreneurship.

3 Model

In this section, I develop a model of entrepreneurship that fits the empirical facts discussed above. The model focuses on learning about relative ability across sectors as the main driver of entrepreneurial dynamics, as this model fits most clearly with the patterns identified in the data; however, I include a simple treatment of financial constraints as well and discuss more generally in the estimation how results would differ if financial constraints were indeed a primary determinant of entrepreneurship.

3.1 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 input K, 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 agriculture.[16] Accordingly, default production will also take some level of input K.

[16]To some degree, some part of household income also derives from wage labor by default in urban regions.

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, K_{iFt} is input in farm production, and η_i^F is the heterogeneous component of farm-specific productivity. 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 input under entrepreneurship, and η_i^E is the heterogeneous component of productivity in entrepreneurial activities.[17]

3.2 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)[18], 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:

$$\eta_i^F = b_F(\eta_i^E - \eta_i^F) + \tau_i \tag{3}$$

$$\eta_i^E = b_E(\eta_i^E - \eta_i^F) + \tau_i \quad , \tag{4}$$

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,

[17]Note that the current set up with the possibility of capital investment in both sectors nests previous treatments in the entrepreneurship literature where the default option does not take capital. Furthermore, the data shows that capital allocations to farm activities in this empirical context are non-zero. While the assumption imposed below that returns to capital in both sectors are the same is restrictive and somewhat unrealistic, this assumption does not force capital allocations to be the same across sectors. That is, capital allocations will still vary across sectors, perhaps most importantly, with relative abilities.

[18]The original exposition of this model of self-selection on comparative advantage can be found in Roy (1951).

accordingly, does not affect the sectoral choice.

The household-specific output gain in entrepreneurship over default production can be redefined to be entrepreneurial *comparative advantage*, η_i, as

$$\eta_i \equiv b_F(\eta_i^E - \eta_i^F). \tag{5}$$

Defining $\phi \equiv b_E/b_F - 1$ and using equations 3 and 4, I can express the heterogeneous components of sector-specific productivities in terms of absolute advantage and entrepreneurial comparative advantage :

$$\begin{aligned} \eta_i^F &= \eta_i + \tau_i & (6) \\ \eta_i^E &= (1+\phi)\eta_i + \tau_i \quad . & (7) \end{aligned}$$

Taking logs of production functions 1 and 2 and substituting in using equations 6 and 7, I get[19]

$$\begin{aligned} y_{it}^F &= \beta_t^F + \rho^F k_{it}^F + \eta_i + \tau_i & (8) \\ y_{it}^E &= \beta_t^E + \rho^E k_{it}^E + (1+\phi)\eta_i + \tau_i. & (9) \end{aligned}$$

For the sake of analytical and expositional simplicity, I will assume that $\rho^E \approx \rho^F \equiv \rho$.[20] 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:[21]

$$\begin{aligned} y_{it} &= D_{it}\Big[\beta_t^E + \rho k_{it}^E + (1+\phi)\eta_i + \tau_i\Big] + (1-D_{it})\Big[\beta_t^F + \rho k_{it}^F + \eta_i + \tau_i\Big] \\ &= \beta_t^F + (\beta_t^E - \beta_t^F)D_{it} + \rho[k_{it}^F + (k_{it}^E - k_{it}^F)D_{it}] + \eta_i(1+\phi D_{it}) + \tau_i \end{aligned} \tag{10}$$

[19] In what follows, lower-case letters denote logs. That is, $y = logY$ and $k = logK$.

[20] Relaxing this assumption will not substantively change the interpretation of the estimation results below, especially once I have allowed for village by time varying intercepts as discussed in greater detail below. That is, to the extent that relative returns to capital across sectors vary only at the village by time level, village by time dummies will account for their role in decisions.

[21] While the theoretical setup treats sector as discrete choice, the sample used in the empirical analysis below includes households who participate in both sectors in a given period. Accordingly, the estimation strategy recovers the returns to participating in *any* enterprise activity as compared to producing exclusively in the farm sector.

3.3 Learning

I 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 introduce an additive productivity shock, ε_{it}, to η_i in equation 10 and assume that $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2 = 1/h_\varepsilon)$. That is, the household only observes the sum of η_i and ε_{it}, but not either individually. The generalized log-linear production function then becomes:

$$y_{it} = \beta_t^F + (\beta_t^E - \beta_t^F)D_{it} + \rho[k_{it}^F + (k_{it}^E - k_{it}^F)D_{it}] + (\eta_i + \varepsilon_{it})(1 + \phi D_{it}) + \tau_i \tag{11}$$

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

$$l_{it} = \frac{y_{it} - \beta_t^F - (\beta_t^E - \beta_t^F)D_{it} - \rho k_{it}^F + (\rho k_{it}^E - \rho k_{it}^F)D_{it} - \tau_i}{(1 + \phi D_{it})} = \eta_i + \varepsilon_{it}, \tag{12}$$

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

$$m_t(l_i^t) = \frac{h m_{i0} + h_\varepsilon(l_{i1} + ... + l_{it})}{h + t h_\varepsilon}, \quad \text{and} \quad h_t = h + t h_\varepsilon \tag{13}$$

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. 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.[22] 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 is valued differentially across the two sectors (e.g. managerial skill). Assuming that the household knows ϕ but not η_i corresponds to assuming the household knows how much each sector values these skills but not their own skill stock. Accordingly, households can learn about their stock through production in

[22]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.

either sector.[23]

For example, suppose that η_i represents the household's managerial skill. The agricultural sector rewards managerial skill in its relation to input inventory management and efficient resource allocation. However, the entrepreneurial sector, corresponding to the household's running a noodle shop for example,[24] rewards managerial ability more heavily. Recent studies on small and medium enterprises in developing countries have emphasized the relative importance of such skill in determining enterprise productivity as well as extensive margin, entrepreneurial participation decisions (e.g. Bloom et al. (2013); Bruhn, Karlan, and Schoar (2010); Calderon-Guemez, Cunha, and De Giorgi (2013)). Accordingly, entrepreneurial earnings relative to agricultural earnings are increasing in managerial skill. The assumptions of the model imply that the household recognizes that entrepreneurship rewards managerial ability more than does agricultural production; however, the household is unsure of its specific stock of managerial skill.

Of course, an excellent manager might still be able to earn more in the agricultural sector than someone with the same access to resources but worse skill in allocating those resources. Therefore, a household that initially believes it is bad at management will operate in the agricultural sector to start, where this lack of managerial skill is less penalized; however, should this household find this period that it is better able to manage its agricultural 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 managerial ability. The mechanism, of course, works in the opposite direction as well. I should note that, to the degree that both sectors reward some skills (e.g. work ethic) *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.

In particular, this deviates from an experimentation or learning by doing framework. In such a framework, the household might know its managerial ability, but is unsure of how much the entrepreneurial sector rewards this ability. This requires the household to actually engage in enterprise in order to learn anything at all about its potential in enterprise. Given that the technologies employed in the entrepreneurial sector (e.g. restaurant, mechanic shop, barber shop,

[23]Specifically, the variance of the distribution of shocks is known but not the center. This is a substantive restriction which can be relaxed in future work, but even the current setup adds richness to the existing literature on the dynamics of entrepreneurship in developing country contexts.

[24]Indeed, shops like this make up a large fraction of household businesses in the sample.

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 (e.g. new crops or new varieties of seed). In these contexts, it is less reasonable to assume that households know which skills the new technology or sector will favor.

Furthermore, I will note that the empirical patterns observed in the data do not appear to support a model of learning by doing. A learning by doing model, as well as a human capital accumulation type model for that matter, would predict that a household would switch into enterprise on the basis of expected performance, choose to stay in enterprise if it performed moderately well, but then most importantly, the household's performance would improve on average the longer it stayed in enterprise, either because households leave as soon as their performance drops or because households accumulate experience that make them more productive in enterprise. Additionally, a learning by doing model would not easily explain the symmetric pattern of switching *out* of enterprise observed in the data without having the households learn about farming as well which is inappropriate in agriculturally intensive developing country settings. Finally, as discussed further below, the specific learning about comparative advantage structure imposed here buys me significant analytical tractability in that it simplifies the solution to a series of one-off static optimization decisions, rather than a dynamic programming problem.

Nevertheless, though I will not discuss alternatives in this study, the empirical strategy developed below will be robust to other learning mechanisms and other models of dynamics. That is, the empirical approach will explicitly estimate the correlations between shocks (both past and present) and sector and input decisions (both present and future). If these correlations do not correspond to the specific model of learning proposed here, or to any model of learning for that matter but rather a model of dynamic easing of financial constraints for example, I would interpret the results as evidence that the specific model developed here is not correct. However, in most cases, the empirical strategy would still recover valid estimates of the average return to entrepreneurship and the rank order of household productivities across sectors (i.e. whether good farm households make good enterprise households). The estimated correlations between productivity shocks and productive decisions will inform the correct model, be it the one proposed here or an alternate model. I will reserve a discussion of instances in which the empirical approach is not robust for the estimation section below.

3.4 Sector and Input Decisions

The timing of the household's decisions is as follows:

1. household i chooses its production technology and the corresponding optimal level of input at the beginning of period t using its current beliefs regarding its comparative advantage in entrepreneurship (i.e. $m_{i,t-1} \equiv m_{t-1}(l_i^{t-1})$)
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 equation 12 and updates its expectation of η_i according to equation 13

I will assume that the price of inputs is r and that households face no cost of input adjustment.[25] I will first consider the case in which households are unconstrained with respect to input expenditure. That is, households can acquire as much input as desired at the given price r. Then, the household's input allocation decision in each sector at any time t can be represented as the solution to the following maximization problem:

$$\max_{K_{ijt}} E_t\Big[e^{\beta_t^j} K_{ijt}^{\rho} e^{\eta_i(1+\phi D_{it})+\tau_i} - rK_{ijt}\Big] \quad j \in \{E, F\} \tag{14}$$

where the expectation is with respect to the agent's information at the beginning of period t. The household's optimal input level in each sector will be a function of $E_t[\eta_i] = m_{i,t-1}$:[26]

$$K_{iEt}^* = \kappa\Big(m_{i,t-1}, \sigma_t^2; \phi\Big) \tag{15}$$

$$K_{iFt}^* = \kappa\Big(m_{i,t-1}, \sigma_t^2\Big), \tag{16}$$

[25] This corresponds well to the empirical context where inputs are largely raw materials and inventory rather than rent or maintenance on land and equipment.

[26] 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. The optimal capital choice functions are presented in section A.1 of the Appendix.

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.[27]

Then, household i will choose to produce in the entrepreneurial sector in period t if $[y_{it}^E(K_{iEt}^*) - rK_{iEt}^*] - [y_{it}^F(K_{iFt}^*) - rK_{iFt}^*] > 0$, and in the default farm sector otherwise.[28] Using 1 and 2 and substituting in for optimal input in each sector using equations 15 and 16, I derive a cutoff rule for entrepreneurship. Remembering that $(\eta_i^E - \eta_i^F) \equiv \phi\eta_i$ and that $E_t[\exp\{\phi\eta_i\}] = \exp\{\phi m_{i,t-1} + (1/2)\phi^2\sigma_t^2\}$, household i will choose to produce in the entrepreneurial sector in period t (i.e. $D_{it} = 1$) if and only if[29]

$$\exp\{\phi m_{i,t-1}\} > \exp\{-(\beta_t^E - \beta_t^F) - \phi\sigma_t^2 - (1/2)\phi^2\sigma_t^2\} \tag{17}$$

The sectoral choice 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 17, I can comment on how the model predicts sectoral choices and incomes will change with the evolution of $m_{i,t-1}$:

$$\begin{aligned} m_{i,t-1} &> \frac{-(\beta_t^E - \beta_t^F) - \phi\sigma_t^2 - (1/2)\phi^2\sigma_t^2}{\phi}, \quad \text{if} \quad \phi > 0 \\ m_{i,t-1} &< \frac{-(\beta_t^E - \beta_t^F) - \phi\sigma_t^2 - (1/2)\phi^2\sigma_t^2}{\phi}, \quad \text{if} \quad \phi < 0 \end{aligned} \tag{18}$$

It is clear 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 particu-

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

[28] Note that the current exposition of the model assumes risk neutrality. While risk aversion in sectoral choice and capital allocations is an important and valuable extension, its inclusion is not trivial and is somewhat outside of the scope of this study (see Karivanov and Townsend (2013)). That is, the use of enterprise as an income smoothing mechanism is studied in greater detail in Adhvaryu and Nyshadham (2013) and Adhvaryu, Kala, and Nyshadham (2013) and is possibly also present in this papers empirical context; however, the use of enterprise as smoothing would generate small, or even negative, returns to entrepreneurship and would not predict convergence. Therefore, the inclusion of this smoothing motive for entrepreneurship would not contribute much to the analysis in this study.

[29] Note that this simplified cutoff rule only holds if ϕ is less than 2 in absolute value. The intuition behind this is that if one sector particularly overweights the skills represented by η_i relative to the other sector this cutoff rule will not perfectly summarize the relationship between the belief updates and the sectoral choices over time. Nevertheless, as shown in the results below, estimates of ϕ fall well within this range in this empirical context, indicating that this simplified cutoff rule well summarizes dynamic sectoral choices.

lar, 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.

Additionally, note that in this model the optimal input expenditure and sectoral choices in period t are also the household's best guess at optimal choices in all future periods. This is due to the fact that I have allowed only for uncertainty regarding η_i and also that I have not introduced any costs of entrepreneurial entry, exit, or input adjustment. An alternate treatment of uncertainty, as discussed above in section 3.3, or the introduction of costs of adjustment would turn this series of one-off optimizations into a dynamic programming problem.

Aside from the obvious gains in analytical and computational simplicity of avoiding the dynamic programming problem, the data from this empirical context and other similar settings show high frequency, *balanced* switching in and out of enterprise, which is inconsistent with the notion of such barriers to entry and adjustment and in support of a model of sorting on comparative advantage. I will also reiterate that the preliminary evidence presented in Tables D.1 and D.2 and discussed in section 2.2 above shows that, although shocks to the global price of rice significantly affect savings, borrowing, and self-reported financial constraints (especially for the most agricultural households), they do not significantly affect entrepreneurship in either a reduced form specifications nor structurally through these financial resource measures in instrumental variables specifications.[30]

[30] Note that I have chosen to model all decisions at the household level. Modeling productive and consumptive decisions at the household level is in keeping with tradition in the development literature. The development literature models decisions at the household level because resource allocations are often difficult to separate among members of a household. That is, labor supply allocations are often badly measured if at all and ownership of family enterprises, let alone investment of shared financial resources into these enterprises, are not easily attributable to individual members of the household. Furthermore, it should be noted that when even one of the input markets is imperfect, the production decisions of the household cannot be separated from its consumption decisions.

The empirical context studied in this paper is no exception. Questions regarding enterprise ownership, expenditures and incomes are asked of the household head about the household as a whole. What little data on economic activities that are collected at the individual level are more accurate along external margins (e.g. whether a particular household member works on the farm, for a wage in the market, or for a household enterprise) than along internal margins (e.g. how much ownership over the farm or non-farm enterprise a particular member has or how much time or money a particular household member contributed to these activities). Additionally, as discussed further below, the labor supply information is limited and labor markets are plausibly imperfect in this context.

3.5 Labor

The model presented 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 greatly 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 K. 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 check the validity of these labor assumptions in this context and present the evidence in section D.2 of the Appendix. First, I show that, though there is some evidence of wage labor in the sample, very little of this activity occurs in entrepreneurial industries, but rather mostly in the default sector. Second, I show that, though labor endowments of the household change across time, these changes are uncorrelated with entrepreneurship decisions of the household across time. Finally, it should be noted that any expenditure on labor (e.g. wages or meals for laborers) both from inside and outside the household are recorded and included in the total input expenditure used in the estimation below. It is only uncompensated labor (e.g. the allocation of the households labor endowment) that cannot be accounted for in the estimation.

This evidence alleviates to some degree concerns about the omission of labor allocation decisions of the household from the estimation. Nevertheless, this omission is a short-coming of the empirical analysis in this study. In particular, I unfortunately have 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.

3.6 Limited Liability and Input Constraints

To address the emphasis placed on credit constraints in the existing entrepreneurship literature, I introduce one form of implied financial constraints through limited liability borrowing and discuss the implications for input and sector decisions in section A.2 of the Appendix. 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 addition to the model 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.

As noted in section A.2, with limited liability borrowing, the optimal input choice, and accordingly the sector choice as well, will depend on the household's 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 for the estimation in section 4.3 below, but will begin by developing the method for the unconstrained case.

4 Estimation

Redefining coefficients in equation 11, I arrive at the estimating equation:[31]

$$y_{it} = \alpha_t + \beta_t D_{it} + \rho k_{it} + (\eta_i + \varepsilon_{it})(1 + \phi D_{it}) + \tau_i + \zeta_{it} \tag{19}$$

[31]Note that the empirical analysis is set forth to estimate the return from engaging in some amount of non-agricultural self-employment activity. While other related questions are certainly important (e.g. returns to sophisticated agricultural enterprise over subsistence farming, returns to engaging in one specific entrepreneurial activity over another), each of these questions require specific care and attention in their study and I have chosen the focus of this study to match what I believe to be the priority among development researchers and policy-makers. That is, given the amount of attention and public resources allocated to credit and training for entrepreneurial endeavors, the specific empirical exercise undertaken in this study is of first order importance.

where $\alpha_t \equiv \beta_t^F$, $\beta_t \equiv (\beta_t^E - \beta_t^F)$, $k_{it} \equiv k_{it}^F + (k_{it}^E - k_{it}^F)D_{it}$, 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 $\beta_t = \beta \ \forall t$.[32]

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 which allows me to consistently estimate β and ρ, recover ϕ, and validate the importance of learning in this empirical context. [33]

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 input expenditure choices for the time being and deal only with the endogeneity in sectoral choice. The method will be extended to allow for endogenous input expenditure choices as well in section 4.3 below. The full model with endogenous input choices is of course the preferred one and the one which I will ultimately estimate below; however, the intuition behind how the empirical strategy will deal with endogeneity in input decisions is largely identical to that regarding endogeneity in sector decisions and so, to start, the approach will be more easily articulated with a smaller choice set.

I know from section 3.3 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

[32]Relaxing this assumption has limited empirical content outside of changes in relative prices across sectors. I will explicitly purge all outcome variables and regressors of variation in means across villages and within village over time in the estimation below using village by time dummies as proxies for relative price shifts. That is, to the extent that relative output prices, or even variations in return to inputs across sectors which I assume away here, do not vary within a village in a single year, extending the analysis to estimate a time-varying β seems of little empirical benefit. Addressing variations in relative prices across villages and over time within the village also alleviates concerns regarding the mapping of choices that are determined by optimizing *profit* functions to the estimating equation which corresponds to a generalized *production* function.

[33]Note also that optimal capital and sectoral choices depend on σ_t^2 as well, as expressed in section 3 above. However, in keeping with standard Bayesian learning approaches, households are assumed to have the same precision in their initial draws and in their productivity innovations. That is, the variance of the prior distribution σ_t^2 evolves deterministically with the number of signals received or the number of periods the household has been producing. Furthermore, the dependance of production decisions on the variance in addition to the mean of the prior distribution is, to some degree, an artifact of the imposed functional form of the production function and provides little economic intuition. Accordingly, I suppress this term in the estimation approach that follows. I can extend the analysis to include the age of the household as a covariate, in the same way I include village by time dummies as covariates below, to account for this deterministic trend in the variance of the prior distribution, but in practice this does not change the results.

that the mean of the prior distribution is a martingale. That is, the law of motion for $m_{i,t}$ is

$$m_{i,t} = m_{i,t-1} + \xi_{it} \quad \Rightarrow \quad m_{i,t-1} = m_{i0} + \sum_{k=1}^{t-1} \xi_{ik}, \tag{20}$$

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

$$y_{it} = \alpha_t + \beta_t D_{it} + (m_{i0} + m_i^{t-1} + \varphi_{it})(1 + \phi D_{it}) + v_{it}, \tag{21}$$

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 by 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 of sectoral choices. In particular, the law of motion of the prior, as expressed in equation 20, 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.[34] 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:[35]

$$m_{i0} = \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \psi_{i0} \tag{22}$$

$$m_i^1 = \theta_0 + \theta_2 D_{i2} + \psi_{i1} \tag{23}$$

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

[34]In the online supplement, 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 input expenditure than the one shown below due to the analytical complexity.

[35]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

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 23 nor the interactions of future sectoral choices (though in a 2 period estimation, this is irrelevant).

Note also that the relative sizes of h and h epsilon will determine the degree to which the initial condition, m_{i0} or subsequent updates, m_i^{t-1}, correlate more strongly with choices across periods. I do not explicitly discuss this relationship further as the estimation will approach this issue agnostically. That is, the estimation will allow the data to show (in the projection coefficients) the degree to which initial conditions and subsequent updates affect choices without restricting a priori the relative magnitudes of these correlations. If, for example, a large dispersion in the initial conditions effectively makes their impact on production decisions negligible, the coefficients in equation 22 will be estimated as indistinguishable from 0, while those from equation 23 might be estimated with larger magnitudes and more precision.

Plugging projections 22 and 23 into equation 21, and grouping terms, I get the following log gross output equations:

$$
\begin{aligned}
y_{i1} \quad &= \quad \alpha_1 + \lambda_0 + D_{i1}\Big[\beta + (1+\phi)\lambda_1 + \phi\lambda_0\Big] + D_{i2}\Big[\lambda_2\Big] + D_{i1}D_{i2}\Big[(1+\phi)\lambda_{12} + \phi\lambda_2\Big] \\
&\quad + (1+\phi D_{i1})(\varphi_{i1} + \psi_{i0}) + v_{i1} \qquad (24)
\end{aligned}
$$

$$
\begin{aligned}
y_{i2} \quad &= \quad \alpha_2 + \lambda_0 + \theta_0 + D_{i1}\Big[\lambda_1\Big] + D_{i2}\Big[\beta + (1+\phi)(\lambda_2 + \theta_2) + \phi(\lambda_0 + \theta_0)\Big] \\
&\quad + D_{i1}D_{i2}\Big[(1+\phi)\lambda_{12} + \phi\lambda_1\Big] + (1+\phi D_{i2})(\varphi_{i2} + \psi_{i0} + \psi_{i1}) + v_{i2} \qquad (25)
\end{aligned}
$$

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 22 and 23 (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:

$$y_{i1} = \delta_1 + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i1} D_{i2} + \nu_{i1} \tag{26}$$

$$y_{i2} = \delta_2 + \gamma_4 D_{i1} + \gamma_5 D_{i2} + \gamma_6 D_{i1} D_{i2} + \nu_{i2} \tag{27}$$

Following Chamberlain (1982, 1984), I will first estimate the reduced form coefficients $\{\gamma_j : j \in [1, ..., 6]\}$ by seemingly unrelated regressions (SUR) and then 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

$$\begin{aligned}
\gamma_1 &= \beta + (1+\phi)\lambda_1 + \phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_{12} + \phi\lambda_2 \\
\gamma_4 &= \lambda_1 \\
\gamma_5 &= \beta + (1+\phi)(\lambda_2 + \theta_2) + \phi(\lambda_0 + \theta_0) \\
\gamma_6 &= (1+\phi)\lambda_{12} + \phi\lambda_1
\end{aligned} \tag{28}$$

It appears from (28) that there are 8 structural parameters to be estimated. However, I will impose the following normalizations:

$$\lambda_0 = -\lambda_1 \overline{D_{i1}} - \lambda_2 \overline{D_{i2}} - \lambda_{12} \overline{D_{i1} D_{i2}} \tag{29}$$

$$\theta_0 = -\theta_2 \overline{D_{i2}} \quad , \tag{30}$$

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.[36]

[36] These normalizations will make estimates of the projection coefficients mean zero and reduce the number of projection coefficients to be estimated by 2, improving efficiency at no real loss of generality or interpretation.

Because this model is just-identified, I would not be able to jointly test the restrictions imposed by this model using an over-identification test. However, in the extension discussed in section 4.3 below, which incorporates endogenous input expenditure choices along with the endogenous sectoral choices and is the preferred model that is ultimately estimated, the model is over-identified and can, accordingly, be tested.

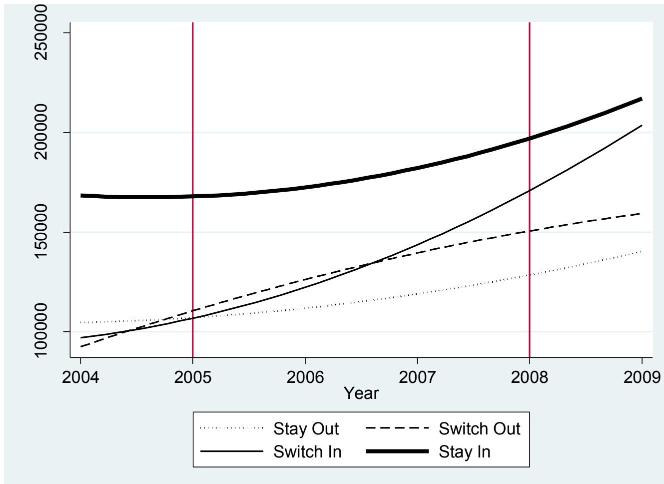
Note that I have not included any exogenous covariates here. In theory, ν_{it} could include, along with v_{it}, any exogenous covariates from equation 21. Though the inclusion of exogenous covariates will affect reduced form expressions 26 and 27, it will not affect the relationships between the reduced 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 4.3.

4.1 Identification Intuition

Note that identification of the structural parameters, such as β, ϕ, the λ's and θ's, comes from a comparison of the log gross output evolutions across households with different sectoral choice histories. That is, I observe in the data the conditional sample mean of log gross output for each entrepreneurship history in each period (i.e. $E(y_{it}|D_{i1}, D_{i2})$). The econometric strategy recovers the contribution that each choice in the optimized history of the household makes to the trajectory of log gross output.

To more clearly convey this intuition, I plot in Figure IV the evolution of realized net incomes (using Lowess smoothing) for each of the 4 possible entrepreneurship histories in 2 periods. The identification comes from comparing across households with marginally different histories (e.g. households that stay in vs. households that switch into the enterprise sector). This amounts roughly to differencing the slopes of the four lines in Figure IV. Note, however, that the preferred specification of the model, as discussed below in section 4.3, fully endogenizes input choices as well which greatly expands the number of type-revealing histories (i.e. the number of line slopes to difference). Additionally, in order to interpret the recovered estimates as structural production function parameters, I use log gross output rather than realized net incomes as the independent

FIGURE IV: REALIZED INCOME OVER TIME BY ENTREPRENEURSHIP HISTORY



variable in the structural estimation below.

In a preliminary inspection of Figure IV, we see that households that stay in the enterprise sector have higher log gross output to start than all other types of households. Additionally, households that stay in enterprise have a log gross output that seems to grow more steeply than does that of households that stay out in both periods or switch out of the enterprise sector in 2008. However, households that switch into the enterprise sector between 2005 and 2008 have low log gross output to start but rise most steeply (have the most convex growth curve), indicating that these households have large returns to entrepreneurship, while other households may not necessarily have the same returns.

4.2 Structural Interpretation of Projection Coefficients

In order to provide structural interpretations of the projection coefficients (i.e. λ's and θ's) I now derive analytical expressions for each parameter. I can express the conditional moments (i.e. $E(y_{it}|D_{i1}, D_{i2})$) plotted in Figure IV 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:

$$\lambda_1 = E[m_{i0}|D_{i1}=1, D_{i2}=0] - E[m_{i0}|D_{i1}=0, D_{i2}=0]; \tag{31}$$

$$\lambda_2 = E[m_{i0}|D_{i1}=0, D_{i2}=1] - E[m_{i0}|D_{i1}=0, D_{i2}=0]; \tag{32}$$

$$\begin{aligned}\lambda_{12} &= \Big\{E[m_{i0}|D_{i1}=1, D_{i2}=1] - E[m_{i0}|D_{i1}=1, D_{i2}=0]\Big\} \\ &\quad - \Big\{E[m_{i0}|D_{i1}=0, D_{i2}=1] - E[m_{i0}|D_{i1}=0, D_{i2}=0]\Big\}\end{aligned} \tag{33}$$

Then, in the context of the model proposed in section 3 and fixing $\phi < 0$ such that good farmers make bad entrepreneurs on average and vice versa, equation 31 says that when $\lambda_1 < 0$ households that engage in enterprise in period 1 and then switch out have higher initial earnings (and beliefs about η_i) than do households that choose to stay out of enterprise in both periods. Similarly, equation 32 says that when $\lambda_2 < 0$ households that choose to switch into enterprise in period 2 have higher initial earnings and beliefs than do households that choose to stay out of enterprise in both periods. Finally, equation 33 says that when $\lambda_{12} < 0$ the gap in initial earnings between households that stay in enterprise and households that switch out of enterprise is smaller than is the gap between initial beliefs of households that switch into enterprise and households that stay out. That is, households that switch out of enterprise in period 2 seem to start out with similar beliefs to those that stay in enterprise; while households that ultimately switch into enterprise appear to have higher initial beliefs than those that stay out.

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

$$\begin{aligned}\theta_2 &= E[m_i^1|D_{i1}=1, D_{i2}=1] - E[m_i^1|D_{i1}=1, D_{i2}=0] \\ &= E[m_i^1|D_{i1}=0, D_{i2}=1] - E[m_i^1|D_{i1}=0, D_{i2}=0]\end{aligned} \tag{34}$$

The expressions in 34 suggest that if $\theta_2 < 0$ while $\phi < 0$, then households that switch into entrepreneurship increase their beliefs about η_i after period 1 by more than do households that choose to stay out; and households that switch out of enterprise decrease their beliefs about η_i after period 1 by more than do households that choose to stay in enterprise.

Note that these λ's and θ are estimated correlations between shocks in period 1 and sectoral choices in period 2. That is, if I estimate $\theta_2 < 0$, then it is the non-entrepreneurial households in the data that had *negative* earnings shocks in period 1 that chose to switch into enterprise in period 2. These estimated signs for both the λ's and θ are consistent with the model of learning about and sorting on comparative advantage in entrepreneurship. A household which does well outside of enterprise in period 1 will learn it is better at farming than it thought and accordingly choose to stay out of enterprise; while a household that does poorly outside of enterprise in period 1 will learn it is worse at farming than it thought and switch into enterprise in period 2. Similarly, a household that does well in enterprise in period 1 will stay in enterprise; while a households that does poorly in enterprise in period 1 will switch out of enterprise in period 2.

On the other hand, note that I am estimating the correlations between productivity shocks in period 1 and choices in period 2 agnostically with the λ's and θ, and then simply interpreting them through the lens of the model presented in section 3 above. I am not, however, imposing the model of learning about comparative advantage in the estimation. Rather, if I find that $\phi > 0$ and $\theta_2 > 0$ (and $\theta_{k2} > 0$ in the extension that follows) the data would be showing that non-entrepreneurial households that received *positive* productivity shocks in period 1 were more likely to switch into enterprise in period 2 which would be inconsistent with learning about comparative advantage and more indicative of a model in which positive earnings shocks eased financial constraints and, in turn, entrepreneurial entry. It remains to be seen in section 5 below which model is supported by the empirical results.

4.3 Endogenous Inputs

As discussed in sections 3.6 and A.2, a household's optimal input allocation and sectoral choice in the presence of financial 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 19 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} (or any other state variable introduced by an alternative treatment of financial constraints, for that matter) 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. That is, the econometric approach as well as the interpretation of β and ϕ will be unaffected by the specific functional form of this relationship, though the predicted signs and interpretation of the λ's and θ's will change as mentioned above.

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 21, I get

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + (m_{i0} + m_i^{t-1})(1 + \phi D_{it}) + \varphi_{it}(1 + \phi D_{it}) + v_{it}, \tag{35}$$

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

$$\begin{aligned} m_{i0} &= \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \lambda_{k1-1} k_{i1} D_{i1} + \lambda_{k1-2} k_{i1} D_{i2} \\ &\quad + \lambda_{k1-12} k_{i1} D_{i1} D_{i2} + \lambda_{k2-1} k_{i2} D_{i1} + \lambda_{k2-2} k_{i2} D_{i2} + \lambda_{k2-12} k_{i2} D_{i1} D_{i2} + \psi_{i0} \qquad (36) \\ m_i^1 &= \theta_0 + \theta_2 D_{i2} + \theta_{k2} k_{i2} + \theta_{k2-2} k_{i2} D_{i2} + \psi_{i1} \qquad (37) \end{aligned}$$

I then proceed as above by substituting these new projections into equation 35 to get reduced form estimating equations similar to equations 26 and 27, but now including input expenditure from each year and their interactions with the history of sectoral choices. The corresponding reduced form regressions are

$$
\begin{aligned}
y_{i1} &= \delta_1 + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i1} D_{i2} + \gamma_4 k_{i1} + \gamma_5 k_{i2} + \gamma_6 k_{i1} D_{i1} + \gamma_7 k_{i1} D_{i2} \\
&\quad + \gamma_8 k_{i1} D_{i1} D_{i2} + \gamma_9 k_{i2} D_{i1} + \gamma_{10} k_{i2} D_{i2} + \gamma_{11} k_{i2} D_{i1} D_{i2} + \nu_{i1} && (38) \\
y_{i2} &= \delta_2 + \gamma_{12} D_{i1} + \gamma_{13} D_{i2} + \gamma_{14} D_{i1} D_{i2} + \gamma_{15} k_{i1} + \gamma_{16} k_{i2} + \gamma_{17} k_{i1} D_{i1} + \gamma_{18} k_{i1} D_{i2} \\
&\quad + \gamma_{19} k_{i1} D_{i1} D_{i2} + \gamma_{20} k_{i2} D_{i1} + \gamma_{21} k_{i2} D_{i2} + \gamma_{22} k_{i2} D_{i1} D_{i2} + \nu_{i2} && (39)
\end{aligned}
$$

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 section A.3 of 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.

4.4 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 productivity shocks (e.g. good rains next year, infrastructure expansion in the village in the near future, rising demand for a specific good in village) 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. Specifically, there are no λ's and θ's included in the estimation to capture correlations between future shocks and past choices. These correlations are assumed to be zero in order to be able to identify the model with multiple endogenous choices and a small number of periods. Specifically, relaxing this assumption further in a model with heteroge-

neous returns leads to an incidental parameters problem causing the model to not be fully identified.[37]

2. Properly Specified Projections - 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 additional endogenous regressors which ought to appear in the structural production function are left out, the projections, as specified, will not be complete and the resulting projection errors will not necessarily be orthogonal to the regressors of interest. That is, if additional productive decisions beyond sector and input expenditure are made by the households on the basis of η_i and are appropriate to include in the structural production functions, but are not included in the projections of m_{i0} and m_{i1}, the estimates of β and ϕ might not be fully purged of endogeneity bias. The most obvious example of such a scenario is if the household allocates its endowment of labor hours across the two sectors and leisure each period. Because I do not observe the allocation of the household's own labor endowment, I cannot include these decisions in the projections. As discussed in section 3.5 above, I have some reason to believe that this may not be such a large concern in my setting.

4.5 Nested Models

In addition to the preferred model presented above, I also estimate nested models which impose additional restrictions on the relationships between η_i and the endogenous choices, D_{it} and k_{it}. Specifically, I estimate 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. These models and the corresponding additional restrictions they impose are presented in section B of the Appendix.

[37] Though this paper contributes to the literature on panel data estimators of correlated random coefficients models by relaxing the strict exogeneity assumption to sequential exogeneity to allow for dynamics, I leave it to future work to relax the sequential exogeneity assumption further to allow for correlations of regressors with both past and future shocks.

5 Results

In this section, I present and discuss results from the empirical analysis discussed in section 4.[38]

5.1 Structural Minimum Distance Estimates

In Table V, I present results from all four models with the endogenous input expenditure as discussed in sections 4.3 and B.[39] I begin with the nested models in order to aid in the interpretation of the added complexity in the preferred model.

5.1.1 Static, Homogeneous Returns Model (CRE)

I present results from the CRE model in column 1. As mentioned above, the CRE model corresponds to a model with static, homogeneous returns to entrepreneurship. In particular, under this model latent ability affects earnings in both sectors equally (i.e. has no effect on returns to entrepreneurship) and the household's perception of this ability does not change over time. Therefore, λ_j and λ_{kj} measure the correlations of the household's sector and input choices, respectively, in period j with the fixed effect. There are accordingly 4 such parameters.

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 input, ρ, is positive and significant at the 1 percent level with a point estimate of nearly .06. The estimate of β is positive and precisely estimated with a point estimate .1858.[40] The χ^2 test statistic corresponding to a joint test of the largest set of 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.

5.1.2 Dynamic, Homogeneous Returns Model (DCRE)

In column 2, I present results from the dynamic CRE model which also restricts returns to be homogeneous, but now allows for households to have imperfect information about this return.

[38] For the sake of comparison, I present ordinary least squares and household fixed effects estimates of the average return to entrepreneurship in section D.3 of the Appendix.

[39] In section D.4 of the Appendix, I present the reduced form coefficients from which I estimate the structural parameters of the econometric models set forth above using minimum distance.

[40] The estimates of ρ and β are quite similar to the results from the household FE regressions presented in column 5 of Table D.6 of the Appendix, as expected.

TABLE V: STRUCTURAL ESTIMATES

	Static, Homogeneous Returns (CRE)	Dynamic, Homogeneous Returns (DCRE)	Static, Heterogeneous Returns (CRC)	Dynamic, Heterogeneous Returns (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).

In the context of this model, the λ's measure the correlation of initial beliefs of households with sector and input decisions, whereas the θ's characterize the degree that and direction in which learning (i.e. the heterogeneity in the update to beliefs between periods 1 and 2) drives changes in sector and input decisions in the future. Specifically, this model introduces two new parameters θ_2 and θ_{k2}, corresponding to the entrepreneurship and input expenditure decisions in period 2, respectively. The estimates of these parameters are small and insignificant. 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.

5.1.3 Static, Heterogeneous Returns Model (CRC)

Column 3 displays results from the static CRC model which allows for heterogeneous returns but restricts this return to be static across time.[41] This model implies that latent heterogeneity will not only affect entrepreneurship decisions in each period, but also the specific history of choices across periods. The static CRC model includes a total of 11 λ's corresponding to the history of entrepreneurial choices, the history of inputs, and their interactions. The estimates of these λ's measure the degree to which the returns to input and sector choices vary across households with different histories of choices. Several of the estimates of the λ's are significant, indicating that earnings vary significantly across households with different enterprise histories. The signs of these parameters are only interpretable in relation to the sign and significance of ϕ.

Recall that ϕ measures the degree to which returns to entrepreneurship vary across households; and, consequently, the degree to which households base their entrepreneurial decisions on this varying return. In the CRC model reported in column 3, the estimate of ϕ is negative but insignificant at conventional levels. Recall that $\phi < 0$ would indicate selection into enterprise on the basis of comparative advantage. On the other hand, $\phi = 0$ would indicate homogeneity in returns to entrepreneurship across households. I interpret the negative estimate of ϕ as suggestive evidence of selection into enterprise on the basis of comparative advantage in entrepreneurship; however, the insignificance implies that restricting that input and sector choices respond to this

[41] This model corresponds to the model developed and estimated in Suri (2011).

comparative advantage in a purely static way leads to a poor fit of the heterogeneity component of the model.

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 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 explain the data much better than do the homogeneous returns models presented in columns 1 and 2.

5.1.4 Preferred Model: Dynamic, Heterogeneous Returns Model (DCRC)

Column 4 reports estimates of the preferred model allowing for dynamic responses to heterogeneous returns. Before discussing the estimates from column 4, I will reiterate the interpretation of the structural parameters from section 4.2. As in the static CRC model, the estimates of the λ's in this preferred model also measure the degree to which the returns to input and sector choices vary across households with different histories of choices. Once again, several of the estimates of the λ's are significant, indicating that earnings vary significantly across households with different enterprise histories. I will discuss the interpretation of the signs of these specific parameters below, after discussing the sign and significance of ϕ.

Recall that ϕ in this preferred model measures the degree to which input and sector choices over time reflect heterogeneity in returns to these choices, now allowing for these choices to respond *differently* to the heterogeneity in returns period to period (e.g. as would be the case if the household were changing its expectation of this heterogeneity over time). $\phi < 0$ indicates, as in the static model in column 3, selection into enterprise on the basis of comparative advantage (i.e. good farmers make bad entrepreneurs). However, in the preferred model, $\phi < 0$ also indicates that households with *low* earnings in the farm sector in period 1 are the households with the *high* returns to entrepreneurship, *and* that these are the precise households that switch into enterprise in period 2. When $\phi < 0$ in the preferred model, the reverse is true as well; that is, entrepreneurial households with low earnings in period 1 are most likely to switch out of enterprise in period 2. Indeed, I find that the point estimate of ϕ in column 4 is large, negative and significant at the 5 percent level with a magnitude of -0.4614.

I will now review the interpretation of the sign of the estimated λ's in this preferred model. When $\phi < 0$, $\lambda_2 < 0$ indicates that households that choose to switch into enterprise in period 2 have higher initial earnings and beliefs (i.e. are closer to the margin of switching) than do households that choose to stay out of enterprise in both periods. $\lambda_{12} < 0$ indicates that households that switch out of enterprise in period 2 seem to start out with similar beliefs to those that stay in enterprise; while households that ultimately switch into enterprise appear to have higher initial beliefs than those that stay out. That is, there is more learning into enterprise than learning out.

Taken together, $\lambda_{k1-2} < 0$ and $\lambda_{k1-12} > 0$, indicate that households that engage in enterprise in both periods have higher initial input levels and lower return per unit input than households that only switch into enterprise in period 2. $\lambda_{k2-2} > 0$ and $\lambda_{k2-12} = 0$ (or only weakly positive) indicate that, by period 2, households that stayed in enterprise both periods and households that only switched into enterprise in period 2 have roughly indistinguishable input levels and per unit input returns by period 2 (with stayer households being insignificantly higher than switchers). On the other hand, $\lambda_1, \lambda_{k1}, \lambda_{k1-1}, \lambda_{k2-1} = 0$ indicate that households that stay out of enterprise in both periods and households that switch out of enterprise in period 2 are roughly indistinguishable in input levels, returns to inputs, and earnings across both periods. These structural estimates correspond remarkably well to the patterns observed in Figure IV.[42]

In this DCRC model, I also recover additional structural parameters corresponding to correlations between sectoral and input choices in period 2 and the productivity shocks observed in period 1. Recall from section 4.2 that if $\theta_2 < 0$ while $\phi < 0$, then households that switch into entrepreneurship increase their beliefs about η_i after period 1 by more than do households that choose to stay out; and households that switch out of enterprise decrease their beliefs about η_i after period 1 by more than do households that choose to stay in enterprise. The estimate of the θ_2, though insignificant at conventional levels is large and negative. The estimate of θ_{k2} is also negative but insignificant.

I interpret these negative θ's (along with the negative ϕ) 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. Given that the estimate of the

[42] A similar picture is obtained for input expenditure across enterprise histories and is available upon request.

ϕ is only significant with inclusion of these learning parameters, the results show that the learning structure is, indeed, important for explaining household behavior in the data. Furthermore, the signs of the more precisely estimated λ's are consistent with the predictions of the learning model.[43]

Finally, in column 4, the estimates of the ρ and β are qualitatively similar to those in column 3; however, the magnitudes 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 larger magnitudes of the ρ and β in column 4 are to be expected given the apparent importance of the dynamics in input and sector choices (i.e. the large discrepancies between λ's corresponding to period 1 choices and those corresponding to period 2 choices). Despite the improved precision of the estimate of ϕ, the preferred model is also rejected with a test statistic of 13 and a p-value of 0.022.

5.1.5 Additional Price Controls (Village x Time Dummies)

Lastly, in Table VI, I present results from the estimation of all four models with endogenous input expenditure, as in Table V, 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 by year 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 V. Controlling for price variation has little effect on the results, other than to reduce the precision of the estimates of the λ's and θ's. 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.

[43]The imprecision in the estimates of the θ's is possibly due to the limited scope afforded the learning structure in a two period estimation. 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 input allocations in 3 periods, I employ a more restrictive treatment of input expenditure 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.

TABLE VI: STRUCTURAL ESTIMATES (VILLAGE X TIME DUMMIES AS PRICE CONTROLS)

	Static, Homogeneous Returns (CRE)	Dynamic, Homogeneous Returns (DCRE)	Static, Heterogeneous Returns (CRC)	Dynamic, Heterogeneous Returns (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)
$\lambda_{k1\text{-}1}$			0.0143	0.0137
			(0.0235)	(0.0247)
$\lambda_{k1\text{-}2}$			-0.0346**	-0.0453**
			(0.0168)	(0.0204)
$\lambda_{k1\text{-}12}$			0.1512**	0.2543
			(0.0739)	(0.1627)
$\lambda_{k2\text{-}1}$			-0.0130	-0.0123
			(0.0153)	(0.0171)
$\lambda_{k2\text{-}2}$			0.0603**	0.0601**
			(0.0253)	(0.0259)
$\lambda_{k2\text{-}12}$			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)
$\theta_{k2\text{-}2}$				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).

FIGURE V: PERCEIVED PRODUCTIVITY GAINS [DYNAMIC CRC: $\beta + \phi m_{i,t-1}$]

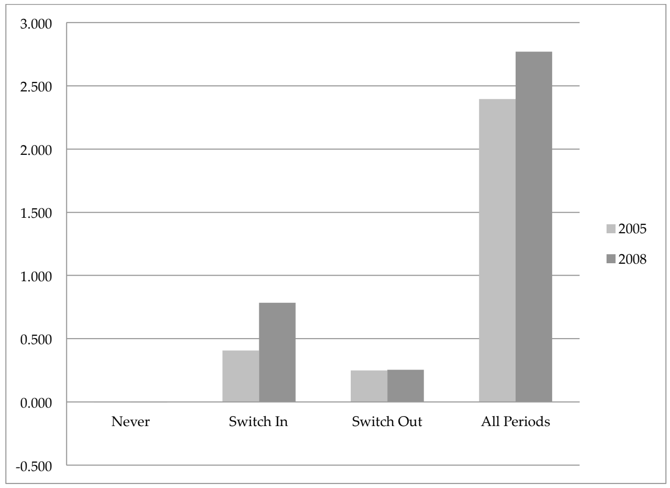


Figure V presents graphically the degree of heterogeneity and learning in the estimated perceived returns to entrepreneurship from the full model with both endogenous input expenditure and price controls (i.e. the dynamic CRC model with learning from column 4 of Table VI). That is, I can calculate from the estimated structural parameters the expected productivity gains from engaging in entrepreneurship that households uses in their entry decision in each period (i.e. $\beta + \phi(m_{i,t-1})$). I graph the average perceived return in each year for each entrepreneurship history, averaging over all input expenditure levels for simplicity. Figure V 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 V are not statistically significant, as mentioned above, but support a learning interpretation for the dynamics observed in the data.[44]

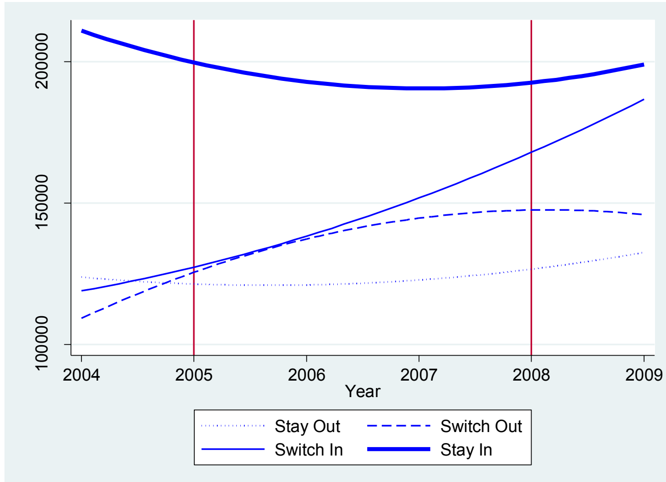
[44] I repeat this exercise for the static CRC model and present it in Figure C.1 in section C of the Appendix.

5.2 Credit Constraints vs. Comparative Advantage

As discussed in section 4.2 above, both models of endogenous easing of financial constraints and learning about comparative advantage produce dynamic, heterogeneous returns to switching sectors. Irrespective of which model is the “true” model driving sector and input decisions of households overtime, average returns to entrepreneurship (β), as well as the degree of heterogeneity in returns (ϕ), are consistently estimated using the proposed econometric approach to the dynamic correlated random coefficients (DCRC) model. However, as presented in section 4.2, using the differential predictions for switching trajectories (and accordingly differential predicted signs of the λ's, θ's and ϕ) given by the competing models of endogenous easing of financial constraints and learning about comparative advantage, I can distinguish between the two in the data.

To the degree that latent heterogeneity reflects predominantly financial constraints rather than relative entrepreneurial abilities, the estimates of ϕ, λ_2, and λ_{12} in both the static and dynamic CRC models (presented in columns 3 and 4 of Tables V and VI) should all be *positive*. That is, if households are predominantly constrained from entering the entrepreneurial sector by finances, then more productive households (those with higher or more positive η_i and therefore higher λ's) should be *more* able to overcome constraints and start successful enterprises, leading to a positive relationship between earnings in the two sectors (i.e. $\phi > 0$). 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. The negative estimates of the λ_2, λ_{12}, θ_2 (along with the analogous input choice projection coefficients) and the ϕ drive the interpretation of latent dynamic heterogeneity as learning about comparative advantage. Overall, the results show, according to the discussion in section 3.4, that the learning about comparative advantage mechanism plays a role in the sorting decision after accounting for financial constraints.

FIGURE VI: EXPECTED INCOME OVER TIME BY ENTREPRENEURSHIP HISTORY



5.3 Expected Incomes by Enterprise History

Finally, I present some graphical evidence in further support of the importance of learning about returns in determining the dynamics of entrepreneurship in this empirical setting. The Townsend Thai Project data includes information on the household's expected income next year, along with the aforementioned information on savings account balances, borrowing activities, and self-reported financial constraints. These data are plotted over time by enterprise history of the household as in Figure IV.

Figure VI depicts the evolution in expected net incomes over time by enterprise history.[45] Note that the pattern of slopes across the histories looks remarkably similar to that of the realized values depicted in Figure IV. In particular, households that stay in the enterprise sector have the highest expected incomes in any period and households that stay out have the lowest. The most important pattern is that households that switch out of enterprise start at roughly similar expectations as those that switch in and those that stay out, but see a less steep rise in expected incomes than those that switch in and a steeper rise in expected income than those that stay out.

I interpret Figure VI as consistent with the idea that households that stay out of, switch out of, or switch into the enterprise sector all start at relatively low expected incomes; however, house-

[45]The expectations data available include expected incomes in a good and bad state as well as on average. The data on expectations in a good state are the least prone to missing and extreme values and are therefore the series used in Figure VI. However, qualitatively similar pictures are obtained when using the other expectations series.

holds that switch out start with steep growth in expectations and ultimately flatten out leading to their exit (concave), while households that switch in start with low growth in expectations and ultimately rise quickly leading to their entry (convex). On other hand, households that stay out see very little rise in expectations and accordingly do not switch their enterprise status. Of course, this figure plots expectations by ex post optimized choice histories and therefore cannot provide strong evidence of causality. Nevertheless, this evidence is in support of the interpretation of the structural estimates above regarding the importance of learning about returns to enterprise.

I repeat this exercise for savings, borrowing, and self-reported financial constraints by enterprise histories of the households and find no clear pattern in support of financial resources or constraints as primary determinants of entrepreneurial entry nor of exit. These plots are presented in Figure C.2 in section C of the Appendix. I also plot both expected incomes and realized incomes by enterprise history on the same graph in order to aid comparisons. I do this both for the whole sample of households as well as, separately, for subsamples of households above and below the median for maximum age of household members. These plots are presented in Figures C.3 and C.4 in section C of the Appendix.

6 Conclusion

Recent experimental interventions offering financial, managerial, and human capital to existing enterprises and potential entrepreneurs provide evidence that interest in adopting these interventions is often low, at least among a large subset of the study sample, impacts are at best heterogeneous and at worst negligible, and attrition from study samples is high indicating a great deal of churning among microenterprises in developing countries. This paper attempts to reconcile these findings with the observed, persistent prevalence of microenterprise in developing countries using a model in which households have heterogeneous returns to entrepreneurship due to differing abilities in enterprise relative to default production and also have imperfect information about their relative abilities.

This model of learning about comparative advantage in entrepreneurship generates dynamics in household entrepreneurial entry and exit decisions that are markedly distinct from those predicted by a model of dynamic easing of financial constraints. Specifically, households constrained from entrepreneurial entry primarily by finances should be *more* likely to switch into enterprise after a positive shock in the default sector; while households constrained primarily by expected ability should learn from the positive shock that they are better at default production than they thought and, accordingly, be *less* likely to switch into enterprise. Using this model and a novel econometric approach to panel data, I estimate the returns to entrepreneurship in Thailand and recover estimates of structural parameters which reveal whether ability or cost constraints dominate the entrepreneurship decision.

The results show a large positive return to entrepreneurship on average; however, the return faced by the household on the margin of switching into the entrepreneurial sector appears to be quite low. The structural estimates suggest that 1) households that are worse at default production (mostly farming) are on average the households that are better at enterprise; and 2) negative shocks in the current sector drive switching out of that sector next period. Taken together, the results show that evolutions in beliefs about comparative advantage in entrepreneurship drive the sorting decision more than does saving or borrowing out of financial constraints. I provide additional, graphical evidence in support of the importance of learning about returns.

Taken together, the theoretical framework and structural results in this paper serve to reconcile the mixed findings of previous studies on enterprise responses to financial, managerial, and human capital interventions as well as to explain additional, less emphasized stylized facts on entrepreneurial switching. Specifically, heterogeneity in returns to enterprise predicts that some households would benefit greatly from consulting, training, and even financing interventions, while others would benefit little and be less likely to adopt or comply. In addition, learning generates dynamics in sectoral choice, especially early in the productive life cycle. This uncertainty about returns and consequently future enterprise activity would further reduce the incentives for households to invest newly available (or newly cheap) credit into enterprises and/or additional effort into improving business practices.

The results of this study have several important implications for policy and future research. Interventions aimed at encouraging entrepreneurship among persistently non-entrepreneurial households might be misguided and inefficient. Some households are simply better at farming and wage employment than entrepreneurial activities. Interventions aimed at improving income trajectories of these "best farmers" could introduce novel technologies or business ideas into local markets and encourage experimentation, but ought not to normatively propose entrepreneurial activity as a necessarily successful endeavor. On the other hand, financial, managerial, and human capital interventions aimed at improving performance among existing enterprises in developing contexts should be targeted to persistently entrepreneurial households that have demonstrated a commitment to the sector. This would not only serve to reduce issues with high attrition and low adoption and compliance, but would likely to lead larger improvements among these more committed and higher ability participants.[46]

[46]This study does not propose to explain the entirety of entrepreneurial switching observed in many developing contexts. Rather it argues that a large portion of switching and the downward trend in switching, particularly among older households that have observed more production signals and households that have persisted in a specific sector longer, is explained by a model of learning about comparative advantage in entrepreneurship. This paper complements related work on entrepreneurial switching in Adhvaryu and Nyshadham (2013) and Adhvaryu, Kala, and Nyshadham (2013). These studies argue that even in the limit, as expected returns and sectoral choices converge, some households might switch in and out of enterprise as a means of weathering shocks to productivity and profitability.

References

[1] Adda, Jerome, Christian Dustmann, Costas Meghir, Jean-Marc Robin. 2013. "Career Progression, Economic Downturns, and Skills." *working paper*

[2] Adhvaryu, Achyuta and Anant Nyshadham. 2013. "Health, Enterprise, and Labor Complementarity." Working Paper.

[3] Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2013. "Booms, Busts, and Household Enterprise: Evidence from Coffee Farmers in Tanzania" *working paper*.

[4] Arellano, Manuel and Stephane Bonhomme. 2011. "Nonlinear Panel Data Analysis." *Annual Review of Economics*, vol. 3: 395-424

[5] Ashraf, Nava, James Berry, and Jesse M. Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *The American Economic Review*, vol. 100, no. 5: 2383-2413.

[6] Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India" *The Quarterly Journal of Economics*, vol. 128, no.1: 1-51

[7] Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2010. "What Capital is Missing in Developing Countries?" *American Economic Review*, vol. 100, no. 2: 629-633.

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

[9] 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.

[10] Calderon-Guemez, Gabriela, Jesse Cunha, and Giacomo De Giorgi. "Business Literacy and Development: Evidence from a Randomized Controlled Trial in Rural Mexico" *working paper*

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

[12] 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.

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

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

[15] 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.

[16] 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.

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

[18] 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.

[19] de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2010. "Who are the Microenterprise Owners?: Evidence from Sri Lanka on Tokman v de Soto." *International Differences in Entrepreneurship,* J. Lerner and A. Schoar (eds.), pp. 63-87.

[20] de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2012. "Business training and female enterprise start-up, growth, and dynamics: experimental evidence from Sri Lanka." *World Bank Policy Research Working Paper* no. 6145

[21] Dupas, Pascaline and Jonathan Robinson. 2013. "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya." *American Economc Journal: Applied Economics.* Vol. 5, no. 1: 163-92.

[22] 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.

[23] 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.

[24] 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.

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

[26] Kaboski, Joseph and Robert Townsend. 2012. "The Impact of Credit on Village Economies." *American Economic Journal: Applied Economics.* vol. 4, no. 2: 98-133

[27] Kaboski, Joseph and Robert Townsend. 2011. "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative." *Econometrica.* vol. 79, no. 3: 1357-1406

[28] Karlan, Dean, Ryan Knight, and Christopher Udry. 2013. "Hoping to Win, Expected to Lose: Theory and Lessons on Micro Enterprise Development." Working Paper.

[29] 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.

[30] Karaivanov, Alexander and Robert Townsend. 2013. "Dynamic Financial Constraints: Distinguishing Mechanism Design from Exogenously Incomplete Regimes." *working paper*

[31] Keane, Michael P. and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *The Journal of Political Economy,* Vol. 105, no. 3: 473-522.

[32] 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.

[33] 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.

[34] McKenzie, David. 2010. "Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?" *World Bank Research Observer* Vol. 25, no. 2: 209-233.

[35] McKenzie, David and Christopher Woodruff. 2012. "What are we learning from business training and entrepreneurship evaluations around the developing world?" *working paper*.

[36] Nyshadham, Anant. 2012. "Describing the Marginal Enrollee: Merit-Based Tuition Subsidies Revisited." Working Paper URL: `http://www.anantnyshadham.com/storage/Nyshadham_MargEnr_sep2012.pdf`

[37] 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,* vol. 144, no. 1: 100-144

[38] Schoar, Antoinette. 2009. "The Divide between Subsistence and Transformational Entrepreneurship." *Innovation Policy and the Economy,* vol. 10, no. 1: pp. 57-81.

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

Appendix

[FOR ONLINE PUBLICATION]

A Omitted Equations

A.1 Capital Choices

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

$$K^*_{iEt} = \left(\frac{\rho}{r} e^{\beta^E_t + (1+\phi)m_{i,t-1} + 1/2(1+\phi)^2\sigma^2_t + \tau_i} \right)^{\frac{1}{1-\rho}} \tag{40}$$

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

$$K^*_{iFt} = \left(\frac{\rho}{r} e^{\beta^F_t + m_{i,t-1} + 1/2\sigma^2_t + \tau_i} \right)^{\frac{1}{1-\rho}} \tag{41}$$

A.2 Limited Liability and Input Constraints

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 K_{ijt} as input into the selected production technology, A_{it} is the household's available savings, and $(K_{ijt} - A_{it})$ is the 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

$$\begin{aligned} D_{it} &= 1: \qquad e^{\beta^E_t} K^\rho_{iEt} e^{(1+\phi)(\eta_i+\varepsilon_{it})+\tau_i} + r(A_{it} - K_{iEt}) \\ D_{it} &= 0: \qquad e^{\beta^F_t} K^\rho_{iFt} e^{\eta_i+\varepsilon_{it}+\tau_i} + r(A_{it} - K_{iFt}) \end{aligned} \tag{42}$$

If the household chooses to default, it receives

$$D_{it} = 1: \qquad e^{\beta^E_t} K^\rho_{iEt} e^{(1+\phi)(\eta_i+\varepsilon_{it})+\tau_i} - \pi A_{it}$$

$$D_{it} \quad = \quad 0: \qquad e^{\beta_t^F} K_{iFt}^{\rho} e^{\eta_i+\varepsilon_{it}+\tau_i} - \pi A_{it} \tag{43}$$

where π is the fraction of assets A_{it} that the household must forfeit as collateral for the defaulted loan.[47]

Then, in equilibrium, a household can only borrow

$$K_{ijt} \leq \left(1+\frac{\pi}{r}\right) A_{it}, \tag{44}$$

where $j \in \{E, F\}$. Then, we have that K_{iEt}^* and K_{iFt}^* are given by equations 15 and 16, respectively, when the credit constraint is not binding. Note that K_{ijt}^* does not depend on assets, A_{it}, in equations 15 and 16. On the other hand, if

$$m_{i,t-1} > \left(\ln\left[(\lambda A_{it})^{1-\rho}\frac{r}{\rho}\right] - \beta_t^F - (\beta_t^E - \beta_t^F) D_{it} - \tau_i - (1/2)(1+\phi D_{it})^2 \sigma_t^2 \right) \frac{1}{1+\phi D_{it}} \tag{45}$$

where $\lambda \equiv \left(1+\frac{\pi}{r}\right)$, 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 input choice, and accordingly 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 $m_{i,t-1}$.

A.3 Minimum Distance Restrictions

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

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

[47]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.

$$
\begin{aligned}
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_6 &= (1+\phi)\lambda_{k1-1} + \phi\lambda_{k1} \\
\gamma_7 &= \lambda_{k1-2} \\
\gamma_8 &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-2} \\
\gamma_9 &= (1+\phi)\lambda_{k2-1} + \phi\lambda_{k2} \\
\gamma_{10} &= \lambda_{k2-2} \\
\gamma_{11} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + (1+\phi)(\lambda_2 + \theta_2) + \phi(\lambda_0 + \theta_0) \\
\gamma_{14} &= (1+\phi)\lambda_{12} + \phi\lambda_1 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} + \theta_{k2} \\
\gamma_{17} &= \lambda_{k1-1} \\
\gamma_{18} &= (1+\phi)\lambda_{k1-2} + \phi\lambda_{k1} \\
\gamma_{19} &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-1} \\
\gamma_{20} &= \lambda_{k2-1} \\
\gamma_{21} &= (1+\phi)(\lambda_{k2-2} + \theta k2 - 2) + \phi(\lambda_{k2} + \theta k2) \\
\gamma_{22} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-1} \qquad (46)
\end{aligned}
$$

The corresponding normalizations of λ_0 and θ_0 are

$$
\begin{aligned}
\lambda_0 &= -\lambda_1\overline{D_{i1}} - \lambda_2\overline{D_{i2}} - \lambda_{12}\overline{D_{i1}D_{i2}} - \lambda_{k1}\overline{k_{i1}} - \lambda_{k2}\overline{k_{i2}} - \lambda_{k1-1}\overline{k_{i1}D_{i1}} - \lambda_{k1-2}\overline{k_{i1}D_{i2}} \\
&\quad + \lambda_{k1-12}\overline{k_{i1}D_{i1}D_{i2}} + \lambda_{k2-1}\overline{k_{i2}D_{i1}} + \lambda_{k2-2}\overline{k_{i2}D_{i2}} + \lambda_{k2-12}\overline{k_{i2}D_{i1}D_{i2}} + \psi_{i0} \qquad (47) \\
m_i^1 &= \theta_0 + \theta_2\overline{D_{i2}} + \theta_{k2}\overline{k_{i2}} + \theta_{k2-2}\overline{k_{i2}D_{i2}} + \psi_{i1}
\end{aligned}
$$

B Nested Models

In this section, I show how the basic framework presented 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 main estimating equation to reflect the particular set of restrictions imposed and, then, redefine the belief projections, estimating equations, and implied minimum distance restrictions, accordingly.

B.1 Heterogeneous Returns with Perfect Information: CRC

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

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i(1 + \phi D_{it}) + v_{it} \tag{48}$$

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^F_{it} and y^E_{it}; 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 input expenditure choices in both periods, and the interaction of input expenditure in both periods with entrepreneurship decisions and their interaction:

$$\begin{aligned}\eta_i \quad = \quad & \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{12} D_{i1} D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \lambda_{k1-1} k_{i1} D_{i1} + \lambda_{k1-2} k_{i1} D_{i2} \\ & + \lambda_{k1-12} k_{i1} D_{i1} D_{i2} + \lambda_{k2-1} k_{i2} D_{i1} + \lambda_{k2-2} k_{i2} D_{i2} + \lambda_{k2-12} k_{i2} D_{i1} D_{i2} + \psi_{i0}\end{aligned} \tag{49}$$

Then, substituting 49 into 48, the corresponding reduced form equations are identical to those from the full model presented in equations 26 and 27; 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.

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

$$
\begin{aligned}
\gamma_1 &= \beta + (1+\phi)\lambda_1 + \phi\lambda_0 \\
\gamma_2 &= \lambda_2 \\
\gamma_3 &= (1+\phi)\lambda_{12} + \phi\lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_6 &= (1+\phi)\lambda_{k1-1} + \phi\lambda_{k1} \\
\gamma_7 &= \lambda_{k1-2} \\
\gamma_8 &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-2} \\
\gamma_9 &= (1+\phi)\lambda_{k2-1} + \phi\lambda_{k2} \\
\gamma_{10} &= \lambda_{k2-2} \\
\gamma_{11} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + (1+\phi)\lambda_2 + \phi\lambda_0 \\
\gamma_{14} &= (1+\phi)\lambda_{12} + \phi\lambda_1 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} \\
\gamma_{17} &= \lambda_{k1-1} \\
\gamma_{18} &= (1+\phi)\lambda_{k1-2} + \phi\lambda_{k1} \\
\gamma_{19} &= (1+\phi)\lambda_{k1-12} + \phi\lambda_{k1-1} \\
\gamma_{20} &= \lambda_{k2-1}
\end{aligned}
$$

$$\begin{aligned}\gamma_{21} &= (1+\phi)\lambda_{k2-2} + \phi\lambda_{k2} \\ \gamma_{22} &= (1+\phi)\lambda_{k2-12} + \phi\lambda_{k2-1}\end{aligned} \tag{50}$$

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

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

$$\theta_2 = \theta_{k2} = \theta_{k2-2} = 0 \tag{51}$$

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 equation 51. 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 equation 51 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.

B.2 Homogeneous Returns with Imperfect Information: DCRE

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

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i + \varepsilon_{it} + v_{it}, \tag{52}$$

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:

$$m_{i0} = \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \psi_{i0} \quad (53)$$

$$m_i^1 = \theta_0 + \theta_2 D_{i2} + \theta_{k2} k_{i2} + \psi_{i1} \quad (54)$$

Notice now that even in the initial belief projection 53, 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:

$$y_{i1} = \delta_1 + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 k_{i1} + \gamma_4 k_{i2} + \nu_{i1} \quad (55)$$

$$y_{i2} = \delta_2 + \gamma_5 D_{i1} + \gamma_6 D_{i2} + \gamma_7 k_{i1} + \gamma_8 k_{i2} + \nu_{i2} \quad (56)$$

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 38 and 39 from the full model, but not in equations 55 and 56 are zero. Therefore, from the 22 reduced form coefficients, I will estimate 8 structural parameters.

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

$$\begin{aligned}
\gamma_1 &= \beta + \lambda_1 \\
\gamma_2 &= \lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + \lambda_2 + \theta_2
\end{aligned}$$

$$
\begin{aligned}
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} + \theta_{k2} \\
\gamma_{3} &= \gamma_6 = \gamma_7 = \gamma_8 = \gamma_9 = \gamma_{10} = \gamma_{11} = \gamma_{14} = \gamma_{17} = \gamma_{18} = \gamma_{19} = \gamma_{20} = \gamma_{21} = \gamma_{22} = 0 \qquad (57)
\end{aligned}
$$

That is, this model imposes 9 additional restrictions on the preferred model:

$$\lambda_{12} = \lambda_{k1-1} = \lambda_{k1-2} = \lambda_{k1-12} = \lambda_{k2-1} = \lambda_{k2-2} = \lambda_{k2-12} = \theta_{k2-2} = \phi = 0 \qquad (58)$$

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 equation 58. The test statistic for this model is distributed χ^2 with 14 degrees of freedom.

B.3 Homogeneous Returns with Perfect Information: CRE

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 or correlated random effects (CRE) model. Under these assumptions, the estimating equation becomes

$$y_{it} = \alpha_t + \beta D_{it} + \rho k_{it} + \eta_i + v_{it} \qquad (59)$$

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

$$\eta_i = \lambda_0 + \lambda_1 D_{i1} + \lambda_2 D_{i2} + \lambda_{k1} k_{i1} + \lambda_{k2} k_{i2} + \psi_{i0} \qquad (60)$$

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

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

$$\begin{aligned}
\gamma_1 &= \beta + \lambda_1 \\
\gamma_2 &= \lambda_2 \\
\gamma_4 &= \rho + \lambda_{k1} \\
\gamma_5 &= \lambda_{k2} \\
\gamma_{12} &= \lambda_1 \\
\gamma_{13} &= \beta + \lambda_2 \\
\gamma_{15} &= \lambda_{k1} \\
\gamma_{16} &= \rho + \lambda_{k2} \\
\gamma_3 &= \gamma_6 = \gamma_7 = \gamma_8 = \gamma_9 = \gamma_{10} = \gamma_{11} = \gamma_{14} = \gamma_{17} = \gamma_{18} = \gamma_{19} = \gamma_{20} = \gamma_{21} = \gamma_{22} = 0 \qquad (61)
\end{aligned}$$

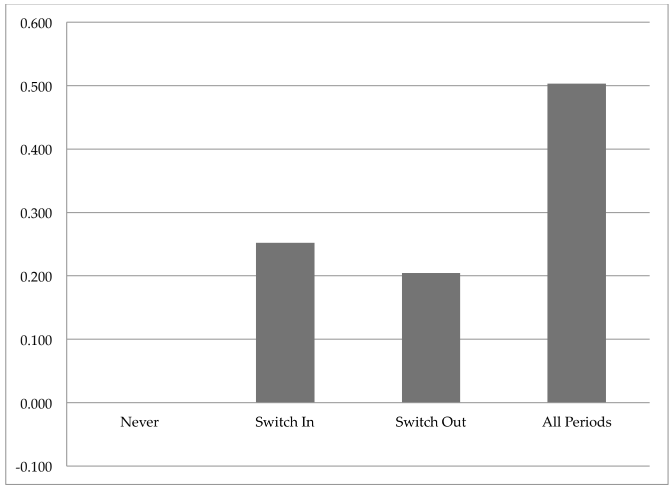
Notice, this model imposes 11 additional restrictions on my preferred model:

$$\lambda_{12} = \lambda_{k1-1} = \lambda_{k1-2} = \lambda_{k1-12} = \lambda_{k2-1} = \lambda_{k2-2} = \lambda_{k2-12} = \theta_2 = \theta_{k2} = \theta_{k2-2} = \phi = 0 \qquad (62)$$

Notice that the set of additional restrictions in equation 62 includes the additional restrictions from both the static CRC model, (equation 51), and the DCRE model, (equation 58). 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.

FIGURE C.1: PERCEIVED PRODUCTIVITY GAINS [STATIC CRC: $\beta + \phi\eta_i$]



C Additional Figures

C.1 Perceived Productivity Gains [Static CRC: $\beta + \phi\eta_i$]

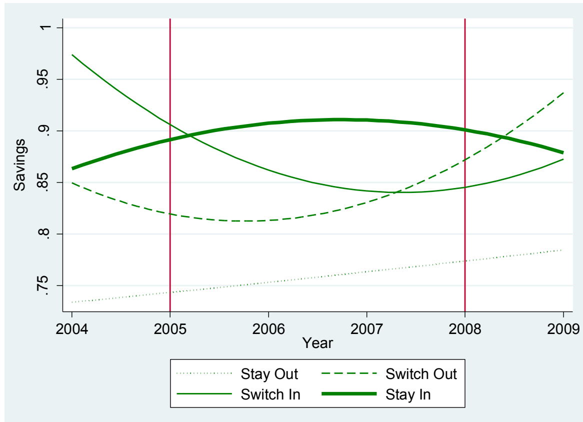
Figure C.1 repeats the exercise from Figure V for the static CRC model with both endogenous input and price controls corresponding to column 3 in Table X. Notice in this model perceived productivity gains will vary by entrepreneurship history, but not within entrepreneurship history over time. That is, 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 X, this variation across histories is statistically significant.

C.2 Savings, Borrowing, and Constraints Across Entrepreneurship Histories

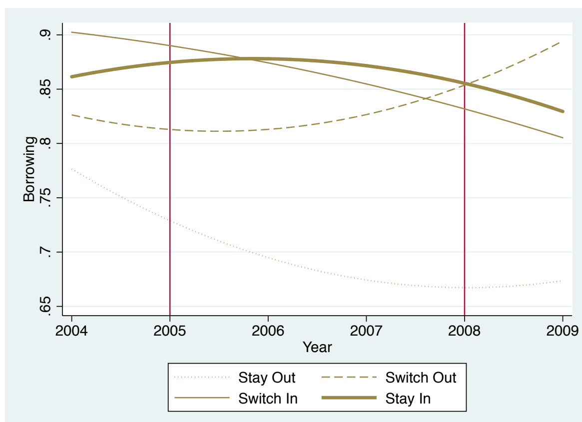
Panels A, B and C of Figure C.2 depict savings, borrowing, and self-reported financial constraints by enterprise histories, respectively. Panel A of Figure C.2 shows that though households that stay in have the consistently highest probability of positive savings and households that stay out the consistently lowest, households that switch in have high savings (indeed, higher than those that stay out) while households that switch out have lower savings. Furthermore, households that switch in have *decreasing* savings while households that switch out have *increasing* savings.

FIGURE C.2: SAVINGS, BORROWING, AND CONSTRAINTS BY ENTREPRENEURSHIP HISTORY

Panel A: Savings Over Time by Entrepreneurship History



Panel B: Borrowing Over Time by Entrepreneurship History



Panel C: Self-reported Financial Constraints Over Time by Entrepreneurship History

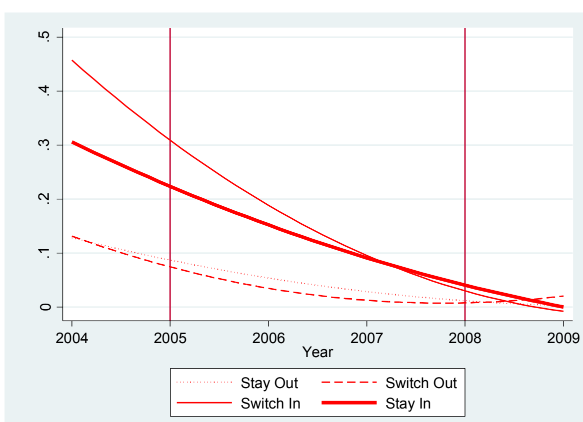
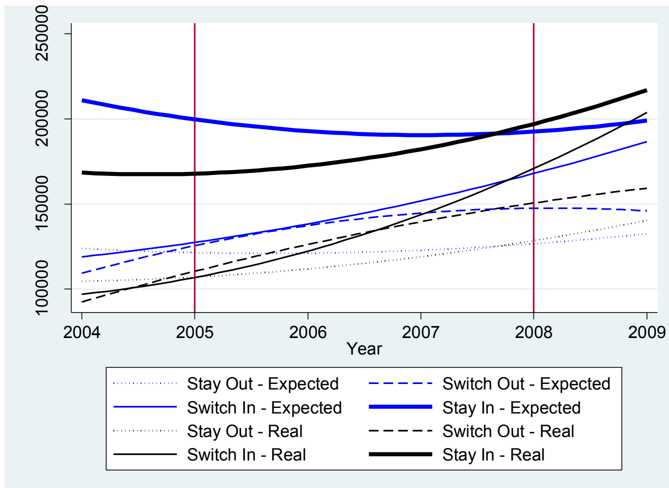


FIGURE C.3: COMPARISON OF EXPECTED AND REALIZED INCOMES



This evidence, similar to that in Tables II and III, is not in support of savings being a strong determinant of entrepreneurship.

Panel B shows that households that stay in do indeed have high probability of borrowing, but so do households that switch in, well before they engage in enterprise. In fact, households that switch into enterprise reduce their borrowing going forward despite switching into enterprise. On the other hand, households that switch out of enterprise start with a lower probability of borrowing, despite being engaged in enterprise, and increase their probability of borrowing as they switch out of enterprise. These patterns in borrowing across enterprise histories, similar to the patterns in savings, do not support the notion that access to or utilization of institutional borrowing is a major determinant of entrepreneurial entry or exit.

Panel C of Figure C.2 shows that indeed households that switch into enterprise seem to have a higher probability of financial constraints to start and the steepest decline in constraints. However, households that stay in enterprise also have high probability of constraints and a steep decline and households that stay out and switch out actually have the lowest probability of constraints in both periods. This evidence does not strongly support financial constraints as the primary determinant of entrepreneurial entry nor of exit.

C.3 Comparison of Expected and Realized Incomes

Finally, Figures C.3 and C.4 plot both expected net incomes and realized net incomes by enterprise history to facilitate a comparison of how expectations evolve with income realizations across households over time. Figure C.3 shows, as discussed above, that expectations track realizations fairly closely. However, it appears that households that switch out overestimate incomes and correct expectations steeply, with their expectations curve being more concave than their realized incomes curve. On the other hand, though households that switch in also overestimate incomes to start, their expectations curve stays convex, explaining their decision to still ultimately switch into enterprise. Households that stay out have weakly convex expectations and realized incomes but fairly flat slopes in both, explaining their decisions to not switch.

In Figure C.4 Panels A and B, I split the sample of households by maximum age in the household and reproduce the plots in Figure C.3. A learning model would predict that older households that have received more productivity signals would have more precise expectations of incomes and only switch sectors due to transitory, uninformative productivity shocks. On the other hand, younger households should be driving the patterns seen in Figure C.3. Indeed, Panel A shows that older households have precise expectations of income across all enterprise histories; while Panel B shows that younger households overestimate incomes by more than do older households and have steeper more convex curves, driving the patterns seen in Figure C.3.

D Additional Tables

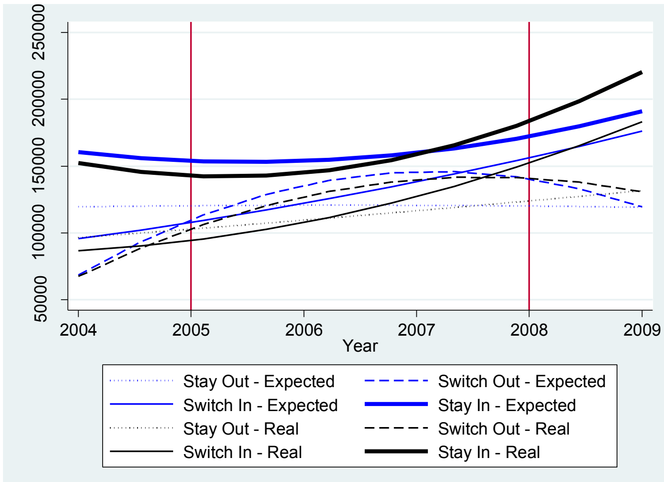
D.1 Preliminary Evidence

In columns 4-6 of Table D.1, I report results from the regression of savings, borrowing, 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 [48]) and its interaction with the global price of rice. Across both sets of regressions, I find that output price shocks significantly increase savings and significantly decrease borrowing and self-reported financial constraints. On the other

[48] 1 acre equals roughly 2.5 rai

FIGURE C.4: EXPECTED AND REALIZED INCOMES BY MAX AGE OF HH

Panel A: Above Median Average Age in Household



Panel B: Below Median Average Age in Household

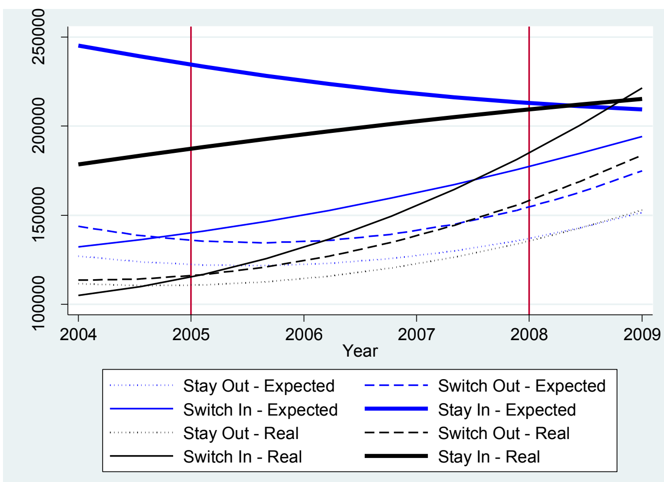


TABLE D.1: AGRICULTURAL PRICE AND SAVINGS

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

	Price x Farm Intensity				Price			
	Any Savings	Any Loans	Self-reported Constraints	Household Business	Any Savings	Any Loans	Self-reported Constraints	Household Business
Price x Farm Acreage	0.000532	0.0230***	-0.0418***	0.000582				
	(0.00782)	(0.00715)	(0.00855)	(0.00910)				
Price	0.0169***	-0.00720***	-0.0586***	0.00155	0.0165***	-0.00405**	-0.0674***	0.000744
	(0.00259)	(0.00236)	(0.00283)	(0.00301)	(0.00209)	(0.00192)	(0.00225)	(0.00244)
Farm Acreage	0.0484	0.0478*	0.213***	0.0813**				
	(0.0315)	(0.0288)	(0.0344)	(0.0366)				
Observations	11,039	11,038	11,039	11,039	11,040	11,039	11,323	11,040
Mean of Dep. Variable	0.798	0.783	0.166	0.457	0.798	0.783	0.166	0.457

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

hand, agricultural output price shocks do not significantly affect entrepreneurship.

Next, I use the price of rice and its interaction with household farm rai to instrument for the savings, borrowing, and constrained dummies in a household fixed effects instrumental variables regression of entrepreneurship on savings, borrowing and constraints, alternately. The results from these regressions are reported in Table D.2. Once again, I find no evidence of an effect of savings, borrowing, and/or financial constraints on entrepreneurship.

D.2 Labor

The model presented 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 greatly 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 char-

TABLE D.2: ENTREPRENEURSHIP DECISION
Effects Savings, Borrowing, and Constraints on Entrepreneurship (HH FE IV)

Saving	0.0973		
	(0.144)		
Borrowing		-0.0793	
		(0.362)	
Constrained			-0.0246
			(0.0365)
Farm Rai	0.0780***	0.0893*	0.0854***
	(0.0280)	(0.0491)	(0.0287)
First Stage - F Stat: Saving	22.09		
First Stage - p-value: Saving	< 0.0001		
First Stage - F Stat: Borrowing		12.98	
First Stage - p-value: Borrowing		< 0.0001	
First Stage - F Stat: Constrained			305.71
First Stage - p-value: Constrained			< 0.0001
Observations	11,039	11,038	11,039
Mean of Dep. Variable	0.457	0.457	0.457

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

acteristics of members of the household and is supplied inelastically across the two sectors in a fixed ratio to K. 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 D.3, 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 D.3, 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

Table D.3: 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 D.4: Changes in Labor Endowments

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

markets exist in this context, there is, at best, an imperfect market for enterprise labor.

Table D.4 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 D.5 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

TABLE D.5: LABOR ENDOWMENTS AND ENTREPRENEURSHIP

	Household Business	
	OLS	FE
Household Size	0.0170	0.00672
	(0.0109)	(0.0188)
# of Males	-0.0145	-0.0180
	(0.0149)	(0.0276)
# of Primary Educated	0.0616***	0.0138
	(0.0112)	(0.0184)
# of Unemployed, Inactive, In School	-0.0526***	-0.0207
	(0.0120)	(0.0167)
Observations	2,206	2,206
Mean of Dep. Variable	0.0456	0.0456

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

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 D.5, 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.

Table D.6: OLS and FE Estimates of Returns to Entrepreneurship

	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
Mean of Dep. Variable	11.71	11.71	11.71	11.71	11.71	11.71

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

D.3 OLS and FE

For the sake of comparison with the structural estimates from the preferred model presented in Tables V and VI, I present ordinary least squares and household fixed effects estimates of the average return to entrepreneurship. In Table D.6, 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 D.6 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 D.6, 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

TABLE D.7: REDUCED FORMS

	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).

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.

D.4 Reduced Form Coefficients

In Table D.7, I present the reduced form coefficients from which I estimate the structural parameters of the econometric models set forth above using minimum distance. In the reduced form specifications, I regress the log of total gross income from each period on the entrepreneurship dummies for each period, the log input expenditure for each period, and the appropriate interactions of these. 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.[49]

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

For an extension to 3 periods, please see the online supplement at:

`http://www.anantnyshadham.com/storage/Nyshadham_DCRC_Supplement_Nov2013.pdf`