THE DYNAMICS OF ENTREPRENEURIAL HUMAN CAPITAL

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Russell Toth

January 2012

THE DYNAMICS OF ENTREPRENEURIAL HUMAN CAPITAL

Russell Toth, Ph.D.

Cornell University 2012

There is broad international recognition of the importance of entrepreneurship for economic growth and development. Yet we still know relatively little about the entrepreneurial process. Existing economic theories are often based on constraints that are external to the entrepreneur and their enterprise, with financial constraints receiving perhaps the most attention. I propose that individuals' endowment and accumulation of entrepreneurship-specific human capital, which I term Entrepreneurial Human Capital (EHC), constitutes a crucial and largely overlooked factor in entrepreneurial dynamics. I build models that capture the relevant mechanisms, and test them using a unique panel dataset from Indonesia.

I begin by revisiting the evidence for financial constraints to household-level enterprise activity. I introduce a new test to the literature using semiparametric econometric methods, and additional tests based on positive wealth shocks. I uncover important heterogeneity in financial constraints, particularly in showing that financing is not the binding constraint to enterprise activity over most percentiles of the wealth distribution.

In the remaining two chapters I test variants of a model of dynamic EHC accumulation. First I exploit a unique natural experiment that provides exogenous assignment of relatively high-ability individuals into self-employment. I find that individuals who unexpectedly entered self-employment have remarkable occupational persistence, along with earnings dynamics, which are consistent

with a theory of EHC accumulation. I also exploit the exogenous nature of the occupational shock to provide causal evidence for an economically-significant value of entrepreneurial experience. Secondly, I develop a full dynamic theory of EHC accumulation through learning-by-doing. The theory explains the bifurcation between subsistence entrepreneurs, who get caught in an EHC poverty trap, and opportunity-oriented entrepreneurs. I test the model with panel data techniques.

Together the results provide complementary support for the concept of EHC and its importance for entrepreneurial dynamics. The results suggest that policymakers who intend to stimulate entrepreneurial activity should be put less of a focus on financing provision for low-skill individuals, and give greater attention to individuals with high entrepreneurial potential, particularly early in their careers.

© 2012 Russell Toth

ALL RIGHTS RESERVED

BIOGRAPHICAL SKETCH

Russell Toth was born on September 16, 1979, in Burnaby, British Columbia (BC), Canada, to Karl and Genevieve Toth. He spent his early years in Burnaby, BC. He attended Nelson Elementary School early in his elementary school years, and was involved with competitions for academically-talented students. In the third grade his family moved further east up the Fraser Valley, to Pitt Meadows, where Russell attended Meadowland Elementary through seventh grade, and then Pitt Meadows Secondary School (PMSS) for grades eight through twelve. At Meadowland Russell was invited to participate in an advanced mathematics program in seventh grade. At PMSS Russell was invited to take accelerated classes in a number of subjects, including English and Mathematics, and he won academic awards as the top student in Social Studies and awards for public service. Before graduation from high school, Russell was awarded an Outstanding Student Scholarship by the University of British Columbia, to pursue undergraduate studies in engineering. However, he declined the offer and chose to relocate to Hungary with his father, arriving in April of 1998.

Once in Hungary, Russell began learning the Hungarian language and attending classes informally at Deák Ferenc High School (a high school in Győr, Hungary, specializing in economics and accounting studies). After 1 1/2 years in Hungary, Russell enrolled in the equivalent of a college diploma program in accounting and economics at Deák Ferenc, which was conducted fully in Hungarian. Russell received 5/5 marks on all of his oral and written examinations at the end of the one-year program. For the following academic year Russell decided to switch to accounting studies from a leading country in the western accounting tradition, enrolling in a distance-learning undergraduate degree program administered by the University of London (with coursework designed

by the London School of Economics and Political Science (LSE)). After a year of independent study, Russell was examined in his four year-long subjects, and he received first-class honors in mathematical economics, and second- and third-class honours in the other three. After a year of the LSE program, Russell elected to return to North America to continue his undergraduate studies at Roberts Wesleyan College (RWC), starting in the fall of 2001.

After 2 1/2 years of further study, Russell completed a bachelor's degree summa cum laude in Humanities at RWC, with a concentration in communication and a minor in accounting, and with with college honors. Russell was awarded the RWC Alumni Association's Outstanding Student Award from his senior class. Having sparked an interest in making a difference in the world through working in international economic development during his undergraduate years and through living overseas, Russell moved on to a master's degree program in Mathematical Economics at Southern Illinois University-Carbondale in the fall of 2003. Russell spent two years in Carbondale preparing to apply for doctoral studies at a top economics PhD program. He was awarded a Master's Fellowship, and wrote a master's paper on foreign aid. In the spring of 2005, Russell accepted a Sage Fellowship to pursue a PhD in Economics at Cornell University. He was married to Sarah (nee Coleman) in June of 2005 in Greenville, Illinois, and together they moved to Ithaca, NY in July of 2005.

Russell began his doctoral studies at Cornell in August of 2005, and studied development economics, economic theory, and behavioral economics under the advisement of David Easley (chair), Christopher Barrett, Ted O'Donoghue, Viktor Tsyrennikov, and Kaushik Basu. He also had significant coursework in econometrics, and regularly attended research seminars in his research fields. Russell had the privilege of conducting overseas field research in Kenya and

Indonesia during his years as a doctoral student, and was involved in non-dissertation research projects on the economics of networks and the economics of pastoralism in East Africa. Russell successfully defended his doctoral dissertation in June of 2011, and submitted the final draft of his dissertation in January of 2012.

I dedicate this dissertation to the memory of my Dad, Tóth Károly. He more than anyone is responsible for my love of learning, my approach to scholarship, and my character as a human being. Through his two hands he made it possible to set me on the path to the type of career he could only dream about for himself. I know he would be happy for me.

ACKNOWLEDGEMENTS

My family and friends have provided tremendous support, encouragement, interest, patience and hospitality throughout my years as a doctoral student. My amazing wife, Sarah, has supported me the whole way through, through good times and bad. I also especially thank my mother, Genevieve Toth, and my father-in-law, Jerry Coleman, for showing special interest in my life as a PhD student, and prayerful support.

It has been a privilege to be mentored and influenced by so many wonderful scholars and colleagues at Cornell. David Easley has always given me just the right amount of space and intellectual freedom, while ingraining a professional ethic. Chris Barrett has read virtually everything I wrote, providing constant feedback and encouragement. Ted O'Donoghue has been a model of professionalism for me, while Viktor Tsyrennikov has forced me to think harder about my assumptions. Kaushik Basu inspired me as a development economist early in my time at Cornell. I am also grateful to have developed warm relationships with many of the staff at the Department of Economics. Finally, I am grateful to the many wonderful teachers who have influenced me throughout my entire academic career.

Max Mihm has been a friend and sounding board during my time as a PhD student, and I have learned a tremendous amount about economics from him. AV Chari helped walk me through my most difficult time as a PhD student. Ervin Starr has grown from an undergraduate professor of mine into a trusted friend, and a source of unique and deeply-informed perspectives on entrepreneurship.

I received a tremendous amount of help in pursuing and carrying out fieldwork for my dissertation, in Indonesia, in the summers of 2009 and 2010. I espe-

cially acknowledge Maria Wihardja and her family for their generous help and hospitality. Of those who provided direct help during my fieldwork I would like to particularly acknowledge Andre Syafroni, and Regina at KPPOD, the Regional Autonomy Watch. I also acknowledge the warm hospitality of colleagues at the Australian National University, the Institute of Technology at Bandung, and the University of Sydney while I was working on the dissertation project.

Many individuals provided formal and informal comments on the ideas and papers in this dissertation, far too many to list here, in discussions, at conferences, and after academic research presentations, in the United States, Canada, Indonesia, and Australia. I thank them for their insights and their patience as I was struggling to formulate my ideas.

I acknowledge financial support during the dissertation stage and in covering direct project costs from a Cornell Sage Fellowship, the Cornell University Institute for the Social Sciences, the Mario Einaudi Center for International Studies, the Cornell University Graduate School, and NSF Expeditions grant 0832782.

Finally, I thank God for providing me with the capability to pursue these studies, and the opportunities that have arisen because of my time at Cornell.

TABLE OF CONTENTS

1 Introduction **1**

2 The Indonesia Family Life Survey: Background, Key Variables, and Business Activity **10**
2.1 Introduction . . . 10
2.2 Overview of the IFLS . . . 11
2.3 Key Aspects of the Data Relevant to the Study of Occupational Choice and Microbusiness Activity . . . 20
2.4 Constructed Variables and Indices: Definition and Summary Statistics . . . 31
2.5 Descriptive Analysis of Enterprise Activity in the IFLS . . . 38
2.6 Conclusion . . . 46

3 Credit Misplaced? Testing for Household-level Financial Constraints to Enterprise Activity **47**
3.1 Introduction . . . 47
3.2 Descriptive Evidence: Wealth, Income and Enterprise Choices . . . 54
3.2.1 Opportunity and subsistence self-employment . . . 55
3.2.2 Heterogeneous earnings dynamics . . . 57
3.2.3 Heterogeneity in the returns to capital . . . 58
3.3 A Simple Model of Credit-Constrained Occupational Choice . . . 60
3.3.1 Outline of the Model . . . 61
3.3.2 Solution of the Model . . . 63
3.3.3 Predictions . . . 63
3.4 Data and Descriptive Evidence . . . 69
3.4.1 Indonesia Family Life Survey: General Background, Characteristics and Context . . . 70
3.5 Assets, Wealth and Entrepreneurial Choice . . . 72
3.5.1 Empirical Model . . . 73
3.5.2 Empirical Implementation . . . 77
3.5.3 Results . . . 78
3.6 Income Shocks and Entrepreneurial Choice . . . 81
3.6.1 Empirical model . . . 85
3.6.2 Empirical Implementation . . . 85
3.6.3 Correlational tests for credit constraints based on positive income shocks . . . 86
3.6.4 Exogenous financial shocks . . . 88
3.6.5 Selection and heterogeneity of effects . . . 89
3.7 Additional Factors in Entrepreneurial Selection and Choice . . . 93
3.7.1 Familial Effects . . . 94
3.7.2 Gender . . . 97

3.7.3 Risk and time preferences . . . 98
3.7.4 Cognitive ability . . . 99
3.8 Conclusion . . . 100

4 Entrepreneurial Human Capital and Entrepreneurial Dynamics: Evidence from a Natural Experiment 103
4.1 Introduction . . . 103
4.2 The Setting, and Entrepreneurial Human Capital . . . 108
4.2.1 Indonesia Background . . . 108
4.2.2 Entrepreneurial Human Capital . . . 112
4.3 A Simple, Dynamic Model of Entrepreneurial Selection, Savings, and Consumption . . . 113
4.3.1 Basic Properties of the Model . . . 117
4.3.2 The Effects of Exogenous Shocks to Occupational Choice Incentives . . . 120
4.3.3 Testable Predictions and Alternative Theories . . . 120
4.4 Design of the Study and Preliminary Evidence . . . 123
4.4.1 Data . . . 123
4.4.2 Preliminary Evidence . . . 125
4.5 Identification Strategies and Empirical Specifications . . . 126
4.6 Estimation Results: Self-employment Persistence and Returns . . . 135
4.6.1 Self-employment Persistence . . . 136
4.6.2 The Dynamics of Self-Employment Returns . . . 137
4.6.3 Selection-corrected earnings dynamics . . . 138
4.7 Alternative Explanations and Further Evidence . . . 140
4.7.1 Capital Stock Lock-in . . . 140
4.7.2 Optimal Industry Selection . . . 141
4.7.3 Changes in Inputs . . . 141
4.8 Conclusion . . . 142

5 A Dynamic Model of Occupational Choice and Entrepreneurial Human Capital 146
5.1 Introduction . . . 146
5.2 Dynamic Model of Entrepreneurial Capital Accumulation . . . 153
5.2.1 Environment and Primitives . . . 156
5.2.2 Bellman Equation . . . 163
5.2.3 Analysis . . . 164
5.3 Identifying the Formation of and Returns to EHC . . . 169
5.3.1 Framework for Distinguishing Innate and Accumulated EHC . . . 171
5.3.2 Identifying EHC in a Panel Regression Framework . . . 173
5.3.3 Description of the Data . . . 176
5.3.4 Results . . . 178
5.3.5 Summary . . . 180

5.3.6 Additional Evidence 182
5.4 Conclusion 184

A Appendix to Chapter 2 **188**
A.1 Overview of Data Available in the IFLS 188
A.1.1 Individual-level data 188
A.1.2 Household-level data (including enterprise data) 189
A.1.3 Community and facility-level data 190
A.2 Figures 192
A.3 Tables 199

B Appendix to Chapter 3 **212**
B.1 Figures 212
B.2 Tables 221

C Appendix to Chapter 4 **244**
C.1 Tables 244

D Appendix to Chapter 5 **251**
D.1 Additional Material 251
D.1.1 Evidence for labor market frictions in EHC transmission . 251
D.1.2 Proofs: The properties of the value function 252
D.2 Figures 256
D.3 Tables 257

Bibliography **264**

LIST OF TABLES

A.1 Summary Statistics on IFLS Rounds . . . 199
A.2 Geographic distribution of firms in the IFLS . . . 200
A.3 Summary statistics on firms, 2008, firms with no employees . . . 201
A.4 Summary statistics on firms, 2008, firms with only family/unpaid employees . . . 202
A.5 Summary statistics on firms, 2008, firms with waged employees . 203
A.6 Summary statistics on enterprise experience . . . 204
A.7 Summary statistics on individuals who are self-employed in the panel . . . 205
A.8 Variance-covariance matrix for key enterprise variables, 2008 . . 206
A.9 Variance-covariance matrix for key enterprise variables, 2008, cont.207
A.10 Variance-covariance matrix for key enterprise variables, 2000 . . 208
A.11 Variance-covariance matrix for key enterprise variables, 2000, cont.209
A.12 Variance-covariance matrix for household variables and enterprise activity . . . 210
A.13 Variance-covariance matrix for household variables and enterprise activity, cont. . . . 211

B.1 Percentiles of wage- and self-employment returns, IFLS, 2008 . . 221
B.2 Wealth and propensity to start enterprise . . . 222
B.3 Wealth per head and propensity to start enterprise . . . 223
B.4 Wealth (in logs) and propensity to start enterprise . . . 224
B.5 Wealth per head (in logs) and propensity to start enterprise . . . 225
B.6 Wealth and startup capital . . . 226
B.7 Wealth per head and startup capital . . . 227
B.8 Wealth and current capital . . . 228
B.9 Wealth per head and current capital . . . 229
B.10 Probit: Effects of income flows on post-2000 startup . . . 230
B.11 Probit: Effects of income flows on post-2000 startup, Part 1 231
B.12 Probit: Effects of income flows on post-2000 startup, Part 2 232
B.13 Probit: Effects of income flows on post-2000 startup . . . 233
B.14 Probit: Exogenous shocks . . . 234
B.15 Effects of income flows on post-2000 startup capital . . . 235
B.16 Effects of exogenous income flows on post-2000 startup capital . 236
B.17 Effects of income flows on post-2000 startup capital, quantile reg. 237
B.18 Effects of exogenous income flows on post-2000 startup capital, quantile reg. . . . 238
B.19 Effects from Unconditional Cash Transfers (UCT), without selection controls . . . 239
B.20 Effects from Unconditional Cash Transfers (UCT), with selection controls . . . 240
B.21 Parent Cross Tab . . . 241

B.22 Parent Effect 242
B.23 Parent Return Effect 243

C.1 Summary statistics on individual entrants 244
C.2 Summary statistics on community-level sources of variation . . . 245
C.3 Panel wage regression 246
C.4 Change in propensity to stay in self-employment after Asian Financial Crisis, 1998 entrants 247
C.5 Change in propensity to stay in self-employment after Asian Financial Crisis, 1999 entrants 248
C.6 Returns to experience, individuals entering self-employment during 1998 249
C.7 Returns to Experience, Individuals Entering Self-Employment During 1999 250

D.1 Evidence for labor market frictions: Dependent variable labor quantity, IFLS4 (2008), OLS 257
D.2 Log-linear fixed effect experience regressions; dependent variable monthly Net Profit 258
D.3 Cobb-Douglas fixed effect experience regressions; dependent variable monthly Net Profit 259
D.4 Cobb-Douglas fixed effect experience regressions; dependent variable monthly Net Profit 260
D.5 Log-linear fixed effect experience regressions, common experience; dependent 261
D.6 Log-Linear fixed effect experience regressions, common start age; dependent variable monthly net profit 262
D.7 The propensity of the household to engage in running an enterprise, 2001-2008, probit 263

LIST OF FIGURES

A.1 Map of Indonesia, highlighting IFLS provinces 192
A.2 The distribution of firms by industry, 2000 and 2008 193
A.3 Household wealth distribution, in logs, year 2000 194
A.4 The relationship between wealth and the propensity to engage in enterprise activity, urban households parsed by wealth percentile. 195
A.5 The relationship between wealth and the propensity to engage in enterprise activity, rural households parsed by wealth percentile. 196
A.6 The relationship between years of experience and the propensity to engage in enterprise activity, 0-50th percentile of wealth distribution 197
A.7 The relationship between years of experience and the propensity to engage in enterprise activity, 0-99th percentile of wealth distribution 198

B.1 Plot of net profit against experience in three enterprise types . . . 212
B.2 The relationship between (lagged) wealth and the propensity to engage in enterprise activity 213
B.3 The relationship between (lagged) wealth per head and the propensity to engage in enterprise activity 214
B.4 The relationship between (lagged) wealth in logs and the propensity to engage in enterprise activity 215
B.5 The relationship between (lagged) wealth per head in logs and the propensity to engage in enterprise activity 216
B.6 The relationship between (lagged) wealth and startup capital . . 217
B.7 The relationship between (lagged) wealth per head and startup capital 218
B.8 The relationship between (lagged) wealth and current capital . . 219
B.9 The relationship between (lagged) wealth per head and current capital 220

D.1 EHC learning function 256

CHAPTER 1

INTRODUCTION

There is growing recognition the world over that vibrant entrepreneurial activity is crucial for economic growth and development. This importance is particularly salient during a time of globalization, when increased competition puts pressure on enterprises at the same time that greater opportunities also present themselves. This level of attention makes our relative lack of knowledge of the entrepreneurial process all the more surprising, and the relative ineffectiveness of numerous policy interventions all the more troubling ([Shane, 2008], [Lerner, 2009]). A central component of policy design meant to stimulate (or at least not inhibit) entrepreneurship is an understanding of the key factors behind the entrepreneurial process.[1] We generally assume that the absence or impediment of certain key factors in entrepreneurs, in their enterprises and externally, will constrain entrepreneurial activity. This attention to constraining factors makes it crucial to understand what the core factors are and how to influence them. In this dissertation, I focus on the role of a key *internal* constraint to entrepreneurship, the accumulation of entrepreneurship-specific skill. I point out that this factor has received relatively less attention in the formal academic literature, and present evidence for its crucial importance in understanding entrepreneurial activity and dynamics.

Existing entrepreneurial and business policy often takes a static view of en-

[1] In this dissertation I employ a rather broad definition of *entrepreneurship*, including most activities related to the startup and ongoing operation of enterprises. However, it will be important throughout to distinguish *subsistence* and *opportunity-oriented* entrepreneurship. The former refers to the observation that many individuals become self-employed not to pursue a meaningful business opportunity, but rather because of low opportunity cost (in particular, lack of wage employment opportunities). The latter refers to individuals who start a business to pursue a meaningful entrepreneurial opportunity.

trepreneurs and their firms. Internal capabilities are taken as given, rather than allowing for a dynamic process of entrepreneurial development. This static approach often drives a search for short-term solutions, and leads to policies focused on singular, external constraints to enterprise activity. The external factor that has received the most focus is financing constraints. Such static approaches parallel the most common academic approaches to the conceptualization of entrepreneurship. The dominant thinkers in economics on entrepreneurship – in particular Schumpeter, Knight, Kirzner and their more mathematically-inclined followers – primarily frame entrepreneurship as an event, seeing the entrepreneurial tendency as being derived from a fixed trait such as risk preferences or alertness to opportunity. There is relatively little attention, particularly in the academic economics literature, to the process of entrepreneurial learning dynamics.[2]

In this dissertation I take some steps in the direction of formalizing and providing empirical evidence for a stock of accumulated entrepreneurial skill, which I term Entrepreneurial Human Capital (EHC). EHC constitutes specialized, entrepreneurship-specific skills, knowledge and other human capital endowments, such as skills in sales, negotiation, product development, risk judgment ([Shane, 2003]), and entrepreneurial social capital. Regardless of one's innate talent, the accumulation of EHC is crucial to entrepreneurial outcomes. I argue that EHC (1) constitutes a central factor behind entrepreneurial activity, (2) that the dynamics of EHC accumulation are crucial to understanding entrepreneurial dynamics more broadly, and (3) that certain stylized facts about entrepreneurial activity that have been interpreted as evidence for other con-

[2] A notable exception to the lack of attention to entrepreneurship-specific human capital is the work of work of Schultz (e.g., [Schultz, 1980]), but his work has received relatively little mainstream attention. Much more recently [Bruhn *et al.*, 2010] points out the lack of attention to entrepreneurship-specific human capital in the economics literature.

straints (financing, in particular) might in fact be driven primarily by EHC.

Throughout the dissertation I build models in the traditional occupational choice framework, which captures the incentive to self-select into self-employment with wage employment as the opportunity cost. I extend the standard framework, where appropriate, to incorporate EHC accumulation. While such applied modeling constitutes a contribution to the theoretical modeling of entrepreneurship, it is not of great independent theoretical interest. Rather, the central contribution of the dissertation is to subject the models to empirical testing.

A probable important factor behind the lack of understanding of the entrepreneurial process is the lack of good data on entrepreneurial dynamics. Though most economists recognize entrepreneurship as a crucial economic activity, it is notoriously difficult to study empirically in part because data acquired prior to important entrepreneurial events (hence allowing before-after comparisons and analysis of dynamic processes) are difficult to come by. Since entrepreneurial activity is relatively rare, a large-population dataset is needed in order to have sufficient sample size to identify a sufficient number of entrepreneurs from the relatively small proportion available in a population sample. Unfortunately, datasets with sufficiently large cross-sections tend to have little information relevant to entrepreneurship, because entrepreneurial activity is almost always of secondary importance to organizations with the wherewithal to conduct surveys with relatively large sample sizes.

I employ a unique, large-scale, household panel dataset from Indonesia, the Indonesia Family Life Survey (IFLS), which enables me to significantly overcome these core data challenges. The IFLS provides a large cross-section,

close to 10,000 households, observed across 4 panel rounds over 15 years. Because self-employment activity is more pervasive in developing countries, self-employment activity received an unusual amount of attention in the IFLS survey questionnaires, particularly in the two most recent rounds. Due to the uniqueness of the data on an international scale, I argue that the findings are more broadly instructive, while at the same time the analysis provides a number of contributions to the relatively shallow literature on developing-country entrepreneurship.

The analysis consists of 4 chapters that can be organized into three main parts. First, because the dissertation is focused on a single dataset, the IFLS, Chapter 2 provides a reasonably comprehensive overview of the IFLS. Chapter 2 is particularly focused on issues relevant to this dissertation – enterprise activity and employment activity. Second, Chapter 3 revisits the evidence on financial constraints to household-level enterprise activity in the developing-country setting. These findings frame the remainder of the dissertation by pointing to important unexplained variation in entrepreneurial choices and outcomes that cannot be accounted for by financing constraints alone. Finally, the two subsequent chapters (Chapters 4 and 5) provide two approaches to modeling and empirical testing for EHC as a relatively unexplored channel to explain variation in enterprise dynamics and outcomes.

More specifically, in Chapter 3 I provide new evidence on financial constraints to household-level enterprise activity in the developing-country setting by conducting two classes of tests of financial constraints. The most basic model of occupational choice in which entrepreneurial activity is potentially constrained by financing predicts a positive relationship between wealth

and enterprise activity, which is most pronounced for less-wealthy households. First, I study the relationship between lagged household wealth, enterprise startup and capital allocation activity, employing a flexible, semiparametric, discrete-response framework that improves on the existing applied literature. I find little relationship between wealth and enterprise activity over the lower 50 percentiles of the wealth distribution, which provides initial evidence that financing is not the key constraint to enterprise activity for less wealthy households.

Second, I study the relationship between positive income shocks (both expected and unexpected) and enterprise activity. If households are credit constrained then access to such cash flows should serve to alleviate the credit constraints, generating correlation between income flows and entrepreneurial choices and outcomes. Yet I find that such positive income shocks, even unexpected ones, show relatively little relationship to enterprise choices, and in any case the noticeable effects are relatively concentrated amongst wealthier households. This again suggests that financing constraints are not the primary driver of enterprise outcomes. I close Chapter 3 with some exploratory evidence on other potential factors behind the variation in enterprise outcomes.

To move from arguing that financing constraints might not be the key driver of entrepreneurial dynamics for most individuals and households, to convincingly arguing that EHC is the key explanation, I need to grapple with two key challenges: difficulties in directly measuring EHC (particularly in the observational data I have available in the IFLS), and the fact that human capital accumulation involves individual choices that introduce the potential for endogeneity in occupational selection and mistaken inferences based on proximate factors.

The remaining two chapters grapple with these challenges through two empirical approaches. First, by exploiting a unique natural experiment that approximates the ideal experiment one might wish to conduct to identify the effects of EHC, and second through using dynamic panel and mathematical modeling techniques to impose useful structure on the data in order to draw inferences.

In Chapter 4 of the dissertation I study entry and persistence in self-employment amongst a cohort of individuals in Indonesia who enter during the 1997-98 East Asian financial crisis. If we entertain the thought experiment regarding what the ideal experiment would be to identify the effects of EHC, it involves randomized assignment of EHC to individuals. While such an experiment is clearly not feasible, perhaps the next best thing is to identify an exogenous source of selection into self-employment, which effectively "exogenously assigns" EHC to individuals through forced learning-by-doing.

I argue that the 1997-98 East Asian crisis provides this kind of natural experiment. In contrast to the effects of large, unexpected shocks in other countries, the crisis in Indonesia seems to have largely pushed individuals closer to the upper end of the wage distribution (i.e., those working in formal sector, higher-skill jobs) into informal self-employment. This turns out to provide a unique setting in which to test the predictions of a theory of EHC that allows for correlation between ability in waged employment and self-employment, as individuals who would otherwise not get exposure to self-employment are forced to do so. The sub-population is of particular interest because it would tend to include a greater proportion of individuals with higher entrepreneurial potential, who might not otherwise obtain entrepreneurial experience due to relatively high opportunity costs to initial experimentation with self-employment.

In order to test the theory and distinguish it from alternatives, I carry out analysis on two important outcomes for the new entrants. First, I study entrepreneurial persistence – the propensity of individuals who enter during the crisis to remain in self-employment in the years after. A static perspective toward entrepreneurship would predict that such individuals should re-integrate into the workforce soon after the crisis ends, as their static human capital endowment must be better suited to wage-sector employment, and all that is needed is for sufficient demand to return in the formal wage-earning sector. What I find is that self-employment is highly (and as I show with more rigorous econometric methods, robustly) persistent amongst crisis-period entrants – 78% remain in self-employment 10 years later, when the average observed 10-year persistence rate is 46%. While this finding is initially surprising, it is less so in light of a theory of EHC accumulation. Under such a dynamic theory the accumulation of a stock variable like EHC could change long-run occupational choice incentives. Furthermore, if we assume a concave learning curve, the marginal choice effects would be strongest for individuals with little prior experience.

In order to further distinguish the theory of EHC, I secondly study earnings dynamics amongst the crisis-period entrants and again I find evidence that is most consistent with the theory of EHC accumulation.

In Chapter 5 I build on evidence from both the preceding chapters and from the emerging literature outside economics on the dynamics of entrepreneurial learning to develop a dynamic microeconomic theory of entrepreneurial human capital accumulation through learning-by-doing. This theory can be seen as building on theoretical work such as [Buera, 2009], extending the canonical

dynamic model of occupational choice under credit constraints to include a dynamic theory of entrepreneurial human capital accumulation. I first develop the formal dynamic programming model, with the key mechanism being the nature of the learning process, in which the accumulation of EHC is an S-shaped function of a measure of the "complexity" of the enterprise that an individual is running. I show that this mechanism qualitatively predicts the stylized findings in the data – low-ability, low-wealth households will operate simple enterprises and hence get "stuck" running simple enterprises, while higher-potential individuals get on a dynamic path of EHC accumulation that leads to intermittent financing constraints.

I then carry out quantitative analysis of the model. I employ panel data techniques on the full IFLS dataset to study the relationship between earnings and various measures of entrepreneurial experience. After using various techniques to check robustness, I confirm a stable and economically-significant positive earnings dynamic that is particularly strong among individuals running relatively more complex enterprises and hence consistent with the theory.

Taken together, these dissertation chapters provide varied evidence for the mechanisms behind entrepreneurial human capital accumulation, and demonstrate their central importance in explaining entrepreneurial dynamics. These chapters demonstrate important nuance in parsing out different cohorts of the self-employed, with a large, relatively poor cohort of individuals having little access to sources of entrepreneurial skill acquisition. This points to a "poverty trap" in entrepreneurial skill affecting much of the population.[3] In constrast to

[3]To my knowledge this particular form of poverty trap mechanism is novel, though it is related to important literature in the economics of development. A couple of seminal papers ([Banerjee & Newman, 1993], [Azariadis & Drazen, 1990]) generate poverty traps out of factors external to the individual agent – individual credit constraints, and social externalities in human capital, respectively. [Galor & Zeira, 1993] incorporates individual-level human cap-

those facing a poverty trap, the relatively select group of higher-potential individuals have a greater tendency to be constrained by financing, yet still depend crucially on the dynamics of entrepreneurial learning.

Such ideas are only beginning to receive serious attention in the broader academic and policy communities, and they have fundamental implications for policy design, particularly in developing countries. This is particularly true in light of emerging evidence that draws into question the overall effectiveness of microfinance as a policy intervention meant to overcome credit access constraints ([Banerjee *et al.*, 2009], [Karlan & Zinman, 2010]). The research in this dissertation suggests that (1) credit access might not be the most pressing constraint to enterprise entry and development for most potential entrepreneurs, and (2) policies must be tailored to the constraints faced by particular groups, with greater attention to the nature of the entrepreneurial development process. Efforts to reduce and remove individual, external constraints on an ad hoc basis have often been less effective than expected, and a strong potential explanation is that the dynamics of entrepreneurial development are more important than particular static constraints.

ital costs as part of the explanation, though external constraints are still needed to generate traps, particularly in the long-run. The closest to the work in this dissertation is perhaps in [Dasgupta & Ray, 1986], which formalizes nutritional traps, as deficiencies in internal abilities perpetuate exactly because such deficiences make independent advancement more difficult.

CHAPTER 2

THE INDONESIA FAMILY LIFE SURVEY: BACKGROUND, KEY VARIABLES, AND BUSINESS ACTIVITY

2.1 Introduction

In this Chapter I provide an introduction to the main dataset that will be employed in the remaining chapters of this dissertation, the Indonesian Family Life Survey (IFLS). The Chapter has four main objectives.

First, I provide background information and context for the dataset, discussing the data collection process and the reliability of the data. Secondly, I describe the raw variables that are available in the IFLS, in particular in relation to the key issues in this thesis: enterprise activity and labor market employment activity, with particular interest in self-employment activity. I then proceed to provide some descriptive information on the raw variables, including summary statistics. Third, I go on to discuss the construction of variables that will be used in later chapters of the thesis, in particular the measures of wealth, self-employment experience and education that I construct. I then provide summary statistics on these constructed variables. Fourth, I provide some preliminary descriptive analysis of self-employment and enterprise activity in the IFLS. I compute variance-covariance matrices that summarize correlations between the main variables in the dataset and I also provide some selected non-parametric descriptive analysis of some of the key bivariate relationships.

In this Chapter I argue that the IFLS is ideally suited for the primary research questions in the dissertation, which are focused on understanding the key fac-

tors behind the emergence and growth of enterprises at the household level. This suitability is due both to the quality of the data collected and the breadth of information available on micro- and small-business activity. This Chapter additionally contributes to our knowledge of household-level enterprise activity, by presenting descriptive evidence from a large, representative sample from an important developing country, Indonesia. While some of the findings in this Chapter are consistent with earlier literature, they also frame the work in the later chapters that supplements or even contradicts existing evidence in the literature.

2.2 Overview of the IFLS

The IFLS has been collected as a longitudinal household and community survey project in Indonesia. There have been data collection rounds in 1993-94, 1997 (with a minor round on 25% of the sample in 1998 in response to the East Asian crisis), 2000-01 and 2007-08.[1] Round-by-round statistics on sample size are summarized in Table A.1. The original 1993 round of the survey (IFLS1) targeted data collection on 7000 households, and obtained just over 7200. It was designed to be representative of 83% of the Indonesian population, following a cluster sampling method of selecting individual villages or regions of cities with urban over-sampling, and then selecting households from population rosters.

[1]Various organizations have been involved in designing and collecting the IFLS: Lembaga Demografi, University of Indonesia (IFLS1-IFLS2), RAND (IFLS1-IFLS4), UCLA (IFLS2), the Population Research Center, University of Gadjah Mada (IFLS3), the center for Population and Policy Studies of the University of Gadjah Mada (IFLS4) and Survey METRE (IFLS4).

There is an extensive list of funding bodies, including USAID, the WHO, the NIA, and the UN; the complete list see: http://www.rand.org/labor/FLS/IFLS/teamfund.htm

For more details on survey design, see [Strauss *et al.*, 2004], [Frankenberg & Thomas, 2000], and [Frankenberg & Karoly, 1995].

The survey has 313 initial locations in 13 of 27 provinces of the country. A map of Indonesia, with the IFLS provinces indicated with a darker shade, is provided in A.1. [2] It generally covers the wealthier provinces in the country, which serves both to generate high population coverage (since the wealthier provinces have higher population densities) and simplify the data collection process by operating in locations with better infrastructure. The survey covered and continues to cover all of Java (the most populous island and economic center, which includes the Special Capital City District of Jakarta), four provinces in Sumatra, part of west Nusa Tenggara (including Bali) and the provinces of South Sulawesi and South Kalimantan.[3]

Subsequent rounds have involved resampling the original households, and then sampling split-offs from the original households where possible. Split-offs might occur due to marriage, coming of age and moving out, or other family breakup. Attrition has been relatively minor, at around 5% between rounds,[4] which matches or improves upon the attrition rates in longitudinal surveys in OECD countries, and reduces concerns about bias from non-random attrition. The tracking effort has involved diligently following household splits outside of the original sample provinces when possible. New households that do not root in the original IFLS1 sample households are not incorporated into new rounds. As of the 2000 round (IFLS3) the dataset covers just over 10,000 households, and over 15,000 by the 2007-08 round (IFLS4).

[2]The public use version of the IFLS does not allow villages to be identified by name, in order to protect respondent privacy. Province identifiers are provided for each village.

[3]There has been discussion of extending the geographical scope of the IFLS in a new data round in 2012 or 2013, perhaps extending as far east as Papua.

[4]The exact attrition statistics are given in [Strauss *et al.*, 2009]. In IFLS2 94.4% of IFLS1 households were re-contacted, while in IFLS3 the re-contact rate of IFLS dynasty households was 95.3%. In IFLS4 the re-contact rate of IFLS1 dynasty households was 93.6%, though the re-contact rate on target households was 90.6% (i.e., some IFLS1 dynasty households led to more than 1 sample household, of which all might not be successfully contacted). Overall 90.3% of IFLS dynasty households appear in all four waves, or died along the way.

Data are collected at the individual, household, and community level and these three sources can be matched together. As noted, the original randomization occurs at the level of selecting villages or portions of cities within the survey provinces. Community-level data are collected from key informants in the village, and cover issues from infrastructure to social services like health and educational institutions to key events in the location such as natural disasters.

The second level of randomization involves selecting households within the locations. Once households are selected, they are invited to participate in an extensive household survey. There are various portions of the survey, one part of which pertains to the household itself (generally directed to the household head), and then sections directed toward adult members of the household, women who have experienced pregnancy, and children under the age of 15. On average about 3 people per household are individually interviewed. The surveys are quite extensive and may take 1-2 days to complete at the household level. The individual surveys contain rich information on a plethora of individual and household characteristics, including education, employment, income and assets, migration, health, and individual and household shocks. Since the full breadth of the survey is beyond the scope of this dissertation, I will focus below on the sections that are relevant to the analysis in the subsequent chapters.

Brief Historical Context

The IFLS has been collected during a period of economic change and political upheaval in Indonesia. When the survey was first initiated, Indonesia had gone through a period of relatively steady economic growth, with per capita income

rising more than fifteenfold from the 1960s until the early 1990s. It appeared that Indonesia was ready to join the group of middle-income countries. Then in 1997 a major currency crisis spread throughout south-east Asia, with Indonesia arguably the worst-hit nation. GDP contracted 12-15% in 1998, a shock on the order of the Great Depression in the United States. The formal financial system was especially hard-hit, with many banks going insolvent.

In the years after the crisis Indonesia went through major political and economic restructuring. President Suharto's resignation in May, 1998 led to democratic elections and policy changes on a number of fronts, including industrial policy and banking. A major decentralization of political powers first came into force in 2003. The effects of this policy are still being studied, but there seem to have been geographically variable outcomes in terms of local governance effects, with effective leaders emerging in some regions and less desirable outcomes in others.

In the most recent years Indonesia has returned to a path of steady economic growth of approximately 6% per annum. Though it falls outside the scope of the IFLS, the country seems to have weathered the 2008 Global Financial Crisis relatively well, managing to avoid the massive negative shock experienced in most countries. Observers have suggested that this success can be attributed to the reforms introduced after the previous crisis.

The IFLS is ideally placed to provide rich micro-evidence on household outcomes during a time of change. Such drastic economic and policy changes can also be exploited as sources of variation of key variables of interest, and indeed such an empirical approach will be pursued in subsequent chapters in this dissertation.

The Data Collection Process and Reliability of the Survey

The IFLS has been designed, collected and cross-checked at high international standards, especially in comparison to many longitudinal developing-country datasets of its time.[5] The original Principal Investigator was Paul Gertler, currently Li Ka Shing Distinguished Professor of Economics at UC-Berkeley Haas School of Business. A number of survey design experts at RAND Corporation and Lembaga Demografi, University of Indonesia, contributed to the questionnaire design and implementation, along with fieldwork and data entry. There has been a steady stream of papers that make use of the IFLS over the years.[6]

Preparation for an individual survey round takes 1-2 years. This involves updating the content of the survey, field testing, arranging logistics, and hiring and training the survey team. For the most recent rounds the fieldwork has been based at SurveyMETRE in Yogyakarta. SurveyMETRE is a professional surveying company that was formed by Indonesia-based experts who had been hired as consultants on previous ILFS rounds. Surveyors are generally recent college graduates from top universities (indeed, Yogyakarta is home to some of the top universities in Indonesia). The potential surveyors are over-recruited, to allow for culling down to select the best performers during the training process, and to provide a pool of backup surveyors should problems arise with the chosen team.

At the household level surveys are conducted by a pair of surveyors, typically with one team member leading the discussion, and the second team

[5]Much of the discussion in this section is based on a conversation with Bondan Sikoki, the field director of many of the IFLS rounds and the Director of Survey METRE.

[6]For a full public listing, see: http://www.rand.org/labor/FLS/IFLS/papers.html. Looking at just the most recent 5 years, for example, the page lists 7 working papers or publications from 2011, 16 from 2010, 20 from 2009, 12 from 2008, and 10 from 2007.

member recording the data. A group of surveyor pairs is dispatched to each province, with a head surveyor overseeing the survey efforts in the region. In the most recent rounds the data collection process has been converted to an electronic format, with new field data immediately uploaded to a central data compilation hub in Yogyakarta. This live process of data transmission allows initial data verification procedures to be carried out while the survey team is still in the field, hence tremendously lowering the cost of re-collecting data when anomalies arise. Individual respondent households are paid for their time for participating in the survey. This has generally been an in-kind payment (e.g., a clock); in the most recent round the value was approximately 40,000 Indonesia rupiah per household (about $4 US).

The survey is generally carried out in the 313 target villages first. Then further tracking is carried out within-province and even in provinces outside the original scope of the IFLS. While this effort is very diligent, which is reflected in the remarkably low attrition rate in the survey, there is inevitable attrition. Attrition can still result from failure to locate households, mortality, and refusal to respond to the survey.[7]

In spite of its reputation as a top-quality developing country panel dataset, the IFLS does have its weaknesses. No matter how careful the survey, measurement error is inevitable, especially in a survey that collects self-reported, recall data that spans decades in some cases. One of the greatest strengths of the IFLS, its comprehensiveness, can also be a weakness when it leads to respondent fatigue and a greater chance of erroneous or inaccurate response.[8] An

[7] Anecdotally I know that there was high respondent refusal amongst Chinese respondents during IFLS4 in North Sumatra, where much government mistreatment has been occurring in the years since decentralization.

[8] I have heard anecdotal reports to this effect, namely that some respondents can become quite exhausted by the end of the interview sessions.

additional great strength of the IFLS, the sample size, can also become a key weakness, due the need for greater decentralization of the survey effort. This means that individual survey teams operate with limited direct supervision, leading to questions about overall effort and diligence.[9] These issues are common to such survey projects and are probably less severe than elsewhere, yet they are worth keeping in mind.

The IFLS and the Study of Enterprise Activity

There are a number of reasons why the IFLS is promising as a data source to look at household-level enterprise activity, as opposed to other data sources that one might consider. First, there is often a selection problem present in firm-level data, with selection biased toward successful firms, and selection biased toward formal-sector firms. This is often because firm-level datasets do not include information on enterprises as of some minimum size cutoff point, for example firms with less than 20 employees as is the case with the primary manufacturing dataset available in Indonesia. In principle, the IFLS includes information on all businesses started at the household level, regardless of survival trajectory.

Second, it is difficult or impossible to look at the *entry* process with firm-level data, especially if one wants to have information on the individual entrepreneur prior to entry. The IFLS allows for many connections to be drawn between individual and household characteristics and enterprise activity, including when the business is not even in operation. These connections are particularly sharp in the case of household enterprises, where there is almost always one primary

[9] I have heard anecdotal reports both about significant variation in surveyor effort, and a significant incident in IFLS4 in which some data were falsified. Fortunately the latter incident also became a proof of the survey's verification safeguards, as the problematic data were identified, removed and re-collected.

individual in charge of the enterprise, so that inferences on individual characteristics and choices can be clearly drawn out.

Third, the panel structure of the data allows for the inclusion of fixed effects in panel regressions, to account for unobserved heterogeneity in location and in time. Finally, the dataset is large enough (in the cross-sectional sense) that the subset of individuals participating in non-farm enterprises is significant, as summarized in Table A.2. There we see that a large number of enterprises are reported in the dataset, with over 30% of households operating at least one enterprise in each survey round.[10]

One concern that might arise about the use of the IFLS is that the businesses would be overly concentrated in a few locations, particularly Jakarta and other major economic hubs. Unsurprisingly, such geographic concentration holds in the case of large, formal-sector and public-sector enterprises. Table A.2 illustrates that in fact there is considerable geographic variation in the location of the generally smaller, informal-sector enterprises in the IFLS, and they are relatively evenly distributed across the IFLS provinces. This suggests that the economic geography of informal sector enterprises is more fitting with the overall population distribution, rather than showing high geographic clustering. Also notable here is that capital levels are quite significant–in the year 2000 based on purchasing power parity terms, $500 in Indonesia is equivalent to about $6000 in the US. In addition, the employment numbers (which sum the number of unpaid, usually household workers, paid workers, and the entrepreneur) indicate that the data allows for a significant look at enterprises beyond single-proprietor microenterprises. An exciting feature of the IFLS is the opportunity to analyze firms on either side of the margin between micro, subsistence enterprises and

[10]The largest number of enterprises attributed to any individual household is seven.

firms which begin to provide wage employment demand.

At the same time, very few of the IFLS firms will be growth-oriented firms that would continue to show substantial growth and move into the medium- or large-firm segment of the firm-size distribution. For policy purposes, practically by definition, growth-oriented firms are of particular interest for employment demand.

An additional drawback of the IFLS is that it is difficult to estimate production function models, for three reasons. First, the dataset does not contain yield data, rather only revenue/profit data. Hence quantity of output is entangled with the sale price of output. While the vast majority of the production literature suffers from the same drawbacks, this is still worth noting. Second, the main firm-level data are only collected within the four main survey-round years, with the possibility that the key variables change substantially within the intervening years (assuming the firm doesn't exit). An annual survey would be preferable in this sense. The third concern is the lack of detailed data on inputs. This has two components. While the data on stock variables (capital, labor) is quite good, the capital data is provided in monetary, rather than quantity, terms. Also, there is relatively little information on flow inputs and inventories.

I look at the distribution of firms by industry in Figure A.2 in the Appendix. Clearly the largest number of firms are in Restaurant and Food Sales, Non-Food Sales, or Transport Services in 2000, with a roughly similar pattern seen in 2008. While these activities have large numbers of enterprises, many of the other activities claim dozens if not hundreds of members.

2.3 Key Aspects of the Data Relevant to the Study of Occupational Choice and Microbusiness Activity

In this section I provide an in-depth review of the contents of the survey that are most relevant to the analysis in the later chapters. I focus on two subsections of the larger IFLS questionnaire: (1) household-level questionnaire on enterprises, (2) individual-level questionnaire on labor market outcomes. A much broader summary of the full range of variables available in the IFLS is summarized in the Appendix B.

Summary of Non-Farm Enterprise Subsection of the IFLS

Information on non-farm enterprises is elicited at the household level. Hence these data are collected from a primary respondent who is the household head or an appropriate person 18 years of age or older who is able to answer the questions. The questionnaire appears in Book II of the survey, section NT.[11]

The nature of the data collected on enterprises has changed over time, both on the extensive and intensive margins. On the extensive margin, IFLS4 and IFLS3 elicit information on *all* household enterprises, while IFLS2 and IFLS1 only elicit information on the "primary enterprise." This can be seen in Table A.1, and it means censoring of observations in IFLS2 and IFLS1 that can affect the analysis that can be conducted. On the intensive margin, the enterprise survey has incorporated more questions over time, and the precision of the data collection has improved. For example, the more recent rounds have included

[11] Book II also includes a section for Farm Business (UT). I do not describe this section of the survey since I do not employ it in my analysis.

questions on topics such as business registration and permits. Regarding precision of data collection, the questionnaire makes of tree-based elicitation of values, to reduce non-response.[12]

The Content of the Questionnaire: Non-farm Enterprises The questionnaire for non-farm enterprises can be summarized as follows:

Ownership and Operational Responsibility.

- Whether business is owned entirely by the household: yes/no.
- Percentage ownership share if answer to prior question is 'no'.
- Whom outside the household shares ownership of the business (if any), and report of whom within the household shares ownership.
- Report of which household members are primarily responsible for running the business.

Business Type and Location.

- Type of business.
- Field of work of business (one of 17 categories).
- Whether the business operates outside of the home: yes/no.

Registration and Permits.

[12]For example, if the repondent is asked about the capital stock in their enterprise, they will first be allowed to report a specific value. If they are unable, they are then given a choice of reporting that the value falls above or below 20 million Rupiah. In the latter case they are subsequently asked if the value is above or below 40 million Rupiah and in the former case they are asked if it is above or below 10 million Rupiah. This approach reduces the missing data problem by allowing for some information extraction.

- Types of permits the enterprise applied for.
- Whether or not the permits were issued: yes/no.
- Time taken to obtain permits, and cost of obtaining permits.

Business Characteristics: Startup

- Start date, how many household members/unpaid workers at startup, how many paid workers at startup, (total) startup capital, source of capital (5 categories: household saving, family, other partners, loans from bank, loans from others).

Business Characteristics: Shut-down

- Whether the business is still operating and, if not: household/unpaid workers in month before shut-down, how many paid workers in month before shut-down.

Business Characteristics: Current

- Number of household/unpaid workers during last four weeks, number of paid workers in last four weeks.
- Assets: total value of all land, building, four-wheel motor vehicles, other vehicles, other non-farm equipment connected to the enterprise.

Earnings and Sales

- Net profit generated in the last 12 months.

- Total revenue in last 12 months (including produce used for own consumption).
- Total expenses spent by the household for the business in the last 12 months.
- Products from the business consumed by the household.
- Money out of the business used for the household.
- Amount of money left over from the business (money or saving).
- Total procurement of goods used in business in last 12 months.
- Total sales of the business in last 12 months.
- Total revenue of rents or shared profit of the goods used in the business in the last 12 months.

Summary Statistics: Non-farm Enterprises Summary statistics on these variables are presented in Tables A.3 to A.5.[13] The table is generated from values available from IFLS4, which contains the largest number of variables and largest sample size, 6186 firms. The tables outline three enterprise types: enterprises with no employees (i.e., single proprietorship), enterprises only employing family/unpaid employees, and enterprises employing waged employees. There are 2711, 2326, and 1149 enterprises in each of these categories, respectively. This stratification provides a natural division of enterprise cohorts and allows us to better understand the heterogeneity in the sample.

Ownership and Operational Responsibility.

[13]Some of the variables are omitted, either because they are discussed elsewhere in this Chapter, or they are difficult to summarize in a table (e.g., because they are categorical variables with a large number of categories).

Perhaps not surprisingly, enterprise ownership is almost universally concentrated within the household for the smaller enterprises in the sample. Only amongst enterprises that have paid employees do we observe a significant proportion (10%) that share ownership with individuals outside the household. For the 169 enterprises which have shared ownership, across all three enterprise type cohorts the median ownership share is 50%. The average household ownership share is around 40%.[14]

Business Type and Location.

The majority of enterprises operate outside the home, with 80%, 70% and 80% for the single proprietorships, family/unpaid worker and paid worker categories, respectively.

Registration and Permits.

Overall, less than 10% of IFLS4 enterprises had ever applied for permits. Unsurprisingly, this activity is again skewed toward larger enterprises, with about 30% of paid employee-hiring enterprises applying for permits.

Of those that did apply for permits, the most common were: permit to own/start business (151), commerce permit (150), other permit (113) and neighborhood permit (29). All remaining permit combinations had less than 20 observations (most just one observation). It appears that firms that applied for permits universally received them. This could suggest under-reporting of application failure. The reported permit costs appear to be driven by some ex-

[14] Although it is not reported here, I have analyzed operational responsibility in other work. I find that the vast majority of enterprises in IFLS4 and IFLS3 are operated either by the head of household or their spouse (over 90% and over 70%, respectively). Next is the child or child-in-law of the respondent, respondent's parents, and parent's-in-law. It is relatively rare to have siblings, grandchildren or grandparents operating the enterprise.

treme, outlier observations in the upper tail. The other quartiles show much lower values, and generally rise across complexity in enterprise type. Without correction for permit type the data should not be over-interpreted, but the application costs seem significant relative to other firm size measures like capital stock.

Business Characteristics: Startup

The data on startup size begins to suggest the tremendous density of very small firms, most with no employees. Even if we focus only on firms that employ paid workers, at the 95th percentile of this conditional distribution a firm only starts up with 9 total employees, including the entrepreneur. The distribution is highly leftward-skewed on size.

The data on capital stock is similar, with a highly leftward-skewed distribution in all three firm type categories. It is notable that startup capital is nearly identical across the distribution for the first two type categories: single proprietorships and firms with family/unpaid employees. Yet firms with family/unpaid employees have double the starting labor complement. This is consistent with the business not being very capital intensive, with a production function that doesn't display significant synergy between capital and labor. The businesses that hire paid employees, however, are much more capital intensive. 5% (more than 100 firms) have a capital stock greater than $6400 USD equivalent.

A third interesting variable is the source of startup capital (which is unfortunately only available in IFLS4). Frequency counts for these are as follows: Household savings alone (2781), family alone (2134), both of these (181), loans

from bank (165), and other owners/partners (112). All remaining individual sources or combinations have less than 100 observations. Clearly, capital for these enterprises primarily comes from own saving and family. Perhaps surprisingly, the relative frequency of these sources does not greatly change if we look only at firms with greater than 5 employees, or with greater than $1000 USD in assets. This is suggestive that household and family wealth and social capital can be extremely important in overcoming credit-market imperfections.

Business Characteristics: Shutdown

In practice the data on characteristics of firms at shutdown are clearly not fully reported. While the year-to-year attrition rate seems to be close to 50% in some years, in IFLS4 (2008) we only have 151 reports on firm shutdown (we would expect the number to be in the thousands). It is not clear what might be causing this missing data problem; perhaps cultural norms against reporting on failure, or perhaps in the case of informal enterprises respondents don't think about a formal shutdown procedure.

What we do observe is that self-reported shutdown is reported only amongst enterprises starting with no employees. Perhaps surprisingly, the size distribution of such enterprises at shutdown appears to first-order stochastically dominate that of the continuing firms. It is hard to tell why this might be the case without further available information. One explanation, which is also consistent with the overall low level of self-reporting on business shutdown, is that those self-reporting shutdown are dominated by individuals who go into retirement. Retirement might be a stage when a survey respondent would have a clearer concept of deliberate shutdown, and such enterprises might tend to be slightly larger if they are more mature and not unwillingly driven out of the market.

Business Characteristics: Current

For the current state of enterprises we have data on labor employed and capital stock. The distribution of labor employed looks similar to the levels at startup for enterprises with no employees, and enterprises that only employ household members. In terms of labor supply, on average these enterprises do not show very much change in size. This could be indicative of a total lack of constraints and uncertainty on the one hand – firms are able to start at their optimal size and remain there. On the other hand, it could be indicative of extreme constraints – firms start as large as they can, and are totally unable to grow beyond that. The latter explanation seems slightly dubious given the freedom that firms have to reinvest profits in the enterprise, but the descriptive evidence is not conclusive.

A more vibrant story is painted by the distribution of capital stock. First, within the current distribution of capital, we see a lot of variation by firm size. The average firm with no employees currently has about 814 USD equivalent in capital, while for firms with family/unpaid employees the value is 867 and for firms with waged employees it is 5786. Looking at percentiles the distributions for enterprises with no employees and those with only family/unpaid employees are not so different, though the latter appears to show moderate stochastic dominance over the former. Consistent with the larger average capital stock, the distribution of capital within the subset of enterprises with waged employees is strongly first-order stochastically dominant of the other two distributions. The 95th percentile of capital stock in the two simpler enterprise categories would not even be at the 75th percentile of the distribution for firms hiring waged employees.

Another way to look at capital stocks is to compare the distribution of capital at startup to current levels. Note that this comparison is not biased by selection because only surviving firms report, though it does bunch up firms of various ages. Quite consistently across the enterprise size categories we see that capital tends to roughly double from startup levels. We again see a strong right-tail effect, with the largest changes quite dramatic – in some cases increases in capital stock on the order of 500%.

These results are interesting on two levels. First, capital stocks seem to increase quite a lot. Perhaps lack of access to financial capital is not such a constraint for these firms. Second, and contrastingly, it is puzzling that capital stock quantities would change so much, while labor employment quantities would change so little. This is consistent with the explanation that the marginal revenue product of additional labor is low if the household has trouble incorporating additional labor that is outside the local social network (due to monitoring and effort-inducement challenges). If the enterprise exhausts easily-available labor on startup, it may be the case that increasing the capital intensity of the existing workforce is the best way to improve returns.

Earnings and Sales

The data on earnings and sales is quite consistent with the data on capital stocks. Values for the two simpler enterprise types, single proprietorships and those only with household/family workers, are quite similar in mean and moments, though the values for household/family tend to slightly first-order stochastically dominate the single proprietorships. Returns for enterprises hiring outside wage workers are generally quite significantly larger, with an impressive upper tail. We can also quickly see the pervasiveness of missing data

problems for other variables – revenues, sales, expenses, etc. This massive attrition introduces potential bias that makes it very hard to look at these other variables.

The data on usage of earnings (products consumed, returns used by household, returns left over) again follows similar patterns across the enterprise types. What is notable, and perhaps surprising, is the number of 0 values for these variables. Up to 50% or more households report 0 values for these variables, which is interesting given that we might expect uniformly high integration of the household with the enterprise. One potential explanation is that such observations are concentrated amongst enterprises producing services, or producing goods that are not for direct household use.

Summary of Labor Activity Subsection of the IFLS

The section of the IFLS on labor market activity is quite extensive. It includes information on job search activity, recent work activity, the primary and secondary employment of an individual (including recall data going back to the previous IFLS round), and additional employment-related information such as on retirement, pension and health insurance. Below I focus on the key variables under labor market activity that are related to self-employment, which I make use of in later chapters.

I will not provide summary statistics of these variables in the present chapter, either because the useful summary statistics will come from aggregations of these variables (as with enterprise experience) in a subsequent subsection, or because these variables will be more usefully discussed and analyzed later in

the thesis.

The Content of the Questionnaire: Labor Activity Relevant to Self-Employment Primary and Secondary Job Details[15]

- Employer's product, sector, number of employees (including a categorical variable).
- Category of employment (includes self-employed with no employees, self-employed with unpaid family worker/temporary worker, self-employed with permanent worker).
- How much net profit did you gain last month, after taking out all your business expenses? Followed by categorical variables if necessary (greater or less than 1 million Rupiah per month, then greater or less than 10 million, or greater or less than 500 thousand Rupiah).
- Approximately how much net profit did you gain last year, after taking out all your business expenses? Followed by categorical variables if necessary (greater or less than 12 million Rupiah per year, then greater or less than 80 million, or greater or less than 6 million Rupiah).

Recall Data on Primary and Secondary Occupation

- Recall data goes back year-by-year, always at least as far back as the year of the most recent IFLS round.

[15]The same questions are asked for both the individual's self-reported primary and secondary occupations (where occupations are ranked according to where the respondent spends the most time). In this thesis I focus on the primary occupation almost exclusively.

- Includes a subset of the questions above, including employer details, occupation type (types of self-employment, or other).

Additional Employment Information

- Questions about the individual's first job, including age, employer, information on employer, hours worked, the occupational category, and approximate wage/salary.

2.4 Constructed Variables and Indices: Definition and Summary Statistics

In the subsequent chapters I will work with a number of constructed variables, including measures of wealth, work experience, and education. In this section I discuss the construction and definition of these variables, and how they are computed from raw variables in the IFLS. I also discuss the construction of a price deflator index variable that I construct in order to make panel variables with values expressed in terms of prices comparable over time.

Household Wealth

I construct the measures of household wealth based on comprehensive set of self-reported household asset values, including both enterprise wealth and direct household assets, in Book 2 of the survey. The interview for Book 2 is conducted with a representative member of the household, generally the household head or the most responsible alternative household member. In the subsequent

analysis I use household wealth as an aggregate variable, and also disaggregate wealth slightly into liquid and illiquid asset holdings (i.e., total wealth = liquid wealth + illiquid wealth). I generally use the lagged value of wealth in regressions, in order to avoid simultaneity bias in regression analysis.

The assets contributing to each measure are as follows:

Liquid wealth. Household poultry, household vehicles, household appliances, savings/certificate or deposit/stocks, receivables, jewelry, household furniture and utensils, other household assets, business four-wheel motor vehicles, business other vehicles, business other non-farm equipment.

Illiquid wealth. House and land occupied by household, other house/building, non-farm land, business land, business building.

The measures are constructed by summing each of these values, and then deflating them as appropriate with the price deflator used throughout the survey.

Self-employment Experience

As already noted, the IFLS provides detailed information on individual's occupations on a year-by-year basis. Most importantly, these measures distinguish three categories of self-employment: running an enterprise with no employees, running an enterprise with only family/temporary workers, or running an enterprise with waged employees. Hence I construct measures of self-employment experience by summing across years of experience in each of these three categories.[16] For each IFLS survey round, an individual self-reports on

[16]While the IFLS distinguishes between experience in primary and secondary enterprises, I exclusively construct measures based on primary occupations, due to the additional complications raised by incorporating secondary occupations (more incomplete records, the need to

year-by-year occupational status at least as far back as the most recent previous round. Hence the experience measures can be constructed in a very direct way for the years covered by survey rounds. Because the recall portion of the survey inquires about whether a given occupation continued from year-to-year (including running a particular enterprise), I further stratify the measures to distinguish years of experience in running one's current enterprise, and total experience running previous enterprises.

To make this clearer, let me provide a simple example. Suppose Iwan appears in rounds 3 and 4 of the IFLS (in 2000 and 2008, respectively). Suppose he enters the workforce in 1998, running a single proprietorship from 1998-2002, then goes to work in a company for three years, then starts a new business in 2005, first alone, then with two paid workers as of 2007, which he continues to run when the survey is collected in 2008. Then as of 2008, he would be recorded as having six years of total single proprietorship experience, and two years of experience running his *current* enterprise with paid employees.

The construction of the experience variables becomes more challenging in cases in which individuals are not covered by the annual recall in survey rounds. For years that exceed the coverage of the survey (in particular, years preceding 1988, the earliest year as of which year-by-year occupational data are available), there are two main sources of information. First, in IFLS1 (1993) there was more useful recall data collected on work histories, at 5-year intervals. Individuals are queried about their place of work on an annualized basis back to 1988, then at 5-year intervals for the preceding 20 years. Beyond that, *each* survey round contains information on an individual's first place of employment, which is particularly helpful in constructing total work histories for individu-

weight experience in primary and secondary occupations by hours spent, etc.).

als who don't appear in IFLS1. In order to back out experience values I look at what the individual was doing at the beginning and end of a non-yearly recall period, and assign the full period to one category or split the period between two categories as appropriate.

So to summarize, I construct 6 variables for enterprise experience, stratifying across current/total experience, and single proprietorship, enterprise with family/temporary workers, and enterprise with paid (outside) workers:

1. Years of experience in running a current (active) single proprietorship.
2. Total years of experience in running single proprietorships.
3. Years of experience in running a current (active) enterprise with family/temporary workers.
4. Total years of experience in running enterprise with family/temporary workers.
5. Years of experience in running a current (active) enterprise with paid (outside) workers.
6. Total years of experience in running enterprise with paid (outside) workers.

Education

The measures of education are constructed by summing total years of education. The IFLS provides a detailed section on education in Book 3A, including dummy variables for which level of education the individual reached (primary, secondary, tertiary, etc.) and how many years were spent at each level. The education experience is further stratified by school types – private, public, special

Islamic schools, etc. I make no corrections for quality or level and take no notice of school type–my measures are purely sums of total years of education.

Price Index

I use a price index that deflates all of the nominal monetary prices (almost always expressed in terms of Rupiah, the Indonesian national currency, in the IFLS) to a comparable value across years. This is taken from national consumer price indices that I obtained from International Monetary Fund statistical extracts, which use 2005 as the base year.

When currency conversions are made (into US dollar terms), I use the a conversation rate applicable to each survey round.

Summary Statistics: Constructed Variables

Summary statistics on selected constructed variables are contained in the Appendix.

Household Wealth

A graph depicting the log of wealth (in US dollars) is given in Figure A.3 in the Appendix. The log transformation pulls the distribution into a fairly normal shape, with a higher peak around the median of the distribution.

If we look at the raw numbers on wealth, broken down by liquid and illiquid, we see that total wealth is dominated by illiquid wealth, covering about 80% of total wealth over most percentiles of the wealth distribution in the year 2000. Average wealth is 26.3 million Rupiah, or about 2630 USD at the time,

with a standard deviation of about 103 million Rupiah. The median of the distribution is 6.22 million Rupiah (or about 620 USD) and the 95th percentile of the distribution is almost 100 million Rupiah (or nearly 10,000 USD).

Liquid and illiquid wealth tend to have about a 20/80 division of these values, though there tend to be more zeroes at the lower end of the illiquid wealth distribution, indicating households that aren't in possession of assets categorized as illiquid.

Self-employment Experience

The summary statistics on self-employment experience are presented in Table A.6. The table aggregates year-by-year counts of individuals' primary occupation over the full panel available in IFLS (1988-2008). Hence an individual could be represented multiple times (at most, 20 times) in the counts. Here the enterprise types follow the codes occasionally used elsewhere in the thesis (1=no employees, 2=only family/unpaid employees, 3=waged employees).

In line with other evidence on enterprise activity in IFLS, the largest number of observations come from experience with enterprises with no employees (over 17,000), with experience in enterprises with only family/unpaid employees close behind (over 14,000). Far fewer observations come from enterprises with waged employees (over 600). It appears that individuals tend to acquire longer experience profiles in the two simpler enterprise types compared to the type employing waged employees, both in the mean (over 6 years to over 4 years) and looking at the percentiles. This is consistent with greater mobility for individuals capable of running more complex enterprises, along with greater turnover faced in that sector.

What we also see in the lower three rows for each enterprise type is that there is mobility between the various enterprise type categories. About 10% of individuals currently running enterprises with no employees have experience in the other two enterprise categories, while the amount is about 20% for individuals currently running enterprises with unpaid/family employees and about 25% for individuals with waged employees. We also see that years of experience in the other enterprise categories is not insignificant, either. This seems to suggest the presence of a fair amount of ”serial entrepreneurship”, with an important cohort of individuals acquiring enterprise experience across various enterprise sizes and types.

Education

Descriptive evidence on education, stratified by individuals running enterprises in the three type categories, is included in Table A.7. The averages are consistent with the idea that most informal-sector enterprises are run by individuals with low education levels. This is particularly true of enterprises that don't employ wage workers, with a little over 2 years of education on average. The results are even starker if we look at the percentiles – over 75% of individuals running the two simpler enterprise types have no education at all, with the average level of education apparently driven by a relatively small subset of individuals with a fair amount of education. Education levels seem slightly higher amongst individuals running enterprises with waged employees, with 12 years of education falling at the 75th percentile of the distribution. The evidence is consistent with [Porta & Shleifer, 2008] who show that most small, informal-sector enterprises are run by individuals with low education levels. They use that evidence to draw doubts on the potential of most informal-sector

entrepreneurs to run relatively large, formal-sector enterprises.

2.5 Descriptive Analysis of Enterprise Activity in the IFLS

In this section I present further descriptive analysis of enterprise activity in the IFLS. I begin with two large tables of correlations between the variables described in previous section. I then proceed to provide some descriptive, non-parametric evidence providing further nuanced evidence on the distribution relationships between key variables for financial capital and human capital.

Tables of Correlations

Enterprise Variables I provide correlations between all of the main variables on individual enterprises in Tables A.8 to A.11, with values drawn from 2008 and one from 2000, respectively. In the interest of brevity, I will focus the discussion on Tables A.8 and A.9 exclusively. To be clear, the variables are all specific to a given enterprise within a given household. Hence a single household may be represented multiple times in the statistics, if it holds multiple enterprises. I will generally focus the discussion on correlations of at least |0.1|, unless a weak correlation is particularly notable. The exception is cases in which strong correlations are trivial, primarily because one variable (e.g., number of family/unpaid employees) is a component of another variable (e.g., total employees).

I proceed column-by-column, to ease exposition.

An enterprise being fully family-owned is interestingly negatively correlated with the amount of family labor and the amount of land invested in the en-

terprise, when we might expect the opposite. It is interesting that full family ownership is weakly related to whether the enterprise runs outside the home – we might expect a negative correlation between the two if having non-family owners makes outside operation more likely.

The labor correlations if an enterprise operates outside the home are not surprising: we would expect less family workers to be involved and more outside wage workers in this case. Also unsurprising is that such enterprises have a larger capital stock, primarily driven by buildings and vehicles. Interestingly, rural enterprises and those outside Java are more likely to operate outside the home. Perhaps this is because of lower population densities in rural areas and outside Java, so the enterprise needs to go "where the market is."

Surprisingly, having a business permit is very weakly correlated with all of the other enterprise variables. This is particularly surprising in regards to capital stock, which is if anything (very weakly) negatively correlated with having a permit. While only about 5% of enterprises have permits overall, we would expect these to be the more significant enterprises, both because the law impinges on enterprises with 10+ employees, and larger enterprises would be more susceptible to enforcement of permit regulations.

I next look at the three startup labor measures. It is notable that household labor and wage labor at startup are weakly negatively correlated, indicating that they are substitutes rather than complements. Unsurprisingly the startup values of labor are strongly positively correlated with current levels of the same. While household/unpaid labor at startup is correlated with current net profit, waged labor is highly positively correlated with current capital levels (0.42) and net profit (0.44). Enterprises that startup with more waged workers are also

likely to show a strong increase in capital stock (0.21). While all of this evidence is correlational, it is consistent with evidence in Chapter 3 of the thesis about much steeper earnings improvements seen amongst enterprises with waged employees.

The amount of startup capital is unsurprisingly highly positively correlated with a number of measures of business size and success, including current capital levels, number of employees, and current returns. Interestingly, there is a fairly strong negative correlation between startup capital and the propensity to show a change in capital since startup (-0.29). This may be indicative of heterogeneous credit constraints on enterprises, with some able to startup close to the optimal capital level, while others are not able to do so (though they are apparently able to subsequently converge to the optimal level).

The evidence on current labor employment is consistent with that for startup labor. First, there is negative correlation within the labor categories – family/unpaid labor appears to substitute for waged labor. Second, while current family/unpaid labor is very weakly correlated with measures of capital stock and returns, waged labor shows a strong correlation with a number of such measures. Current waged labor is also fairly strongly correlated with a positive change in capital stock, indicating a complementary between the two, and perhaps constraints preventing the optimal level of capital from being reached initially.

The next six columns pertain to capital stock. Unsurprisingly capital stock values tend to be positively correlated with each other, indicating complementarities, though there is tremendous variation in the level of correlation, sometimes as low as zero, sometimes close to 0.7. Interestingly quantity of capital

is weakly correlated to the average return to capital, seemingly consistent with CRS production technology. On the other hand, a number of capital stock measures are quite strongly positively correlated with average returns to labor, indicating fairly strong complementarity between the two.

Finally, we have a number of returns measures and constructed measures. Perhaps most surprising here is that average returns to capital show little relationship to other variables, while average returns to labor are quite strongly positively correlated with the returns measures. The remaining correlations are relatively unsurprising, including the strong tendency (0.44) of Javanese enterprises to be urban.

Household Enterprise Activity I provide a table of correlations amongst household and enterprise activity variables across Tables A.12 and A.13. Where an individual is referred to (i.e., for demographic variables such as age, education, etc.) it is always the household head who is recorded. I will generally focus the discussion on correlations of at least |0.1|, unless a weak correlation is particularly notable. The exception is cases in which strong correlations are trivial, primarily because one variable (e.g., number of family/unpaid employees) is a component of another variable (e.g., total employees).

I proceed column-by-column, to ease exposition.

The first set of variables are choice variables summarizing various enterprise-activity choices. First, we have a variable for whether a household started up an enterprise (and did not own one previously). I ignore unsurprising correlations within these outcome variables themselves (amongst which we naturally expect high correlation). The first interesting thing to note, although

the correlation is rather weak, appears to be a story about suboptimal overcapitalization of enterprises amongst households new to enterprise activity: there is a positive correlation (0.06) between first-ever enterprise startups and the capital stock in the enterprise, but a negative correlation with current capital stock. This story even persists, though more weakly, for households that had prior experience. Quite interestingly, there is very little correlation of enterprise activity with lagged values of household wealth. We see that education of household head is actually positively correlated with propensity to start an enterprise, which is somewhat surprising if we think that most enterprise startup activity comes from subsistence self-employment. Of further interest is a negative correlation with measures of household labor supply (male and female). This relationship persists, though at slightly weaker level, for households that already owned an enterprise previously.

Next come measures capturing enterprise returns and capitalization. Unsurprisingly, total net enterprise returns are quite strongly positively correlated with measures of capital invested in the enterprise, lags of household wealth, and education of the household head. Also unsurprising is the fairly strong positive correlation of household wealth and the amount invested in the enterprise. This can be contrasted with the lack of correlation noted above between entry and wealth – what these results suggest is that household wealth matters much more for the intensive margin (how much to invest in the enterprise) than the extensive margin (whether to start an enterprise at all). It is interesting to contrast this finding with the role of education of the household head, which shows a notable positive correlation on both the intensive and extensive margin.

The next three columns summarize measures of household wealth, includ-

ing an aggregated measure, and then a breakdown to liquid and illiquid capital. Perhaps unsurprisingly, household wealth is positively correlated with household size and household head education.

The subsequent three columns summarize measures of household size. Most of the correlations are trivial based on reasonable expectations for demographics, though it is interesting that urban households have significantly more female, adult household members. Also interesting is that household heads with experience with more complex enterprises are likely to have more male adult members. This is consistent with evidence from other settings in which it seems that having more male household members allows enterprises to expand due to having more available supervisory labor.

The subsequent four columns summarize characteristics of the household head: age, education, gender and whether married. None of the correlations seem particularly interesting, other than for demographic purposes. Finally, we have four columns of measures of enterprise experience of the household head, and then two final variables capturing urban location and Java location. Most notable is that total years of experience is negatively correlated with experience in each of the three enterprise categories. This is puzzling at first, but makes sense in light of a simple adding-up condition: if individuals have many years of experience in one particular enterprise type, all things equal they should have fewer years of experience in another enterprise type. Hence the sample is likely segmented between two types of individuals: ones with many years of experience in one enterprise type, and ones with many years of total overall experience (spread across the types). The latter group, with lots of years of total experience, will have years of accumulated experience likely to be negative correlated with

experience in the other categories, which have individuals who are more specialized.

Non-parametric analysis

This section summarizes some descriptive results that are generated with bivariate non-parametric methods. The first set of figures focus on the relationship between wealth and enterprise startup activity, while the second set focus on the relationship between enterprise experience and the propensity to further engage in enterprise activity. The graphs are generated using a non-parametric Lowess[17] tri-cube smoother, which overweights relatively close observations in generating the value of the function.

Wealth and enterprise activity Graphs summarizing bivariate relationships between (lagged) wealth and enterprise activity are presented in Figures A.4 and A.5. Figure A.4 presents the relationship between wealth and enterprise startup amongst urban households in the IFLS, while Figure A.5 presents the same for rural households. The three sub-figures in each figure look at different subsets of households, corresponding to the lower 50, lower 75, and lower 99 percentiles of the wealth distribution.

These graphs capture the relationship between household wealth in 2000,

[17] Lowess stands for "locally weighted scatterplot smoothing." It involves fitting low-degree polynomials to the observations near to a particular datapoint using weighted least squares through a nearest neighbors algorithm.

The tri-cube smoother specifically uses the weighting function:

$$w(x) = \left(1 - |x|^3\right)^3 I\left[|x| < 1\right],$$

where x captures the distance between the particular datapoint and another point in question.

and the propensity to start enterprises in 2001-2008 (and have them survive until 2008). Almost uniformly I find a relatively weak relationship between wealth and the propensity to startup. If we look at households below the 50th percentile of the wealth distribution, we see almost no relationship between wealth and the propensity to startup, except perhaps at the very lower end. This is surprising if we think that relatively poor households are most constrained by financing in their ability to pursue enterprise activity. Looking up to the 75th percentile, we do see somewhat of a positive relationship between wealth and enterprise startup, though it is still weak. Looking all the way up to the 99th percentile most of the graph appears flat again, and there is even a negative relationship for rural households.

Human capital (experience and enterprise activity) Graphs summarizing bivariate relationships between human capital (measured in terms of experience, as discussed above) and enterprise activity are presented in Figures A.6 and A.7. Each Figure is broken into three sub-images. The first looks at the relationship between entrepreneurial experience of the individual running the business and propensity to start an enterprise for experience in enterprises with no employees, while the latter do the same in relation to experience with enterprises with only family/unpaid employees and enterprises that employ wage workers.

While experience could be proxying for a number of things, including outcomes orthogonal to human capital such as accumulated capital stock, the story presented by these figures is consistent with a story about human capital accumulation. The propensity to continue to engage in enterprise activity is quite strongly increasing in experience, and this relationship is strongest for individuals running enterprises that employ wage workers. According to the theories

to be presented later, these are the individuals most likely to accumulate significant entrepreneurial experience. The individuals who are likely to be the most mobile also show great persistence, as success appears to beget success. We will look at this suggestive evidence more rigorously in later chapters.

2.6 Conclusion

In this Chapter I provide an overview of the Indonesia Family Life Survey, particularly in relation to enterprise and employment activity, which will receive much more detailed attention as this thesis proceeds. I describe how the IFLS is a rich source of information on individual and household-level enterprise activity.

I also provide some initial, descriptive evidence on enterprise activity in the IFLS. What we see is consistent with a few different theories: that many informal sector enterprises are run by lowly-educated individuals, that the set of growth-capable small enterprises is relatively small, that enterprises face important constraints in their ability to startup and grow, and that the accumulation of human capital specific to enterprise activity through experience (learning-by-doing) may be an important factor in enterprise outcomes.

In the remainder of the thesis I develop these themes in much more detail. In Chapter 3 I focus on the question of who really is constrained by financing, and who might be constrained by other factors. In Chapter 4 I focus on a particular factor outside of financing, by developing and testing a theory of entrepreneurial human capital accumulation. In Chapter 5 I provide further tests based on a fully dynamic version of the primary model.

CHAPTER 3

CREDIT MISPLACED? TESTING FOR HOUSEHOLD-LEVEL FINANCIAL CONSTRAINTS TO ENTERPRISE ACTIVITY

3.1 Introduction

Across developing countries roughly half of the workforce is primarily self-employed. Since such enterprises account for such a large proportion of the active labor force, their outcomes have welfare implications on a large scale. Yet the vast majority of such enterprises are small, often with no employees, and fail to grow. Hence there is tremendous interest in the factors that shape the formation and performance of such enterprises. A key focus of the academic literature on this topic has looked at the role of financing (or lack thereof) as a stimulant or a constraint to enterprise activity, particularly in light of the international spread of microfinance as a means of alleviating financing constraints to small-scale enterprise activity. Recent, large-scale randomized trial studies have drawn into question the effects of microfinance in stimulating small-scale enterprise activity ([Banerjee *et al.*, 2009], [Karlan & Zinman, 2010]). In this paper I take a step back from the microfinance-focused literature and present a number of tests of financial constraints to small-scale, developing-country enterprise activity, employing an unusually rich household panel survey from Indonesia. I delve into the heterogeneity across the self-employment distribution, testing the hypothesis that most low-wealth households are not primarily financing-constrained in their enterprise activities.

Financing constraints are difficult to empirically identify in observational data, because they involve both supply-side and demand-side factors that are

difficult to disentangle.[1] The supply of financing – whether by a bank, a non-bank financial institution, an angel investor or one's family or friends – depends on the potential lender's opportunity cost of capital and belief in the promise of the potential borrower's project and chance of repayment. On the other hand, the demand for financing depends on the potential borrower's own beliefs about project quality, aspects of the borrowing contract such as pricing, timing, and collateral conditions (which raise the prospect of risk rationing), and the opportunity cost of capital to the owner. Because of these concerns, I do not look directly at financing flows coming from lending-and-borrowing activity. Rather, I study financial transfers of which the timing is exogenous to the recipient household.

The empirical tests are derived from an extended version of the standard model of financing-constrained occupational choice.[2] This standard model involves looks at the choice between waged employment and self-employment of an individual with fixed values of wealth and entrepreneurial ability, and a fixed value of access to outside financing. I extend the model to allow for two types of potential entrepreneurs – a low-ability type and a high-ability type. I motivate this with evidence from my primary dataset and other studies, which point to significant heterogeneity among the self-employed. It has been suggested (e.g., [Schoar, 2010]) that the set of self-employed might be best characterized as a mixture of "subsistence" and "opportunity-oriented" individuals. The subsistence (low-ability) types have low opportunity costs to entering self-employment, and would probably be pulled into wage work by greater opportunities in that activity. The high-ability (opportunity) types start a business to

[1] [Karlan & Zinman, 2009] provide an innovative approach to disentangling some of these factors in the context of consumer lending.

[2] A full exposition is given in [Evans & Jovanovic, 1989].

pursue a genuine entrepreneurial opportunity, and may have a relatively high opportunity cost to doing so. The main result in the model is to show that the high-ability types may actually be relatively more constrained by financing, even though they are likely to be wealthier and more able to self-finance. While perhaps not profound, this result contrasts sharply with results derived from this model in existing literature.

I then examine the possibility that wealthier households are relatively more financing-constrained, and other predictions of the theoretical model, on a large-scale household panel dataset that has been collected over 15 years in Indonesia. I study two types of outcomes: decisions to startup enterprises and run enterprises (extensive margin), and the capital invested in the enterprise (intensive margin). The latter outcome is particularly notable; while many datasets have information on occupation and hence allow for the study of extensive margin effects, far fewer match individual variables to enterprise variables, and in particular capital investment in the enterprise, allowing for the study of intensive margin effects. For example, [Bianchi & Bobba, 2010] study the effect of plausibly exogenous financing shocks due to the Mexican Opportunidades program on enterprise formation, but cannot analyze capital investments. Hence it is not clear if enterprises are formed due to financing constraints or as a more-desired occupation to be enjoyed while excess funds are available.

I study the First, I study the relationship between assets and wealth as stock variables, and enterprise decisions. These stock variables provide a long-run measure of the household's financing base, which can be thought of as a relatively broad measure of liquidity. Such resources might be directly employed in financing an investment, or used as collateral to obtain additional financing. If

financing matters for enterprise outcomes, then we would expect a high correlation between such wealth measures and enterprise activities.

In order to deal with simultaneity concerns between enterprise activity on the one hand and measurement of wealth and assets on the other, I employ lagged values of asset and wealth variables, as in other studies using similar methods (e.g., [Paulson & Townsend, 2004], [Hurst & Lusardi, 2004]). In order to more flexibly study the relationship between wealth and enterprise activity across the wealth distribution, I employ a novel semiparametric technique that allows me to improve on previous tests in the literature by more flexibly estimating the role of wealth in enterprise outcomes. The semiparametric approach allows me to identify a clear non-linear effect–over the lower 60 percentiles of the wealth distribution there is little relationship between lagged wealth and future enterprise activity. As of about the 60th percentile wealth becomes a key factor, and then tails off again at the upper tail of the wealth distribution.

Second, I study the relationship between income shocks and the same enterprise decisions. While assets provide a positive, long-term measure of a household's financial base, income provides a more direct measure of temporal shocks. This analysis takes on two strands. First, I study liquidity-based tests that exploit data on income shocks that are potentially anticipated. These include exogenous transfers such as conditional and unconditional cash transfers, and other sources of government transfers. As I discuss in the context of the theoretical model, if a household is unconstrained by financing then the timing of such transfers should not be related to enterprise decisions. On the other hand, if there is a correlation between the two, then it suggests that households are forced to depend on such financing sources in order to undertake enterprise

activities.

The second income-based test exploits plausibly-exogenous transfers through sources such as bonus payouts, insurance payouts and lottery payouts. Here the transfers are more clearly unanticipated. While the decision to engage in activities that make such transfers accessible might certainly be correlated with individual characteristics, since the receipt of such payouts is random we expect that the *timing* of such receipts is exogenous, and hence provides a randomized source of exogenous variation in financing. Again, what we expect is that financing-constrained households would be responsive to such transfers.

An additional goal of the paper is to look at heterogeneity in the response to income shocks, particularly on the wealth dimension. This is motivated by papers such as [Banerjee *et al.*, 2009] and [Karlan & Zinman, 2010], which find that the effects of financing shocks on enterprise activity are smaller for poorer households, even though returns to capital tend to be higher in firms run by poorer households. Hence an additional goal of this paper is to provide evidence across the wealth distribution and over time regarding the role of income shocks. While the simplest version of a theory of financing constraints predicts that wealthier households should be less responsive to wealth transfers, all things equal, there is actually suggestive evidence that this prediction does not hold – that in fact wealthier households are sometimes more responsive.

The analysis in this paper is related to a large, emerging literature on financing constraints and household enterprise activity in developing countries. [de Mel *et al.*, 2008] find 55-63% annualized returns by providing random shocks of $100-200 USD in cash or in kind to microenterprises in Sri Lanka, with the sample limited to firms with no paid employees and a maximum capital

stock of $1000 USD. [Udry & Anagol, 2006] calculate returns on investment in pineapple production in Ghana, finding mean returns as high as 250% per annum, on plots of a fraction of a hectare. The observation that some enterprises appear to have high returns to physical capital and yet many fail to grow has been interpreted by a number of papers as evidence for financial constraints (e.g., the review [Banerjee & Duflo, 2010]). In theory, enterprises should equalize the marginal cost of capital with its marginal return, which implies that if the smallest firms have the highest marginal returns they should be the most responsive to positive financing shocks. Hence these seminal papers and others have been interpreted to indicate that there are significant financial constraints to developing country enterprise activity.

Yet, clearly, high marginal returns do not necessarily imply financial constraints – it may be that unmeasured human capital constitutes an additional, valuable stock of capital in the enterprise, for example ([Naude, 2008], [Udry & Anagol, 2006]).[3] Also, access to capital in itself might be symptomatic of insufficient human capital. Indeed, [Ikhwan & Johnson, 2009] present evidence showing that potential entrepreneurs in Indonesia significantly underestimate their access to financial capital (as verified by assessments from bank loan officers).[4] While lack of access to capital from formal financial institutions might be evidence for financing constraints, it is also consistent with an efficient financial market that selects reasonably well on ability for many self-employed ([McLeod, 1980]), at least in expectation ([Ghatak *et al.*, 2007]).

Others have argued that finding high returns to capital might be due to the fact that exogenous assignment of financing can act as a guide to sub-

[3][Udry & Anagol, 2006] acknowledge that "(t)hese returns are not adjusted for risk. . . [and] it is not possible to distinguish the returns to entrepreneurship from those to capital".

[4]See also [Astebro & Bernhardt, 2005].

jects on how to allocate investment, overinflating the estimated treatment effect ([Bruhn *et al.*, 2010]). Even in the US or other developed economies with apparently more efficient credit markets, the majority of micro and small enterprises are self-funded or are initially unable to raise capital from formal financial institutions. There is also little evidence on the *dynamics* of micro and small enterprise returns – it may be that returns are quite volatile over time and while some enterprises could have high average returns today, this could be quite different later. It is an open question why, if the finding that average returns are well above the market interest rate truly represents a financing distortion, why the self-employed do not leverage such high returns to own-save and increase the stock of enterprise capital.

Hence it is still an open question to what extent financing constraints are actually binding. This Chapter contributes to the literature by providing evidence on financing constraints generated from natural exogenous variation available in observational data. It exploits a dataset of unique scope and timespan relative to existing studies, which allows a more nuanced picture to be derived about financing constraints across the wealth distribution. It provides a theoretical rationale for the possibility that less wealth households might actually be less constrained by access to finance, and provides empirical evidence in support of this hypothesis from a number of approaches.

The paper proceeds as follows. I begin by presenting descriptive evidence on entrepreneurial heterogeneity in Section 3.2. This motivates a simple model of financing-constrained occupational choice that highlights the occupational choice effects of wealth and asset stocks on the one hand, and windfall income flows on the other, in Section 3.3. I then further discuss the data in Section 3.4.

In Section 3.5 I present wealth and asset-based tests of financing constraints, while in Section 3.6 I present income shock-based tests. In light of the results in Sections 3.5 and 3.6, in Section 3.7 I present some additional evidence on the factors behind entrepreneurial choice going beyond financial constraints. Section 3.8 concludes, while tables, figures and additional derivations are presented in Appendix B.

3.2 Descriptive Evidence: Wealth, Income and Enterprise Choices

In this section I provide descriptive evidence about the nature of heterogeneity in enterprise activity. This evidence is strongly suggestive of the hypotheses that I study in more rigorous empirical frameworks later in the Chapter; namely, that it is important to distinguish subsistence and opportunity-oriented enterprise owners, and financing is not the primary constraint to enterprise activity for most households.

The evidence is based on the primary dataset that I will make use of for the later analysis in the paper, which is a large-scale household survey dataset from Indonesia that I described in Chapter 2 and which I will describe further in Section 3.4. First, I compare the earnings distributions for wage employed and self-employed individuals, showing that the wage earnings distribution first-order stochastically dominates the net profit distribution for self-employment over about the lower 90 percentiles of their supports. This finding is particularly puzzling since we might expect a compensating differential in earnings from self-employment due to its additional risks. Second, I provide descrip-

tive evidence on capital holdings and average capital holdings, again providing evidence for significant heterogeneity.

3.2.1 Opportunity and subsistence self-employment

I provide evidence that there is significant heterogeneity amongst micro and small enterprise owners. In particular, I argue for the distinction between ”subsistence” and ”opportunity-oriented” enterprise owners, to distinguish between those who have few alternative opportunities and hence are in some sense ”forced” into self-employment, versus those who enter to pursue a meaningful business opportunity.

First, in Tables A.3 to A.5, we see that most enterprises operate with negligible capital stock, with a relatively small proportion of enterprises operating with considerable capital stock. More than 75% of microenterprises operate without any significant business assets at all – they report zero holdings of land, buildings, machines and vehicles, usually just reporting small stocks of equipment or other working capital. Similar patterns hold for other aspects of enterprises, such as labor, employment, returns, etc. – there is dramatic skewness in the distribution. This provides simple evidence for important heterogeneity in the (realized) distribution of self-employed individuals and households.

I provide additional evidence for the subsistence-opportunity distinction through studying the wage premium. Similar to studies from more developed countries, I find that the wage employed have a significant earnings premium over the self-employed over most of both earnings distributions. This can be seen in Table B.1, which summarizes the value of earnings at a number of per-

centiles of each earnings distribution, focusing exclusively on individuals who report the earnings corresponding to their self-reported primary occupation,

Up to the 90th percentile the Wage earnings distribution first order stochastically dominates the Net profit distribution. At the 90th percentile the two distribution functions cross, at $162 USD per month. Above the 90th percentile the earnings distribution from enterprise activity quite dramatically exceeds that from Wage employment, nearly doubling in value as of the 99th percentile.

Similar evidence from more developed economies has been interpreted as a puzzle in the existing literature, because we would in general expect to see that the differential should go the other way ([Hurst & Lusardi, 2004]). Self-employed individuals should be compensated for presumably taking on greater risk, and hence over most if not all of the earnings distribution there should be an earnings premium for the self-employed. Hence such evidence has been used to motivate the idea that entrepreneurs have very strong preferences for non-tangible entrepreneurial benefits, such as independence, or behavioral biases, but I argue that in the developing country setting there is a more intuitive explanation. In a less developed economy there is generally a surplus of labor, much of which gets allocated to subsistence enterprise activity. Hence there are some individuals who are self-employed due to a low opportunity cost to self-employment, while a relatively small subset of individuals enter self-employment to pursue an opportunity, at a relatively high opportunity cost.

Such heterogeneity amongst the self-employed is supported by other, recent empirical literature. [de Mel *et al.*, 2010] find that 2/3 to 3/4 of microentrepreneurs (individuals running enterprises with few, if any, employees) from a survey sample in Sri Lanka have personality traits much more like those of

wage workers than of larger firm owners. [Porta & Shleifer, 2008] study new cross-country data on smaller-scale entrepreneurship that has been collected by the World Bank in its Enterprise Surveys. They argue that there is a significant division between individuals who operate informal sector enterprises and those who operate formal sector enterprises, and that there is little prospect for the vast majority of informal-sector self-employed to "move up" and run formal sector enterprises.

While the evidence on earnings distributions provides one source of preliminary evidence, in the remaining two subsections I focus on heterogeneity *within* the set of enterprise owners.

3.2.2 Heterogeneous earnings dynamics

In Figure B.1 I non-parametrically plot experience-earnings (net profit) profiles across these three qualitative enterprise categories, using a Lowess tri-cube smoother. The first group is the set of enterprises running with no employees, the second group is the set of enterprises only employing family/temporary workers, while the third is the set of enterprises that hire workers who are paid a wage.

There we see that all three groups enjoy an average increase in earnings over time as the sample is not corrected for selection on survivors. While perhaps remarkably all three enterprise types start off with roughly similar earnings, the rate of increase in earnings is substantially higher for those running the enterprises we would expect to be most complex (those that actually hire outside wage workers). This observed bifurcation in returns is suggestive of the select

group of individuals running more complex enterprises "pulling away" from the much larger group of individuals running enterprises in the other two categories. This is consistent with greater dynamism and perhaps even greater ambition among the set of more complex enterprises, along with other explanations such as more rigorous market selection pressures on larger firms.

It is interesting to note that while there seems to be a strong bifurcation of fortunes amongst enterprises, this seems to similarly be the case amongst individuals, in Table A.6. Looking at the year 2008, for example, we see that 17% of individuals currently running enterprises with waged employees had run an enterprise with no employees in the past, and 15% had run one with household/unpaid workers (non-exclusive). In fact, 10% had run *both* types of enterprises, indicating a substantial proportion who had followed an incremental entrepreneurial career trajectory – starting off with one of the simpler enterprise types and then "moving up". Yet at the same time, these figures show that the majority of individuals who end up running enterprises hiring waged workers immediately jump into that form of enterprise activity.

3.2.3 Heterogeneity in the returns to capital

In the IFLS it is difficult to rigorously measure the marginal returns to physical capital. However, I can provide suggestive evidence if we take average returns to capital (net profits divided by the value of the stock of physical capital in the enterprise) to be a reasonable proxy for marginal returns. This is a reasonable proxy assuming a constant returns to scale production function.[5] If we look just at micro-firms with three or fewer employees, we see that in the 2008 cross-

[5]Existing work on smaller firms in Indonesia suggests that assuming CRS is reasonable.

section 41% of enterprises have average returns to capital exceeding 10%, the national interest rate. 10% have greater than 60% returns, while 5% have greater than 108% returns. More detailed evidence is provided in Tables A.3 to A.5. The row for 'Unit returns to capital (%)' stratifies the same information by firm type for 2008 – firms with no employees, only family/unpaid employees, or those that hire waged employees. We see that, interestingly, the unit returns to capital in enterprises with no employees stochastically dominate the other two distributions. Indeed this is also true of unit returns to labor. And such firms seem to have greater proportionate increases in capital over their life cycle, at least at the upper end.

Greater returns to capital than some cost-of-capital benchmark is a necessary, but not sufficient, condition to imply that firms should invest more in capital. It may be that capital is lumpy – either physical capital itself, or the endowments that labor that complement the capital. While physical capital might be quite decomposable, it may be that moving from zero to one employee, for example, is quite drastic for a microfirm. Acquiring the complementary capital to the extra work could be relatively quite a significant step. Hence it might be optimal for firms to operate with relatively high marginal returns to capital. Another issue is volatility – to confidently invest in new physical capital, a firm would ideally have the sense that they will make profitable use of the capital year-by-year. If not, a bad year could wipe them out, even if one average they are better off with more capital. As noted in the Introduction, heterogeneity in returns to capital may also be indicative of heterogeneity in other endowments that are less observable. Finally, it may be that poor households simply have higher opportunity costs – they would rather invest new infusions of capital in other ways, such as in immediate consumption, or education, or health.

All in all what the evidence from the IFLS shows us is that there is tremendous variation in average returns to capital across micro and small firms. Even for those which seem to have returns which might justify greater investment, there may be completely natural market and internal frictions, and dynamic returns processes, which justify the current level of capital. A key goal of the current paper is to provide more evidence on the drivers of heterogeneity in enterprise investment activity.

3.3 A Simple Model of Credit-Constrained Occupational Choice

In the previous section I provided evidence that suggests the following hypothesis about small-scale entrepreneurial activity: the distribution of self-employed individuals can be broadly characterized by two ”types” – low-ability types who are largely involuntarily self-employed, and higher-ability types who have a significant opportunity cost to engaging in self-employment, yet still do so.

In this section I outline a model of credit-constrained occupational choice with this main hypothesis in mind, and describe the relevant predictions in the context of this paper. In particular, I focus on predictions related to the role of household wealth in enterprise choices, and positive financial shocks for the same. The subsequent analysis will provide evidence on these relationships in the context of the IFLS.

3.3.1 Outline of the Model

The standard model of occupational selection involving entrepreneurship has the individual choose between assigning a single unit of labor between one of two discrete occupations, wage work or self-employment.[6] The solution to the individual's occupational choice problem involves identifying the returns to each occupation, and then by simple comparison identifying which occupation has a larger return. The return to waged employment is denoted by a wage parameter, w.

A standard version of the value function, R, for self-employment is as follows:

$$R(W, \theta_i) = \begin{array}{l} \max_k pf(k, \theta_i) - (k - W)r \\ \text{s.t. } 0 \leq k \leq \lambda(W, \theta_i) \end{array} . \tag{3.1}$$

Here W represents the household's ex ante wealth, θ_i represents "entrepreneurial ability", p is the output price, f is the production function (taking physical capital, k, and θ_i as inputs), and r represents the market price of capital. The subscript i indicates that θ_i can take one of two values: $i = H$ refers to higher-ability types, and $i = L$ refers to lower-ability types, with $\theta_H > \theta_L > 0$. I assume that f is strictly increasing in both arguments, jointly concave, with the standard Inada conditions. It is notable that, for simplicity, I abstract from labor as an input. λ is a credit constraint which will be explained shortly.

I denote the level of k chosen to solve the above problem by $\bar{k}(\theta_i)$. The objective function allows for two classes of capital allocations: (1) setting $\bar{k}$ less

[6] Such a model has made prior appearances in papers such as [Lucas, 1978], [Kanbur, 1979], and [Evans & Jovanovic, 1989].

than W, in which case the household entirely self-finances business capital, with residual wealth ($W - \bar{k} > 0$) earning return r, and (2) having $\bar{k}$ greater than W, in which case the household must be borrowing capital (quantity $\bar{k} - W > 0$) at rate r. There is a potential limit on the household's ability to access outside capital beyond its own endowment W, which is expressed by the constraint $k \leq \lambda(W, \theta_i)$. Defining λ as a function is the one departure in the model above from the standard literature.[7] The mathematical properties of λ are as follows: λ is increasing in both arguments (strictly in W), $\lambda(W, \theta_i) \geq W$ (one can always self-finance), and λ is concave in both arguments. Specifying the constraint with these properties suggests that the ability to access outside capital depends on these endowments, perhaps through having greater collateral (W) and/or greater reputation (θ_i).

Having defined the payoffs, the occupational choice problem can be expressed simply as follows:

$$\max \{R(W, \theta_i), w\}. \tag{3.2}$$

If one wants to incorporate the role of risk and risk preferences in the model, it is possible to add a stochastic variable in the production function and define a utility function over returns. I abstract from this extension since the crux of the analysis in this paper is at the relationship between financing and occupational decisions, and good data on production risk are not available.

[7] Oftentimes λ is a parameter on W, with corresponding constraint λW, as in, e.g., [Evans & Jovanovic, 1989].

3.3.2 Solution of the Model

The model can be solved by backward induction. First, we derive the solution to the self-employment returns maximization problem. The key choice variable is k. If the (unconstrained) optimal value of k, $k^*(\theta_i)$,[8] is greater than $\lambda(W, \theta_i)$, then we set $\bar{k}(\theta_i) = \lambda(W, \theta_i)$, and say that the financing constraint is binding.[9] The implication is that the individual would select a larger capital stock to employ in the enterprise, were that to be available. If the constraint doesn't bind, then $k^*(\theta_i) = \bar{k}(\theta_i)$. Having solved for the optimal value of k, the value of $R(W, \theta_i)$ is then fixed. The occupational choice is then simply a matter of comparing $R(W, \theta_i)$ to w.

3.3.3 Predictions

In order to derive empirical implications of the single-agent model above, we need to consider the population distributions of the variables of interest. A key aspect of these models that often goes undiscussed is the relationship between W and θ_i. Often the relationship between these variables goes unaddressed, though the more reasonable assumption is that W and θ_i are fairly strongly positively correlated – on average it should be the case that more talented individ-

[8]The mathematical characterization of $k^*(\theta_i)$ is that

$$pf_1(k^*(\theta_i), \theta_i) = r.$$

[9]The mathematical characterization of this case is that

$$pf_1(\lambda(W, \theta_i), \theta_i) > r.$$

Notice that this condition formalizes the classic testable implication of a credit constraints model – that the marginal revenue product of capital is strictly greater than the marginal cost of capital.

uals would tend to have higher wealth.[10] This same reasoning applies to the wage, w – we would expect higher-ability individuals to have greater wage-earning opportunities, all things equal (i.e., each individual has some underlying "basic talent" that is applicable across both enterprise activity and working for a wage). Hence I assume that W, w and θ_i are all positively correlated in the discussion below.

Wealth shocks and enterprise activity

The first implication of the model pertains to the relationship between positive wealth shocks and enterprise activity choices (decision to enter, decision to invest, etc.). Here let us assume that we are considering a specific individual (with fixed θ_i, w, W) and that the individual faces a (positive) financial shock. In this case the implication is clear – all things equal, the positive shock should make the individual more likely to start a business, or to expand a given business. Hence if financing constraints bind, at least for some individuals, we should expect to observe a higher-than-usual positive response in terms of intensive and extensive margin enterprise choices in response to a positive wealth shock like the receipt of government transfers or lottery winnings.

Of course, there is some subtly here. First, transaction costs imply that individuals may not be long-run financing constrained, but they can appear to be constrained in the short-run if the cost of obtaining capital on their own is higher than the marginal value of the needed capital. Hence a sufficiently

[10] [Buera, 2009] provides microfoundations for the correlation between W and θ by adding an infinite-horizon dynamic extension to the standard model, allowing households to save over time (and hence own-save out of financial constraints). Under this dynamic extension the correlation between W and θ comes from the fact that higher-ability entrepreneurial types will have an incentive to own-save in the presence of financing constraints, hence building up asset reserves corresponding to their ability level.

small response to positive financing shocks might be uncovering basic transaction cost frictions. Second, it is important to think about whether a transfer is anticipated or not. If anticipated, an individual may actually make a business activity response *in advance* of the transfer – for example, perhaps the credible promise of an external financial transfer allows them to borrow funds from a social network member or money lender. This anticipatory effect could dampen the inferred role of credit constraints if we only look at enterprise response *after* the shock has been observed.

The second implication of the model pertains to the distribution of the effects of wealth shocks in the model – namely, for which individuals do we actually expect financing constraints to bind? This is the most novel prediction of the theoretical analysis.

It can be the case that high-ability households are relatively more responsive to positive financial shocks in terms of fixed capital investment. That is, formally, it can be the case that $(dk(\theta_H)/dW)/k(\theta_H) > (dk(\theta_L)/dW)/k(\theta_L)$.

This Proposition is somewhat surprising: it essentially says that high-ability individuals would be relatively more responsive to positive financial shocks, even though (with wealth correlated with ability) we would expect them to be less likely to be bound by financial constraints, all things equal.

The possibility that the case outlined in the Proposition holds depends on how the returns function varies in k versus how tight or slack the financing constraint condition, $0 \leq k \leq \lambda(W, \theta_i)$, is. In the extreme, suppose that the financing constraint is only binding for the high-ability household. This is possible if there is a relatively large gap between θ_H and θ_L, and λ is relatively invariant

in θ_i.[11] In that case, $k(\theta_L)$ will be much smaller than $k(\theta_H)$, so it is possible that $k(\theta_L) < \lambda(W, \theta_L)$ while $k(\theta_H) > \lambda(W, \theta_H)$, even though $\lambda(W, \theta_H) > \lambda(W, \theta_L)$. The value of $k(\theta_L)$ does not depend on wealth–the optimal value of the capital stock only depends on the parameters of the profit maximization problem. Hence even if there is a (positive) shock to W (for example, an exogenous transfer like from a government program), the low-ability self-employed individual may not respond by increasing capital stock. Meanwhile, since the high-ability individual actually faces a financing constraint, s/he will be responsive to the positive wealth shock.

Although other cases are possible this extreme example is sufficient to illustrate the claim in the Proposition. It is then an empirical question whether such relationships between wealth, wealth shocks and enterprise choices will be borne out in the data.

Note that the Proposition can be easily extended to the extensive margin question of occupational choice. The relevant of the simple, static model becomes a bit more tenuous in this case, though, because realistic predictions may rely on additional assumptions about the lumpiness of capital and the nature of the household's savings problem. As already discussed, the financing constraint can be critical in determining occupational choice–if it binds, it may drive down returns in self-employment, $R(W, \theta_i)$, to the point that $R(W, \theta_i) < w$. If the credit constraint binds, then it is possible for a wealth shock to change the value of R, which hence makes it possible to flip the inequality so that $R(W, \theta_i) > w$. If the credit constraint is only binding for the higher-ability household, then we

[11] To be complete, the result also relies on W not being too much larger for the high type than the low type (so self-financing isn't sufficient), and w not being too much larger for the high type than the low type (so the high type actually has an incentive to enter self-employment and realize the demanded capital level).

can immediately see how occupational choice might change.

In addition, however, are two considerations. First, higher-ability and wealthier households might be more likely to invest in firms that involve lumpier forms of capital, such as buildings and machines. Hence credit constraints may bind on them in different ways – as long as credit constraints don't allow them to invest at the level of minimal efficient scale then they may be deterred from investing. This consideration is not reflected in the model, and in any case not empirically testable with the data at hand. Second, wealthier households may be able to follow the strategy of starting an enterprise today, even if it is initially not very profitable, then plow profits back into the business to converge upwards to minimum efficient scale. The household's ability to do this depends on the nature of its savings problem, which is again not reflected in the model and not easily tested.

In practice I will carry out empirical tests that consider both versions of the Proposition in constructing dependent variables (intensive and extensive margin), through binary and continuous-dependent-variable models. The considerations I have just described should be borne in mind when interpreting the results from the binary-choice, extensive margin analysis.

Wealth stocks and enterprise activity

While the above Proposition pertains to the relationship between financial shocks (interpreted as shocks to the stock of wealth) and enterprise choice, the model also has implications for the relationship between the stock of wealth and enterprise choices. The key subtlety, however, arises from the role of w as

a measure of opportunity cost. First, assume that w is homogenous across the population. Then individuals' decisions between wage and self-employment depend on the returns to self-employment, which as we have seen depend on whether or not the financing constraint binds. It is always the case that high ability individuals have higher-returns from self-employment all things equal, $R(W, \theta_H) > R(W, \theta_L)$, since all key parameters of the model (θ_H, W, and λ) go in the same direction. Then the question is how the returns compare to w. We have three cases: (1) $w > R(W, \theta_H) > R(W, \theta_L)$, so no one would enter self-employment, (2) $R(W, \theta_H) \geq w > R(W, \theta_L)$, so only high-ability types would enter self-employment, or (3) $R(W, \theta_H) > R(W, \theta_L) > w$, so everyone would enter self-employment.

Incorporating the possibility that w varies by individual adds a layer of complication and potential censoring. Namely, let us now revert to the original assumption that w and θ_i are positively correlated, so wages for high-ability types are higher than for low-ability types, and denote these wages by $w_H > w_L$. In this case it can be that $w_H > R(W, \theta_H) > R(W, \theta_L) > w_L$. Namely, the individuals who can obtain the highest returns from self-employment will not enter, because their opportunity cost is even higher. Notice that this possibility does not change the unambiguous *latent* positive relationship between wealth and enterprise activity – it is just that w may act to obscure this relationship. This is the extreme case of negative selection, and it has a number of implications.

First, revealed preference alone is not a sufficient justification to judge the efficiency of the occupational choice configuration in any economy. Namely, the observed self-employed may not at all be those who are most skilled in that occupation. Second, the fact that there are many small, subsistence enterprises

that do not grow may not be an indication of exogenous constraints – it may just be that a negative selection mechanism is effectively forcing low-ability types into self-employment. This has the additional direct implication that attempts to remove posited external constraints may not be very effective. Indeed, recent literature, mostly based on randomized-controlled trials (RCTs),[12] has shown a number of interventions, from credit to business registration, to be relatively ineffective in changing the activities of apparently low-ability self-employed and potential low-ability self-employed.

Hence, unfortunately, the relationship between wealth and enterprise activity is ambiguous, due to the possibility that w is correlated with W. It is not clear that the propensity to start a business will be increasing in W, even if credit constraints *do* bind, if w is rising even faster and removing the incentive to switch into enterprise activity. Hence if we hypothesize that low-income economies are characterized by having large numbers of relatively low-ability individuals with weak employment prospects, we might expect to observe a large pool of subsistence self-employed who show little responsiveness to wealth, while a relatively small cohort of higher-ability individuals is more responsive to financing shocks. The empirical analysis below will look into these empirical relationships in more detail.

3.4 Data and Descriptive Evidence

In this section I provide a discussion of the dataset employed in the empirical analysis, the Indonesia Family Life Survey, pertaining to the issues at hand in

[12] Papers include [Banerjee *et al.*, 2009], and [Karlan & Zinman, 2010], [Carter & Olinto, 2003], among others.

this chapter. A more comprehensive discussion of the dataset is provided in Chapter 2.

3.4.1 Indonesia Family Life Survey: General Background, Characteristics and Context

My primary dataset is the Indonesia Family Life Survey (IFLS), a large-scale household survey from Indonesia with data collection rounds in 1993-94, 1997, 2000 and 2007-08.[13] For the intervening years between survey rounds, significant retrospective data are collected in the subsequent round. The dataset was designed to be representative of 83% of the Indonesian population in 1993. It covered 13 of 26 existing provinces in 1993, generally higher-population provinces in the western parts of the country, covering all of Java, most of Sumatra, and additional provinces in Kalimantan, Sulawesi and Bali. There was oversampling of urban locations and locations outside of Java, which is the country's main economic hub.

In the survey rounds data were collected at the individual, household, and community levels, and these three sources can be matched together. At the household level there is a wealth of information on issues such as consumption, assets, and labor market outcomes, both at the household-level and for select individuals within the household. The community and facilities survey collects information on issues such as aspects of the physical and social environment, infrastructure, employment opportunities, food prices, and access to health and

[13]Various organizations and researchers have been involved in designing, collecting and funding the IFLS. For more details, see [Strauss *et al.*, 2009], [Strauss *et al.*, 2004], [Frankenberg & Thomas, 2000], and [Frankenberg & Karoly, 1995].

educational facilities.

The original round of the survey in 1993-94 (IFLS1) surveyed 7224 households. Subsequent rounds have involved re-sampling the original households and then sampling all split-offs from the original households. Attrition has been relatively low, especially for a survey of this scope, at around 5% between rounds. Overall 87.6% of the original households appear in all four rounds. The sample expands in each subsequent round, as splits from the original households are tracked and surveyed. In addition, the proportion of household members directly interviewed also increases across rounds. By the 2007-08 round (IFLS4) the survey covers 13,535 households, with 44,103 individual interviews.

The IFLS covers a period of tremendous economic dynamism and upheaval in Indonesia. The decades preceding the 1990s were a time of significant demographic, social and economic change, under the regime of President Suharto. Per capita income had grown significantly since the 1960s, and hence massive improvements occurred on a number of dimensions of living standards. The poverty headcount ratio declined from 40% in 1976 to 18% in 1996, and improvements of similar scope were seen on indicators such as infant mortality, primary school enrollments, secondary school enrollments and fertility rates.

In the late 1990s Indonesia was caught in the grips of a major economic crisis, which affected much of Asia. The crisis first began to hit the financial sector in July 1997, but many of its effects on the real economy took until later in the year or into 1998 to take effect. Indonesia was worst-hit of all the Asian countries, experiencing a 13.5% decline of GDP and massive currency devaluation in 1998. Under tremendous political and economic pressure, President Suharto stepped down after 30 years in power in May of 1998. IFLS2 was fielded in 1997, just

prior to the onset of the crisis. A 25% sub-sample (IFLS2+) was collected a year later, and has been used to assess the early effects of the crisis; however, these data are not publicly available.

What followed the crisis was a period of political and economic transition. The authoritarian government of the Suharto era gave way to a new period of democratic elections under a multi-party system. A number of economic and political reforms were carried out, perhaps most significant among them a large decentralization of powers to the regions, which took effect in 2002-03. As these reforms came online and there was broader regional recovery, the economy began to grow quickly again, with GDP growth rates between 4.5% and 5.5% in the years until 2007. Indeed, Indonesia was one of only three major economies (China and India being the others) that experienced positive economic growth (over 4%), during the global financial crisis of 2008-09. The most recent available round of the IFLS, IFLS4, was collected just prior to the onset of the crisis, in 2007-08.

3.5 Assets, Wealth and Entrepreneurial Choice

In this section I carry out asset and wealth-based tests of financial, studying the relationship between stock variables in the household's financial portfolio and intensive and extensive margin enterprise choices.

3.5.1 Empirical Model

Theoretical Setup

The empirical model is based on regressing an enterprise decision variable, such as the (binary) decision of a household to startup a business (extensive margin), or the (continuous) amount to invest in startup capital (intensive margin), on a measure of wealth and a vector of household characteristics. I allow the wealth variable to enter the regression equations through a semiparametric estimation approach that allows the relevant enterprise decision to be a fully-flexible, non-linear function of the wealth variable.

My approach is motivated by recent papers that point to a highly non-linear relationship between (lagged) household wealth and enterprise outcomes. One set of papers uses non-parametric techniques to model the relationship between wealth and enterprise outcomes (e.g., [Paulson & Townsend, 2004]). While highly flexible, such approaches suffer from the well-known dimensionality problems in non-parametric kernel approaches that importantly do not allow for additional linear controls to enter the regressions. Hence such evidence is highly descriptive and may suffer from significant omitted variable bias.

An additional set of papers uses flexible parametric models (usually a 4th or 5th-order polynomial functional form) to model the relationship between wealth and enterprise choices (e.g., [Hurst & Lusardi, 2004]). By working in a more tractable parametric framework these empirical models allow for additional linear controls to enter the model. However, given the significant non-linearities uncovered in this literature, we might still be concerned about over-smoothing of non-linear effects given the imposition of a parametric specifica-

tion.

The approach described in this section manages to combine the virtues of each approach – the great flexibility of the non-parametric approach in capturing the effects of the main explanatory variable(s), along with the ability to tractably control for additional observable heterogeneity in a parametric framework. The approach is built on a semi-parametric approach in which non-linear effects can be incorporated tractably through linear combinations of flexible parametric functions.

I will describe the modeling framework in the context of a binary choice model, in particular focusing on how the semiparametric component is incorporated in the model. The outlined approach applies similarly to setups with a continuous dependent variable, with the only difference being the familiar difference between a binary choice model and a linear model. Hence I omit the development of the case of a continuous dependent variable, simply pointing out the obvious adjustment to the model setup.

In the empirical model y is the latent (unobserved) variable representing the discrete enterprise choice, taking the value 1 if the individual chooses the enterprise activity at hand, 0 otherwise. I drop the subscript i representing individual observations to simplify notation. y^* represents the true threshold condition, with the threshold normalized to 0. y^* is assumed to be a function $f^*(W, x) - \varepsilon$ of wealth, W, and other observable covariates, x. The choice-based version of the empirical binary choice model is then as follows,

$$y = \begin{cases} 1 & y^* \geq 0 \\ 0 & o.w. \end{cases} = \begin{cases} 1 & f^*(W, x) \geq \varepsilon \\ 0 & o.w. \end{cases}, \tag{3.3}$$

where ε represents the distribution of unobservable influences on choice, and noise.

My objective is to estimate f^*, given an assumed distribution on ε. I make the following (linearizing) functional-form assumption on f^*:

$$f^*(W, x) = g^*(W) + \beta^{*\prime} x,$$

where g represents a (possibly highly nonlinear) function of wealth and $\beta' x$ represents the contribution of other covariates in the usual (linear) parametric way.

We can also interpret this model in terms of the probability of choosing $y = 1$,

$$\begin{aligned} P[y = 1|W, x] &= P[g^*(W) + \beta^{*\prime} x \geq \varepsilon|W, x] \\ &= F_\varepsilon^*[g^*(W) + \beta^{*\prime} x]. \end{aligned} \tag{3.4}$$

For identification purposes, assume F_ε^* is standard normal, which gives location and scale, g^* is real-valued continuous but unknown, and the full set of covariates has full support. provides general identification results on this class of models.

Estimation

The estimation procedure, allowing for the flexible estimation of $g^*(W_i)$, is carried out using a semiparametric approach through the use of penalized splines. The essence of the penalized spline approach is that in addition to the usual regression optimization problem fitting a function of covariates to a response

variable, it also constructs a penalty matrix, with a parameterization determining the weight of the penalty. Penalization accounts for the fact that such flexible estimation is susceptible to the overfitting trap, and hence "wigglyness" of g is penalized. The estimation is carried out in the R statistical package, using the 'mgcv' (generalized additive model) package developed by Simon N. Wood of the University of Bath.

As one might expect, the results of such models cannot be fully presented using the conventional regression table, since the estimands of interest are not merely regression parameters. The conventional linear estimation coefficients (β^* above) can be presented in standard regression tables, with conventional standard errors and attendant tests of statistical significance. The presentation of the estimated function (g^* above) is similar to a non-parametric approach – graphical. We can display a graph of the estimated function over its domain, and then report on certain properties of the function. The standard presentation is an "influence graph" – a function which plots the influence of variation in the key right-hand side variable (here, W_i) on the outcome represented on the y-axis.

Continuous Dependent Variable Case

The continuous variable case is very similar to the above, essentially just removing the complication of the model structure needed for a binary choice model. The empirical model hence takes the form,

$$Y_i = \beta' x_i + g(W_i) + \varepsilon_i, \tag{3.5}$$

in the case of a continuous left-hand-side variable like startup capital, where Y_i is a continuous outcome (e.g., startup capital, or current capital in the enterprise), g is an arbitrary function of wealth, W_i, $\beta' x_i$ represents additional linear covariates, and ε is a normally-distributed error parameter.

3.5.2 Empirical Implementation

In practice I need to account for the concern that at least part of W can be simultaneously determined with the outcome variable. As in previous papers (e.g., [Paulson & Townsend, 2004]) I exploit the panel nature of the data to take the lagged value of W as a proxy for current capital, which is clearly independent of current decisions. The lagged value of wealth is the value of wealth in the previous round of the survey. So, for example, in studying enterprise decisions in IFLS4 (2007-08) I would be using wealth from IFLS3 (2000) on the right-hand side. All values in the study are converted to 2005 dollars, so they are on a common index.

In practice I measure W through a broad set of measures of household wealth available in the IFLS, including the value of large assets such as land, buildings, and automobiles, major household durables like appliances, and additional forms of wealth such as jewelry and bank accounts. I also include business wealth.[14]

[14]Specifically, I include all of the following variables in the measure of wealth, which I distinguish here by plausibly liquid and plausibly illiquid wealth variables:

Liquid wealth. Household poultry, household vehicles, household appliances, savings/certificate or deposit/stocks, receivables, jewelry, household furniture and utensils, other household assets, business four-wheel motor vehicles, business other vehicles, business other non-farm equipment.

Illiquid wealth. House and land occupied by household, other house/building, non-farm land, business land, business building.

The measures are constructed by summing each of these values, and then deflating them as

I also provide two measures of wealth: (1) wealth in levels (i.e., total wealth of the household), and (2) wealth per head, with total wealth of the household divided by adult equivalent units. The latter measure provides a useful correction that controls for heterogeneity induced simply by household size. It turns out that this distinction is not so important empirically.

3.5.3 Results

All results appear in the Appendix. There are four sets of results: (1) the effects of wealth (in levels) on enterprise startup, (2) the effects of wealth (in logs) on enterprise startup, (3) the effects of wealth on startup capital, (4) the effects of wealth on current capital (as of 2008). For each of these four sets of results I present six graphs – three of which are from the results with wealth in levels, and three from the results with wealth per head, giving the graphical representation of the result of the semiparametric fit. Since the results from wealth per head are generally qualitatively consistent with those from wealth in levels, I focus here on the results from wealth in levels. Error bands at two standard deviations are also included. The corresponding tables for these sets of results appear in Tables B.2 to B.9, which include the estimates for the parametric coefficients, and information on the quality of fit.

First, I look at the relationship between wealth and enterprise activity in terms of the propensity to start an enterprise, in Figures B.2 to B.5 and corresponding Tables B.2 to B.5. More specifically, I study the propensity of a household to start a new enterprise between 2001 and 2008, conditional on their wealth in 2000. This follows [Paulson & Townsend, 2004], who study the re-

appropriate with the price deflator used throughout the survey.

lationship between wealth and enterprise activity through bivariate, nonparametric analysis and multiple regression analysis.[15] I find that the hypothesized relationship, that the propensity to startup an enterprise should be monotonically increasing in wealth, does not seem to hold.

Over approximately the first 60 percentiles of the wealth distribution there is almost no relationship between wealth and enterprise activity. This suggests the poorest households, which we would expect to be *most* constrained by access to finance, are not more likely to startup enterprises based on incremental increases in wealth. One possible explanation is that starting any enterprise requires significant, "lumpy" investments, and so the non-response simply identifies the subset of households of such low wealth that they cannot afford the minimal necessary fixed investment. However, given that the relationship between wealth and startup capital is similarly flat for poorer households, this explanation requires strong assumptions about the "lumpiness" of necessary capital. Also, the majority of enterprises in the data operate with little formal capital stock at all, suggesting that many enterprise forms are attainable with little to no physical capital.

Interestingly, the relationship between wealth and enterprise startup propensity is much stronger over the middle of the wealth distribution, while again tapering off in the higher wealth percentiles, before the data become sparse. The evidence above about the 60th wealth percentile is quite consistent with the evidence from [Hurst & Lusardi, 2004], who find that financing does not seem to be a significant hindrance to most enterprise startups in the US.

[15] [Paulson & Townsend, 2004] find some evidence that is supportive of the standard model of credit-constrained occupational choice, though their dataset is dominated by firms with 0, 1 or 2 employees.

Tables B.6 to B.9 move to continuous left-hand-side variables, with Tables B.6 to B.9 presenting the corresponding regression coefficients. In Figures B.6 and B.7 we have the startup capital of the enterprise as the key dependent variable, while in Figures B.8 and B.9 we have current capital in the enterprise. In all cases we see little activity over the first 50 percentiles of the wealth distribution, then see greater response at higher percentiles.

These findings are supported by recent evidence in the literature. [Carter & Olinto, 2003] show that the enterprise investment response emanating from a policy change meant to strengthen property rights (and hence the ability of households to collateralize assets) is concentrated amongst wealthy households. [de Mel *et al.*, 2008], [Banerjee *et al.*, 2009], and [Karlan & Zinman, 2010] present evidence on random financing shocks to microentrepreneurs and randomized rollout of microfinance services, showing in part that the response is concentrated among higher wealth households or those who are already engaged in enterprise activity. This points to the idea that there must be something more at work in driving enterprise behavior than just financing constraints.

The results on the estimates from the parametric terms in the models are consistent with existing literature. Notably, the dummy for urban location is has the largest economic significance on the propensity to startup, indicating greater urban enterprise activity and perhaps churning, which is consistent with more competitive markets. There is a concave relationship between age and the propensity to startup, which is highly consistent with existing literature. Notably, education bears a positive relationship to startup. The availability of household labor (which also implies the need to find work for additional labor) is unsurprisingly positively related to startup. Looking at the continuous

dependent variables, we see that education of the household head and available male labor are the only two variables that significantly (positively) predict startup capital. Education of the head and available labor are perhaps unsurprisingly related to current level of capital.

While the evidence that emerges from this analysis is interesting, it suffers from the drawback that, in a dynamic sense, wealth and enterprise decisions are still co-determined. [Buera, 2009] best highlights the omitted variable bias that can result from studying the relationship between wealth and enterprise activity. Namely, in a dynamic model with savings, if financing constraints bind then households with relatively large stocks of entrepreneurial skill have a relatively larger incentives to ”save up” for enterprise investments (particularly if those investments are generally ”lumpy”). But this implies that households with high ability should (1) be more likely to engage in enterprise activity, and (2) be relatively more wealthy, even before entering self-employment. If measures of ability are omitted from the regression, we can draw the false impression that it is wealth (and hence financing) alone that is driving entrepreneurial dynamics, when in fact wealth is endogenous in this longer-term problem.

In order to partially deal with this issue, in the next section I move away from stock-based measures of financing.

3.6 Income Shocks and Entrepreneurial Choice

In this section I present tests of financial constraints based on various income flows. This includes flows that are plausibly anticipated or conditional on some characteristics of the individual, including government transfers like uncondi-

tional cash transfers and non-government wealth transfers. I also look at the role of income shocks of a more random nature, such as bonuses, lottery winnings and insurance payouts.

The basic idea behind these tests is related to a large literature that attempts to test for financing constraints using exogenous income shocks. The idea is that if a household is financing-constrained in business activity, then a positive income transfer should lead to an increased propensity to startup or invest in a business. In a world in which all of the financing needed for the business is freely available, we would not expect positive financial shocks to have an effect on the business, because using the financing on the business would not be productive. The financing could be more efficiently used on household consumption, for example.

I initially look at correlations between positive income shocks and enterprise response across a variety of specifications. While overall significance of effect and direction are merely indicative in this case, it is still helpful to look for significant effects and direction of effects. At this stage I find that some financial flows seem to have effects on enterprise activity (unconditional cash transfers, and the more random financial flows noted above like lottery winnings and bonuses) and others don't. I then drop my pursuit of the seemingly non-effectual flows, and attempt to address the role of selection in drawing more rigorous inferences from each of the two seemingly effectual flows.

The simple inference that "if positive shocks lead to more investment, the household is financing constrained, and if not, then not" is complicated by at least two factors in observational data. The first is the role of anticipation. If a positive financial shock is anticipated, then it might be possible for the house-

hold to smooth income by contracting on the expected financial flow (e.g., borrowing from a money lender or family member). This would bring the financing effect forward, which would mute the observed effect of the financing transfer itself. Hence the implication for inferences from shocks is that it should be plausible that a transfer is unanticipated, or else financial markets should be sufficiently imperfect that *ex ante* contracting is not fully available. However, if a kind of transfer is plausibly anticipated, it is sufficient to show that there is heterogeneity in effects, particularly if we find little effect at the low end of the wealth distribution. This is because the poor would be most likely to face distorted financial markets and hence greater difficulties in contracting on anticipated financing.

The second complicating factor is the role of opportunity cost, particularly heterogeneity in opportunity cost. Namely, it may be that in response to a financial transfer a household does not invest in business activity, not because the business could not plausibly make use of capital at market rates, but because some other need in the household carries even greater marginal return. It could even be the case that wealthier households invest relatively more in enterprises even if they are relatively less financing constrained, because poorer households have more pressing needs to initially attend to than the enterprise.

This issue is challenging, even in the context of randomly-assigned financing treatments. Any subsequent inference on treatment response that varies according to an observable variable like wealth may be driven by heterogeneous opportunity costs that are correlated with wealth. This means that the selection of the study population can have important effects on results, even if enterprise returns and costs of borrowing are roughly similar across settings. I attempt

to control for this unobserved heterogeneity through identification approaches that are well-suited to the financial flows that seem to be effectual. While this cannot rule out heterogeneous response that is correlated with wealth, inferred heterogeneity in treatment response can still indicate that certain households have relatively higher opportunity costs and enterprise may not be the highest-value investment. This has important implications for maximizing efficiency in program targeting.

An additional potential complication may arise due to size of needed investment. Namely, it may be that a poor household would like to invest positive income shocks in the business, but the desired investment is lumpy and its price exceeds the value of the transfer, making the desired purchase unattainable. While this possibility is present and difficult to rule out with certainty, it does not seem totally plausible. In the data we see that lower-income households commonly run simple, subsistence enterprises with little fixed capital. It seems unlikely that such enterprises, if financing constrained, would not be able to identify incremental investments that increase productivity.

Hence the results in this section should be interpreted with these caveats. Selection control strategies have the potential to account for the potential role of unobserved heterogeneity in driving observed outcomes, increasing our confidence in the results.

3.6.1 Empirical model

I begin by providing correlational evidence on the potential role of positive income shocks, employing binary-choice empirical models of the form

$$Y_i = \Phi\left(X_i\beta + s_i\gamma\right), \tag{3.6}$$

and continuous models of the form

$$Y_i = X_i\beta + s_i\gamma + \varepsilon_i, \tag{3.7}$$

where Y_i represents the binary enterprise formation decision or the continuous starting capital decision, Φ is the functional for a binary choice model (I employ probit), X is a vector of controls, and β is a vector of estimands. s_i represents a vector of (lagged) shocks to household income, while γ captures the effect of the shock(s) on the household decision. In the analysis that appears herein s_i is primarily coded as a dummy variable.

3.6.2 Empirical Implementation

Income Shocks

Here I summarize the kinds of income shocks, the key right-hand variables, considered in the analysis. First, note that I date the transfers around the cutoff year of 2000. In general, transfers are classified as pre-2000, or '07-'08, representing

the years in which the transfers occurred. The categories of transfers are:[16]

- Gov. trans. = sundry transfers of money from the government, apart from conditional cash transfers and unconditional cash transfers.
- Non-gov. trans. = an aggregation of transfers received from non-government sources; most commonly family and friends.
- Uncond. cash trans. = unconditional cash transfers.
- Exog. trans. = the most plausibly randomized forms of wealth transfer – lottery, bonus, and insurance payouts.

Further Issues

Sometimes I distinguish "ent. 1" and "ent. 2" – this captures whether the startup decision pertains to the first enterprise owned by the household, or the second.

3.6.3 Correlational tests for credit constraints based on positive income shocks

Results are reported in regression Tables B.10 to B.18. In Tables B.10 to B.14 I employ binary models (probit), so I report both the coefficient and the marginal effect. Tables B.15 to B.18 contain the results on the starting capital decision.

In the probit formation setup, I find that the relationship between (lagged)

[16]Conditional cash transfers are excluded from this part of the analysis as they are more difficult to code – they only began to appear post-2000 (i.e., 2003, 2005) and hence are entangled with enterprise decisions made in the post-2000 period.

sources of potentially expected income transfers and enterprise activity is generally negligible for government transfers and non-government transfers

The strongest noted effects on startup are obviously for unconditional cash transfers. In Table B.10 the effect seems very large – roughly a 50% marginal effect increase in propensity to startup an enterprise subsequent to an unconditional cash transfer. This is also confirmed in Tables B.11 and B.12, which expands the set of transfer measures. Interestingly, the wealth interaction for unconditional cash transfers seem to be negligible. However, it is also notable in Tables B.11 and B.12 that *future* receipts of unconditional transfers (in 2007-08) are *also* correlated with enterprise activity, and negatively.[17] This finding suggests two things. First, there is something special about the set of individuals who tend to receive unconditional cash transfers, even after other controls, that leads them to serially obtain cash transfers and be active in business. Second, given the negative coefficient on the 2007-08 transfers, if those trapped in poverty are more likely to be cash transfer recipients in the future, then those individuals seem to be significantly less likely to engage in enterprise activity.

Now I look at starting capital decisions in Tables B.15 to B.18. We first can see that as in Table B.15 I break out the results by percentile and find that once interacted with household wealth, there seems to be a positive effect of government transfers on enterprise startup activity near the lower end of the wealth distribution, with this effect increasing in wealth. A less detectable effect also exists for the (quadratic term in) non-government transfers in the same table. However, the dominant effects are again for unconditional cash transfers. While there is no effect in Table B.15 in the first-order term, there does seem to be a "convex" effect

[17]The strategy of incorporating future lags into such regressions is analogous to Hurst and Lusardi (2004).

in wealth. This is interesting because it suggests a negative marginal propensity to engage in self-employment for the poorest of the poor, but that this effect is removed as we move up the wealth distribution. Effects in the other percentiles are not statistically significant though this may be partly because individuals in those wealth percentiles are less likely to receive the transfers. In Table B.17 we see little effect of unconditional cash transfers, at least in the mean regressions displayed in Table B.15.

3.6.4 Exogenous financial shocks

I exploit more plausibly exogenous sources of income shocks, such as lottery winnings, insurance payouts, and bonus payouts. This use of exogenous shocks is analogous to tests in the recent literature, which have looked at the effects of exogenous transfers of financing through natural experiments or randomization (e.g., [de Mel *et al.*, 2008], [Banerjee & Duflo, 2010]). I do find evidence that such income transfers seem to (positively) predict enterprise activity. However, in addition I present evidence that the propensity to respond to such exogenous wealth shocks is "increasing" in wealth. That is, fitting with other literature on exogenous transfers, it appears that the response in terms of setting up or investing in enterprises is actually concentrated amongst higher-wealth households. A sample of this evidence is presented in Table B.10. I report the coefficient estimates and marginal effects from interaction terms between (lagged) wealth and the receipt of positive shocks, in probit enterprise startup regressions. I look at the startup of both a household's first and second enterprise. For the first enterprise, I find that the positive relationship between wealth and the enterprise activity only disappears beyond the 99th percentile of the wealth

distribution, which directly opposes the prediction of the standard model of financing-constrained enterprise activity. Interestingly, for the second enterprise the expected negative relationship between wealth and startup returns, which reinforces the idea that more experienced entrepreneurs are relatively more constrained by access to finance.

3.6.5 Selection and heterogeneity of effects

The preceding analysis provides suggestive evidence about the role of financing in enterprise outcomes, by looking at correlations between the receipt of positive income shocks the timing of which is exogenous to the household, and enterprise outcomes. The concern with drawing causal inferences from this analysis is of course that treatment assignment is likely to be correlated with observable and unobservable characteristics of the household. Based on the correlational evidence, I will focus on two forms of transfers, which show strong effects in correlation: unconditional cash transfers and what were called exogenous transfers (lottery winnings, bonuses, insurance payouts). For each of these two forms of transfers, I develop an identification strategy that is appropriate to the selection processes that are likely at work.

Exogenous financial shocks

In the of transfers like lottery winnings, bonuses and insurance payouts, the IFLS does not provide information on the full potential recipient population. That is, we do not know which households enter lotteries, take out insurance, or are in occupations where bonuses are available, just which households get

such payouts. Given that the receipt of such transfers likely varies from year-to-year, it is very likely that there are individuals in the sample who have potential access to such transfers, but do not happen to receive payouts in a certain year or set of years. Hence in attempting to construct a counterfactual to the group of households that receive the transfers, I employ a propensity score matching strategy which balances the characteristics of the treatment and control groups on observable factors that seem likely to be related to selection into access into these kinds of transfers.

Fortunately, the IFLS provides a rich set of observables on which to statistically match subjects in the propensity score approach. The key identifying assumption in the PSM approach in a causal framework is that conditional on available balancing variables, it is as if the treatment is randomly assigned. Since this assumption is not possible to test the credibility of a PSM exercise will rely on the plausibility of the balancing variables chosen.

The variables I focus on are as follows. First, it seems very likely that selection into accessing "exogenous transfers" is related to risk and time preferences. Fortunately, the IFLS4 contains a number of questions on risk preferences and time preferences (8 for each) based on hypothetical scenarios, and I balance on each of the response variables within the household head.[18] Second, I also balance on additional characteristics of the household head: age and years of education. Finally, I balance on characteristics of the household: a dummy for urban status and lagged wealth. In the stage in which I calculate the propensity score, I require balancing to hold at the 0.01 level, and the listed control variables gen-

[18] An example of a risk question is the choice between an option of earning 800 thousand Rph. per month for sure, versus a gamble between 1.6 million Rph. per month and 400 thousand Rph. per month with equal probability on both outcomes. An example of a time question is for the subject to suppose they have won the lottery and can choose between 1 million Rph. today or 3 million Rph. in 1 year.

erated balanced treatment and control groups for ”exogenous transfers” under this criterion.

I use the stratification method to calculate the average treatment effect. The program generates 7 blocks, with 8725 observations in total, with 95 treatment and 8630 control. As noted, mean propensity score is not different for the treatment and control groups in each block, at the 0.01 level.

The estimated ATT is 0.069. The standard errors need to be bootstrapped, and this procedure yields a standard error on the estimated parameter of 0.044, with a t-statistic of 1.578, indicating a statistically-significant, positive treatment effect of ”exogenous transfers” on enterprise participation response.

However, the estimated effects are much smaller than the correlations indicated in the preliminary analysis, where the marginal effect of receiving exogenous transfers on enterprise participation was on the order of approximately 0.4. It appears that the selection correction has removed a significant amount of the relationship between the receipt of exogenous transfers and enterprise response. In tandem with the suggestive evidence that responsiveness to financial transfers is increasing in wealth, it appears that much of the response is driven by relatively wealthier households. These households would be more likely to have the disposable income to participate in lotteries, purchase insurance, and might be more likely to be in occupations in which bonuses are available. Apparently it is this group that is driving much of the treatment response, in line with emerging results in the literature on microfinance and microentrepreneurship.

Unconditional cash transfers (UCT)

In the case of unconditional cash transfers it is possible to control for treatment assignment based on observables, at least in principle. The targeting of UCTs in Indonesia was based on an objective index,[19] with most if not all of the criteria appearing in the IFLS survey. Of course the credibility of individual assignment that can be explained by observables rests on the assumption that the program was relatively well-targeted. It also rests on the assumption that lagged values of the targeting indicator variables (recorded in IFLS3 around 2001) provide good measures for the indicators in 2005 and later, when the Indonesian government implemented a major cash transfer program. Conditional on these assumptions holding, selection into the UCT treatment can be directly controlled by observables. Indeed, regression analysis on selection indicates that the selection equation into receipt of the UCT has relatively high R-squared.

Results from the UCT approach are presented in Tables B.19 and **??**, which do not and then do control for selection into the UCT, respectively. There we see that once program selection is controlled for there is no statistically-significant effect of receipt of UCT on enterprise activity.

[19]There are 14. criteria for receiving cash transfers: size of house (square meters), flooring material of house, material used for walls of house, sanitary facilities in house, source of drinking water, source of main lighting, kind of fuel used for daily cooking, source of main lighting, how many times a week the family buys meat/chicken/milk, how many times per day the family eats, how much new clothes the family buys for a majority of household members per year, financial ability to go to the clinic if sick, main job of the head of family, and possession of specific assets worth over 500.000 rupiah (about $50 USD) – savings, gold, color TV, livestock.

The household is also asked about the name of the head of family, education level of the head of family, number of family members, children aged 7-18, and females 10-49 in the household, and whether they are married.

3.7 Additional Factors in Entrepreneurial Selection and Choice

The existing analysis thus far is consistent with the hypothesis that lower-income households aren't primarily constrained by financing in their entrepreneurial activities. There seems to be little relationship between (lagged) wealth and enterprise choices in the lower percentiles of the wealth distribution, while exogenous income shocks seem to have more of an influence on higher-income households. This then raises the obvious question: if heterogenous financing constraints, impinging more severely on lower-income individuals, do not explain observed enterprise patterns, then what else could it be?

In this section I provide some initial evidence in this direction, exploring a number of potential factors. This starts to provide evidence that accounts for the limitations of the empirical analysis that we can carry out on observational data, by attempting to account for some potential key unobservables that are available.

The first three factors are primarily lifestyle and context-based. First, I look at the role of family, with the idea in mind that perhaps something about entrepreneurship is passed between generations. Indeed, in much of the developed-country literature on entrepreneurial selection, having a parent (particularly a father) who is an entrepreneur seems to be the strongest predictor of entry into enterprise activity of the common measurables. Second, I look at the role of gender. Perhaps it is the case that some of the micro-enterprises we observe are started by women just looking for a side-business to be run out of the home, while watching children. Such businesses are severely limited in their ability to grow and may not be meant to grow at all, due to the woman's other

time commitments and responsibilities. This hypothesis is consistent with findings in the literature on microfinance activity, that when an impact of microfinance is observed, it is usually more concentrated amongst male borrowers (see, e.g., [de Mel *et al.*, 2008]). Third, I look at the relationship between non-farm enterprise activity and farm-based enterprise activity. This is motivated by the possibility that non-farm enterprise activity amongst agricultural households may just be a revenue-diversification activity, and not a primary focus for income-generation.

The second set of factors look more internally at the individual, providing attempts to measure unobserved preferences and human capital. The fourth factor I consider is the role of ”behavioral” factors – risk and time preferences. Note that the strength of the inferences that can be drawn from these variables is limited by the fact that they are only collected in IFLS4 (2008). Fifth, I look at the role of raw cognitive ability – this could be a significant source of variation in entrepreneurial outcomes if raw cognitive ability is an important factor in entrepreneurial activity. Finally, I consider an alternative source of entrepreneurial skill accumulation – direct experience. The idea is that learning-by-doing is critical to building entrepreneurial skill. Hence individuals with more experience, especially in running more complex (which I proxy by size) enterprises should be relatively more successful.

3.7.1 Familial Effects

I hypothesize that the familial channel is the primary institution for the transfer of human capital specific to entrepreneurship (taking direct learning-by-doing

to be about accumulation rather than transfer of human capital specific to entrepreneurship), particularly in the developing-country setting. If frictions in the labor market create a hindrance to "outsiders" working in family enterprises, children could be end up much more likely to work in the family enterprise. But if this is the case, then the child may be more likely to accumulate valuable human capital specific to entrepreneurship. We see some evidence for this in recent analysis using US data. [Dunn & Holtz-Eakin, 2000] and [Fairlie & Robb, 2007b], [Fairlie & Robb, 2007a] look deeper into the strong propensity of children of self-employed individuals to become self-employed themselves. While having self-employed parents is perhaps the strongest predictor of a child's self-employment propensity in the empirical analysis of entrepreneurship, such a correlation could be driven by at least two channels. First, entrepreneurs tend to be wealthier and hence parents might help their children overcome financial constraints (or directly transfer the enterprise itself). Second, there may be some kind of valuable non-monetary transfer between parents and children. The authors' findings strongly favor the second hypothesis. Direct transfer of the enterprise between parents and children is actually quite rare (less than 5% of children's enterprises start this way). Overall wealth of parents, or financial transfers from parents and children, do not seem to predict enterprise activity or the success of the children in enterprise activity. In addition, the success of parents in enterprise activity strongly predicts children's success, even after controlling for wealth. Also, whether or not the child actually worked in the family enterprise while growing up predicts both greater propensity for self-employment and later success. This empirical work provides strong evidence that something valuable is transferred between entrepreneurial parents and their children, though of course it is still not clear from this work

what is being transferred, whether human capital, or perhaps genetic code.[20]

I present preliminary evidence related to familial effects in Appendix B. I show that there is a significant increase in the propensity of children to become self-employed based on having self-employed parents. I begin by presenting a cross-tab that bins the observations of parents and children according to whether or not parents have certain kinds of enterprise experience and the subsequent outcomes of their children in Table B.21. The results are clearer in the Table B.22, which collects probit regression results reporting the propensity of children to be self-employed as a function of their parents' self-employment status. We see that the marginal effects are substantial relative to the baseline self-employment propensity in the sample, which is about 10%. If one's father was most recently operating a single-proprietor enterprise, then one is 33% more likely than the average to engage in self-employment. The comparable effect is 60% for father's status running an enterprise with household members working in the enterprise. Though the effect for having parent self-employed in an enterprise hiring in waged workers is not significant (probably due to sample size concerns) we still say a large, marginal effect on the order of a 50% increase. The mother effect is analogous. Perhaps surprisingly, the marginal effect on the interaction of having *both* parents self-employed is negative. It is possible that having both parents self-employed is correlated with other statistical patterns (e.g., higher poverty) that would bias the coefficient. But even then the large individual effects seem to overwhelm this larger, apparently negative overall effect.

As an initial attempt at teasing out whether the "parental transfer" is primar-

[20]There is emerging work on the role of genetics in entrepreneurship. For one of the first published papers see [Nicolau *et al.*, 2008]. This paper, based on twin data, finds that about half of entrepreneurial propensity can be explained by genetic factors.

ily financial or something else, I look at the subsequent earnings of children in self-employment as a function of their parents' self-employment status in Table B.23. I find that children of fathers who run enterprises with wage workers earn on average about $127 US equivalent per month more than those with parents in other categories. Unfortunately, other coefficients are not statistically significant.

These results provide preliminary evidence that parental skill matters both for their children's later propensity to be self-employed, and there performance therein.

3.7.2 Gender

To consider the role of gender I first focus on households headed by men and women, and only the primary enterprise run by the household. This removes potential biases from comparing primary and secondary enterprises, and maintains the focus on the primary breadwinner. Across all measures I find that female-run business are smaller and less prolific – they are less likely to operate outside the home (68% for female-run to 83% for male-run), less likely to apply for business permits (47% to 52%), startup with less household/temporary and wage-workers (0.62 to 0.67 and 0.2 to 0.8, respectively) and startup with 30% the level of physical capital stock, similarly have less current workers in both categories (0.59 to 0.68 and 0.27 to 1.08) and currently have less than half the capital stock, and have 71% the amount of earnings. Along these same lines, we find that if the primary enterprise is run by the male household head it is larger and

more prolific on all expected measures[21] than if it is run by another household member, which is almost always the wife.

3.7.3 Risk and time preferences

To look at the role of risk and time preferences in enterprise outcomes I employ the same risk and time measures that were employed earlier in this study. The idea that risk and time preferences would matter for enterprise activity seems intuitive. Running a business can be an incredibly risky and uncertain activity, and so attitudes to risk may govern individuals' willingness to participate or what actions they take. I also expect time preferences to be relevant, since individuals' patience might determine their willingness to stick with an idea or wait for a business to grow.

These variables are based on the standard approaches for eliciting risk and time preferences: subjective choices when faced with hypothetical gambles and time receipt of money. The results come with a key caution: the risk and time preference variables are only available in IFLS4. If we think that risk and time preferences are stable over time then this is not a problem, but we might expect that risk and time preferences vary over time, in ways correlated with actions (including enterprise activity choices). In any case, it is interested to look at the results in a descriptive sense.

Initial results on selection into self-employment indicate that selection into entrepreneurship is positively correlated with a willingness to take risk. Of the two risk-choice questions that lead to responses that are statistically signifi-

[21] Only smaller in terms of household/temporary workers employed, which could be due to children working in the family enterprise.

cantly correlated with entry into self-employment (after controlling for a range of covariates), both involve the subject choosing the "more risky" option by taking the gamble. In terms of time preferences, only the response to one question is statistically significant. This coefficient estimate suggests that selection into self-employment is associated with patience, as the response correlated with selection into self-employment involves making the choice to wait for a payout rather than taking a smaller payout today.

While this preliminary evidence on risk and time preferences comes with the clear caveat that the risk and time preference questions may vary over time as a function of agent outcomes, it is still suggestive. As expected, we find that selection into self-employment seems to be positively correlated with a willingness to take risk, and patience.

3.7.4 Cognitive ability

It seems intuitive that raw cognitive ability for matter for enterprise activity. Individuals with greater raw cognitive ability might be able to process information more quickly and solve challenging real-world puzzles. On the other hand, it may be that if "push" self-employment is important, that lower cognitive ability might be correlated with an inability to obtain steady wage employment.

The IFLS contains a number of measures of raw cognitive ability, which fall in two categories of about 10-15 questions per subject. The first category of questions is based on matching shapes. The subject is presented with an image with a piece missing, and invited to propose a fill-in piece from a menu of 3-4 options. The second set of questions are basic math questions, involving basic

operations like additional, subtraction, multiplication and division. The measures of cognitive ability show significant variation across the sample, which a very small proportion of subjects truncated at 0 or 1.

Initial results on raw cognitive ability and self-employment indicate that selection into entrepreneurship is quite negatively correlated with raw cognitive ability (of the household head). The marginal effect of -0.2 (on a variable on a 0-1 scale) is statistically significant at the 10% level, with a z-value of 0.087 (with a number of other demographic and locational controls included). This initial result suggests that the negative selection story may be the more prevalent one for enterprise activity. It may be that individuals with high cognitive ability have greater opportunities to pursue higher-return wage earning opportunities, and hence shun self-employment on average.

3.8 Conclusion

Recent literature on the relationship between financing and small-scale enterprise activity at the household level in developing countries raises some important questions. In contradiction to the standard model of credit-constrained occupational choices, we find that financial transfers do little to spur significant microenterprise activity. We particularly see that such transfers fail to spur the growth of microenterprises that might increase employment demand. In this paper I delve into the heterogeneity in financial effects in greater detail, carrying out a number of tests of financial constraints to small-scale entrepreneurship at the household level.

First, making use of detailed information on households' durable and busi-

ness assets I study the relationship between (lagged) wealth and measures of enterprise activity, such as startup and investment. The results from this exercise indicate that there is little relationship between (lagged) household wealth and enterprise activity at the lower 60 percentiles of the wealth distribution. This sharply contrasts the predictions of standard models, which predict that less wealth households should be most severely subject to financial constraints.

Second, I carry out liquidity-based tests of financing constraints using information on exogenous (but potentially anticipated) financial flows. I find that most flows are unrelated to enterprise choice, though receipts of unconditional cash transfers are important predictors of later enterprise activity. Third, I exploit plausibly exogenous financial shocks, such as lottery winnings and insurance payouts, to further test for the presence of financial constraints. I find a small but significant, positive net effect of such transfers on later enterprise activity in most specifications. I run a further specification for these transfers that employs a Propensity Score Matching approach to account for potential unobservable heterogeneity in the propensity to get access to such transfers, which implies a much more moderate effect of financing constraints.

Finally, I interact transfers with wealth levels to uncover heterogeneity in the role of financing constraints across the wealth distribution. In a number of specifications I find that responsiveness to financial transfers is increasing in the wealth distribution, in contradiction to standard models that predict that poorer households are most constrained by financing. I also find greater responsiveness amongst existing entrepreneurs, which is consistent with a model of entrepreneurial learning.

Taken together, these tests provide a nuanced picture on the role of financ-

ing as a constraint to household enterprise activity in the developing-country setting. Importantly, the results suggest that financing is not the main binding constraint to enterprise activity at the lower end of the wealth distribution. This in itself is an important policy implication in a world in which there is tremendous effort spent on getting financial resources to microentrepreneurs with the express goal of promoting growth-oriented enterprise activity. Of course, these results do not by any means show that microfinance or other such programs cannot be productive in supporting a range of other household activities. Yet they point to important caution in thinking of financing as the answer to entrepreneurial development among the poor.

Furthermore, the results indicate that wealthier households and existing enterprise owners are actually more constrained by financing. This suggests the hypotheses that such individuals may be endowed with greater entrepreneurial skills, which are largely overlooked in the current policy mix. I briefly look at the potential role of familial transmission of entrepreneurial skill in the final section of the paper. The results are also fitting with the model of [Ghatak *et al.*, 2007], which suggests that developing-country lending might be distorted by an overabundance of subsistence self-employed who might be better-served in switching into waged employment. These issues, also discussed in [Schoar, 2010] will require further study.

CHAPTER 4

ENTREPRENEURIAL HUMAN CAPITAL AND ENTREPRENEURIAL DYNAMICS: EVIDENCE FROM A NATURAL EXPERIMENT

4.1 Introduction

The question of what drives entrepreneurial dynamics, particularly entry and enterprise growth, is of long-standing interest to policymakers and academics in both the advanced and developing world. Such issues have taken on a new urgency in the developing world due to the pressures and opportunities of globalization and an increased recognition of vibrant private-sector enterprise activity as a source of economic growth and poverty reduction. In the academic literature there has been a recent, burgeoning interest in the factors behind entrepreneurial dynamics and firm performance. This has partly been motivated by the large interest, and subsequent disappointment, in microfinance as a stimulant of widespread growth-oriented entrepreneurship ([Banerjee *et al.*, 2009]; [Karlan & Zinman, 2010]), and also a growing interest in cross-country differences in firm productivity ([Bloom *et al.*, 2010]).

While much existing literature has focused particularly on firm-level financing access constraints as the key factor behind enterprise dynamics and outcomes, there has been growing awareness that this might not tell the full story. Attention has begun to shift to the human capital and managerial skill and knowledge of the individual entrepreneur and the firm (see, e.g., [Bloom *et al.*, 2010], [Bruhn *et al.*, 2010]), and the distinction between the much larger cohort of subsistence entrepreneurs and the significantly smaller cohort of higher-potential, transformational entrepreneurs ([Schoar, 2010]). While

existing theories largely take the entrepreneurship-specific human capital of the entrepreneur as a fixed parameter and focus on other factors behind enterprise dynamics such as learning and credit-savings interactions (e.g., [Jovanovic, 1982], [Buera, 2009]), in this chapter I derive and test predictions of a simple theory of entrepreneurship-specific human capital accumulation.

To test the theory and provide additional evidence on the returns to entrepreneurial experience, I exploit a uniquely-suitable natural experiment, the 1997-98 East Asian financial crisis. Key to identifying the stock of entrepreneurial human capital is the ability to disentangle it from other (generally unobservable) factors generating selection into entrepreneurial occupations, such as the role of a market opportunity, a "good idea," or unobserved ability or information. The crisis provides a plausibly unanticipated shock that generates increased entry into self-employment, which is orthogonal to a number of potential confounding factors. I focus on Indonesia, where the crisis had its most drastic effects among all countries. During the crisis period the returns to informal self-employment relative to private sector wage employment shift sharply in favor of self-employment for a significant cohort of individuals, providing an exogenous source of selection into self-employment. In addition, this effect is much more pervasive at the upper end of the earnings distribution ([Thomas *et al.*, 2000]), allowing us to focus on a higher-ability cohort of potential entrepreneurs.

We can think of this as a convenient natural laboratory in which to study the factors behind entrepreneurial dynamics. A cohort with higher potential to be transformational entrepreneurs, who might not normally consider running a business, are suddenly forced to do so. They generally start enterprises

in the informal non-farm sector, which was relatively unaffected by the financial crisis relative to the formal sector. Many have little prior self-employment experience. This focus on high-ability types is important in light of recent literature pointing to significant heterogeneity in the entrepreneurial potential of individuals (e.g., [Barrett *et al.*, 2005], [de Mel *et al.*, 2010], [Porta & Shleifer, 2008], [Schoar, 2010]). Such work suggests that the subset of individuals with the potential to run relatively larger, growth-oriented enterprises is a relatively small subset of the entrepreneurial talent distribution. Additionally, much of the existing developing-country literature has focused on the self-employment experiences of low-ability individuals, or draws inferences on higher-ability types from samples that are subject to non-random selection and recall biases. Finally, while recessionary and crisis events in more developed economies also have been shown to increase self-employment, such cases are less useful for the purposes of this chapter because the selection effect tends to be concentrated on lower-ability individuals, and is distorted by social safety nets such as unemployment insurance, severance packages, firing restrictions, and the like.

My primary dataset is the Indonesia Family Life Survey (IFLS), a panel survey which follows a nationally-representative sample of around 10,000 households and spans the crisis years. A particularly striking and puzzling finding in the data is that self-employment activity is remarkably persistent amongst those who enter self-employment during the crisis. This is true even in the years after the effects of the crisis have dissipated and the cohort of crisis entrants might have been expected to return to waged employment. About 78% of individuals who enter self-employment during the crisis are still running a business as their primary occupation 9 years after the crisis, whereas the usual 9-year persistence rate is around 46%.

I conduct more detailed empirical analysis in order to show that this persistence is robust. To do so I estimate counterfactual models of self-employment participation, to compare predicted self-employment participation to realized outcomes. I show that even after controlling for other covariates, in particular individuals' expected wage (as a measure of opportunity cost), individuals are estimated to be three to ten times more likely to engage in self-employment than under the counterfactual. I interpret this as evidence that after the initial, unexpected entry into self-employment, the individuals' stock of entrepreneurship-specific human capital has increased to the point of altering long-run occupational choice incentives. This explanation is also consistent with the jump in persistence (from 46% to 78%) amongst crisis-period entrants. If we make the natural assumption that the learning curve is concave and hence relatively steep early in one's entrepreneurial career, then we would expect the largest jump in persistence amongst a cohort of individuals with little prior experience who are exogenously pushed into self-employment.

Since alternative theories also predict persistence, however, I also look at unique predictions of the theory of entrepreneurial human capital accumulation regarding earnings. Competing theories that take the stock of entrepreneurial skill as fixed generate sub-optimal initial physical capital allocations due to uncertainty over own-ability ([Jovanovic, 1982]) or credit-market constraints ([Buera, 2009]). Under such theories firm growth (and persistence) is input-driven, as the stock of labor and capital in the firm grow to match the ability endowment of the entrepreneur. However, such theories require certain patterns in the co-movement of inputs and returns to hold (under a reasonable specification of the production function). I show that the trajectory of returns is best explained by the human capital-acquisition theory, as it exceeds potential

benchmarks including the dynamic expansion of inputs.

Finally, having provided evidence in support of the theory of entrepreneurial human capital accumulation, I proceed to directly estimate the causal effect of entrepreneurial experience on earnings. I use self-employment experience as a proxy for entrepreneurship-specific human capital acquisition through learning-by-doing, since underlying entrepreneurial ability is not directly observed. While a substantial literature in labor economics that similarly proxies human capital accumulation through years of exposure generally points to a positive effect of human capital on entry and earnings,[1] rigorous empirical evidence focused on human capital specific to entrepreneurship is much less common. In this case the separation between formal-sector labor market churning and outcomes in the informal sector is used as the key exclusion restriction in an instrumental variables setup. The evidence is supportive of the quantitative importance of the role of human capital acquisition, as experience is shown to strongly increase earnings, with the best estimate suggesting on average a 3% boost in net profit for each additional year of experience.

The chapter makes a number of contributions to the literature. It provides new evidence and an explanation for the surprising persistence in entrepreneurship that we see amongst higher-ability individuals after the Indonesian financial crisis. It suggests that a theory of transformational entrepreneurship should reflect the role of learning-by-doing in driving entrepreneurial dynamics. It also provides evidence on the value of entrepreneurial learning-by-doing for the unique cohort of crisis-period entrants, providing some of the first estimates on the value of such human capital in the literature.

[1] See [Card, 1999] for a review of studies taking years of education as a proxy for generalized human capital accumulation, and [Angrist, 1990] and [Behrman & Rosenzweig, 1999] for studies of human capital accumulation through work experience.

The chapter proceeds as follows. I begin with a qualitative description of the setting, with a particular focus on the informal sector in Indonesia and the effects of the crisis, along with a discussion of entrepreneurial human capital, in Section 4.2. I then outline a simple, dynamic model of entrepreneurial selection, savings and consumption in Section 4.3, and derive testable predictions that can be used to distinguish it from competing theories. I provide descriptive evidence on the dataset and outline the identification strategies in Section 4.4. The empirical analysis has two foci: evidence in favor the theory of entrepreneurial human capital accumulation, and causal evidence on the effects of entrepreneurial experience on earnings. The results are presented in Section 4.6, while various threats to analysis and robustness checks are presented in Section 4.7. Section 4.8 concludes, while additional content appears in C.

4.2 The Setting, and Entrepreneurial Human Capital

4.2.1 Indonesia Background

Indonesia is the world's fourth-largest country by population, and the largest Muslim democracy, though civil society is relatively secularized. In 1970 it was one of the world's poorest countries by any measure. However, it enjoyed average economic growth of 4.5% per year between the mid-sixties until the 1997-98 Asian Financial Crisis, and was on the verge of joining the middle income countries. In 1998 GDP dropped by 14% at the height of the crisis. After the end of President Suharto's reign during the crisis, the country began a political transition, which has involved full, democratic elections, regulatory reform, and

decentralization of power. By 2000 GDP growth recovered to 5% and was fairly steady around 5-6% until 2008.

Geographically, the country is spread out over thousands of islands in the world's largest archipelago. The country is highly diverse ethnically, religiously, linguistically and economically, yet unified by a common major language and national institutional structure. The island of Java, which contains the capital city of Jakarta, is the central economic hub. Per capita gross domestic product in purchasing power parity terms was $4000 USD in 2009, putting Indonesia at 155 in the world in this category (for comparison, the equivalent value for the United States is $46,400), while the value of the Gini coefficient was 39.4 in 2005, which is about average in international comparison ([CIA, 2011]).

Indonesian Labor Market Status and Trends

In most less-developed economies more than half of the workforce is engaged in operating or working in microenterprises,[2] which generate roughly half of GDP. A negligible proportion of such enterprises manage to grow beyond subsistence scale. In more developed economies the contribution of microenterprises to employment and GDP is closer to 15%, while an active and large small and medium enterprise sector that is absent in most developing countries contributes close to half of GDP. In Indonesia more than half of the workforce has typically been involved in working in or running micro, small or medium enterprises. The vast majority of such enterprises are informal sector firms with less than 10 employees.

[2]Indonesia's official enterprise size cohorts are defined as follows: microenterprise (1-4 workers), small enterprise (5-19 workers), medium enterprise (20-99 workers) large enterprise (100+ workers).

The formal sector was expanding in Indonesia in the years leading up to the crisis. From the mid-1980s until the late 1990s agricultural employment declined from 55 to 41% of the workforce, while manufacturing employment increased from 8 to 19% over the same period. Workforce participation rates of women were also increasing in the years leading up the crisis, from about 30 to 37% in the urban sector, though there was a steadier level of participation in rural areas of around 55%. This increase in urban employment was enjoyed in both the wage and self-employment sectors. We also notice that women are much more likely to work as workers in household enterprises in rural areas, at a 20-30% rate. Overall we see that labor force participation is relatively stable leading up the crisis, with a small uptrend toward formal sector activity.

The 1997-98 Crisis

The study of the labor market and self-employment effects of the crisis is facilitated by the availability of two excellent micro-datasets, which is unusual for a developing country. SAKERNAS is a labor-force survey that is collected by the Indonesian government statistical service, BPS, and is a large-scale, cross-sectional labor force survey. The Indonesian Family Life Survey (IFLS), is a panel dataset that was collected over multiple years, including the years spanning the crisis. For the study of the effects of a large-scale economic crisis, the dataset is particularly exciting because it included rounds just before the crisis hit, in 1997, and then a one-year-later follow-up. The IFLS is the primary dataset that will be used in the subsequent analysis in this chapter.

[Smith *et al.*, 2002] and [Thomas *et al.*, 2000] provide evidence on the labor market effects of the crisis. It is broadly recognized that Indonesia was the coun-

try worst-hit by the crisis, and that it was an unexpected event. The primary direct victim of the crisis was the banking and financial-services sector, much of which was left out or reorganized. The banking sector fell into disarray, and this led to a shortening of credit. While some of the early post-crisis research suggested that the crisis caused massive unemployment, in fact this claim does not hold up in the micro-data. What we see instead is significant churning in occupational allocations, with one important movement being from private sector waged employment to self-employment. The government sector seems to be relatively well-sheltered from the effects of the crisis.

Consumer prices began to spiral upward in 1998, at the rate of 80% in that year according to CPI. Hence a number of price subsidies were removed, such as on rice, oil and some fuels. All of this uncertainty and economic pain led to the fall of President Suharto in May 1998, with multi-party elections and the return to relative stability in 1999. The shock to relative prices that the crisis brought about did have some beneficiaries – exporters, export producers and the like. Those producing services and non-tradeables likely did less well, though on the other hand the informal sector was also better-sheltered from the crisis, by being more independent from formal sector financial institutions ex ante.

The labor market and consumer effects were a derivative of the impacts on firms and the price rises. On average real wages collapsed by 40% between August 1997 and August 1998, and these effects reached most sectors of the economy. However, informal sector effects were less pervasive, particularly amongst rural, self-employed males. Of greater interest to the current study is the resulting *relative* price changes, as reported in [Thomas *et al.*, 2000]. In particular, there is strong evidence that the main relative price shock during the crisis was in ex-

panding the relative margin between waged employment and self-employment. Self-employment broadly became relatively more attractive in comparison to private, waged employment, on the order of a 25-60% shift in relative returns depending on sector, gender and urban or rural location. In addition, this effect seems to be more strongly concentrated at the upper-end of the wage distribution ([Smith *et al.*, 2002]; [Thomas *et al.*, 2000]), and we see the most significant occupational churning from private wage to self-employment at the upper end of the wage distribution.[3] This suggests that it might be relatively high-ability individuals who were induced to enter self-employment during the crisis.

Hence it appears that the crisis can be interpreted as a large, unexpected shock to the choice margin between private wage employment and self-employment, which hits the most able formal-sector workers the hardest. This was due both to a significant hit to private wage returns, along with the observation that informal-sector, self-employment activity was generally more sheltered from the crisis. This exogenous and unanticipated shift in the choice margin appears to have induced sectoral restructuring toward self-employment activity.

4.2.2 Entrepreneurial Human Capital

Entrepreneurial human capital (EHC) constitutes specialized, high-level entrepreneurship-specific skills and knowledge, such as in selling, negotiating, product development, risk judgment ([Shane, 2003]) and entrepreneurial social capital. Above and beyond heterogeneous *ex ante* endowments of innate EHC, perhaps due to genetic inheritance or early upbringing (i.e., dynastic transi-

[3] [Poppele *et al.*, 1999] argue that the main effect of the crisis was on the urban elites.

tions), I hypothesize that EHC is significantly accumulated through *direct* exposure to entrepreneurial activity. Such entrepreneurial capabilities are distinct from other, generalized forms of human capital such as education, life experience, and experience in waged employment. A primary channel for acquiring EHC is learning-by-doing (i.e., running an enterprise, the focus of this chapter).[4] EHC cannot be transacted in the marketplace separately from the individual endowed with it and public institutions for the transmission of EHC (such as the formal education system for wage work) are generally absent. Hence dynamic occupational selection incentives can play a crucial role in individuals' ability to accumulate EHC.

4.3 A Simple, Dynamic Model of Entrepreneurial Selection, Savings, and Consumption

I outline a simple, forward-looking model of individual occupational choice that captures the theoretical mechanism that I will test for in the data. The key feature of the theory is that it allows for entrepreneurial human capital accumulation through direct learning-by-doing. This is meant to capture the accumulation of *entrepreneurship-specific* human capital and business capital, such as in product development, marketing, risk judgment and business-relevant social network connections, through first-hand exposure to entrepreneurial activity. Much of the existing literature takes entrepreneurial human capital as fixed,[5]

[4]Other channels for EHC transmission that one might consider include transmission of EHC in the family (e.g., learning from one's parents, if they are entrepreneurs), or learning through work experience in another firm.

[5]The seminal, early reference on job- and occupation-specific human capital is [Becker, 1964]. Surprisingly little work has been done to formally extend such ideas to entrepreneurship, though less formal work exists in the economics literature in the work of [Schultz, 1980]; see

and then studies dynamics emanating from the physical capital accumulation (savings) choices of the entrepreneur,[6] or learning about the value of the unknown entrepreneurial ability endowment in a Bayesian learning framework.[7] In the empirical analysis I will argue that the learning-by-doing framework best matches the patterns in the data. Hence I begin by providing a simple formalization of the learning-by-doing framework, then derive testable predictions of the modeling frameworks that allow me to disentangle them in the data.

The agent is initially endowed with a stock of liquid wealth, W^0. While the agent can be thought of as capturing a household unit, the exposition will describe the model for a single individual. In each of two periods, $t = 0, 1$, the individual makes a choice between one of two occupations – waged employment, denoted by w, or self-employment, denoted by s. Human capital specific to each occupation is modeled by stock variables for each occupation.

Wage earnings, $y(\theta_w^t)$, are a function of the agent's stock of wage-specific human capital, θ_w^t, which can accumulate due to work experience. Let $\Phi_w\left(\theta_w^0, \cdot\right)$ denote the transmission of wage-specific human capital between the two periods, where the second argument of Φ_w records the agent's occupational choice in the first period. Then θ_w^0 denotes the initial endowment of wage-specific human capital, and $\theta_w^1 = \Phi_w\left(\theta_w^0, \cdot\right)$ denotes the stock of human capital in the second period. I assume that work experience has value, that is, that $\Phi_w\left(\theta_w^0, w\right) = \theta_w^1 > \theta_w^0$.

also [Klein & Cook, 2006]. Two exceptions, though less general in scope, are [Otani, 1996] and [Iyigun & Owen, 1998].

[6] The literature focusing on occupational choice and the dynamic savings problem has primarily been motivated by an attempt to rationalize otherwise suprisingly strong inequalities in the aggregate wealth distribution. In such models individuals with (unobserved) high entrepreneurial skill have an incentive to save much more than others, which can generate significant wealth inequalities in a dynamic setup. See, e.g., [Cagetti & Nardi, 2006] and [Buera, 2009].

[7] The early, seminal paper in this line is [Jovanovic, 1982]. [Taveras, 2010] carries out a calibration exercise on a similar model to show that a number of stylized facts that have been taken as evidence of credit constraints in prior literature can in fact be rationalized in a model of Bayesian learning about entrepreneurial skill if learning is sufficiently slow.

For simplicity, $\theta_w^0 = \Phi_w\left(\theta_w^0, s\right)$ (i.e., if the agent does not acquire wage work experience then the stock of wage-specific human capital does not change). Note that this rules out the possibility that human capital relevant to wage employment can be acquired in self-employment. While it would be interesting to consider an extension that weakens this assumption, it would raise additional empirical challenges to identify cross-occupational-relevant human capital accumulation. Finally, I assume that Φ_w is increasing in its first argument.

The setup for self-employment is analogous. Self-employment earnings are influenced by the agent's stock of entrepreneurial human capital, θ_s^t. $\Phi_s\left(\theta_s^0, \cdot\right)$ denotes the transmission function, where $\theta_s^1 = \Phi_s\left(\theta_s^0, \cdot\right)$. Analogously, I assume that $\Phi_s\left(\theta_s^0, s\right) = \theta_s^1 > \theta_s^0$, that $\theta_s^0 = \Phi_s\left(\theta_s^0, w\right)$, and Φ_s is bounded.[8] The profit function is as follows, for $t = 0, 1$,

$$\begin{gathered}\pi\left(\theta_s^t, W^t, p^t, p_k^t, p_l^t\right) = \max_{l \geq 0} p^t f\left(\theta_s^t, k, l\right) - p_k^t\left(W - k\right) - p_l^t l \\ s.t.\ 0 \leq p_k^t k \leq \lambda W^t\end{gathered} \tag{4.1}$$

where p^t is the price of a single output in period t, k is capital, l is labor, p_k^t and p_l^t are their respective prices in period t, and f is an increasing, concave production function. I assume that the the firm is a price-taker. The constraint set $k \in [0, \lambda W^t]$ is standard in the literature and captures credit constraints – the stock of physical capital employed in the enterprise may be constrained by own-funding constraints if there are frictions in credit markets and other financing sources are not available. That is, it may be that the optimal stock of capital, k^*, is strictly greater than λW^t, so that the firm is constrained from employing the optimal capital stock.

[8] Perhaps most important here is the second part of the assumption, which implies that entrepreneurial skills aren't acquired in wage employment.

The timing of the model is as follows. The agent first draws human capital endowments θ_w^0 and θ_s^0 from the joint distribution η, with support on $\mathbb{R}^2$. This allows for arbitrary correlation between the two, which can be interpreted as capturing greater general ability in the individual, and can exacerbate selection effects as I will demonstrate below. These occupation-specific human capital stocks are known at the beginning of each of the two decision periods. Given these known human capital stocks, the agent makes a discrete occupational choice, between wage-employment and self-employment, w or s, in a forward-looking way in the first period. If the agent chooses self-employment, she makes a decision about the labor and capital inputs to the enterprise, k and l. The intertemporal connection between the two periods is given by the human capital transmission functions Φ_w and Φ_s as described above, along with the intertemporal savings problem. Denote the savings choice by x, where it must be that the value of x is less than the sum of wealth the agent opens the first period with, W^0, and earnings (y or π). The residual of the savings choice is consumption, which is evaluated in the strictly increasing, concave utility function U. At the end of the second period the agent is taken to consume all remaining wealth.

Formally, then the agent faces the following decision problem in the initial

period, which is summarized by the value function V_0,

$$
\begin{aligned}
& V_0\left(\theta_w^0, \theta_s^0, W^0\right) && (4.2)\\
= \; & \max\left\{ \max_{0\le x\le y+W^0} U\left[y\left(\theta_w^0\right)+W^0-x\right]+\beta\max\left\{U\left[y\left(\theta_w^1\right)+W^1\right], \right.\right. && (4.3)\\
& U\left[\pi\left(\theta_s^0, W^1, p^1, p_k^1, p_l^1\right)+W^1\right]\Big\}, \\
& \max_{0\le x\le \pi+W^0} U\left[\pi\left(\theta_s^0, W^0, p^0, p_k^0, p_l^0\right)+W^0-x\right]+\beta\max\left\{U\left[y\left(\theta_w^0\right)+W^1\right], \\
& \left.\left. U\left[\pi\left(\theta_s^1, W^1, p^1, p_k^1, p_l^1\right)+W^1\right]\right\}\right\} \\
= \; & \max\left\{ \max_{0\le x\le y+W^0} U\left[y\left(\theta_w^0\right)+W^0-x\right]+\beta\max\left\{U\left[y\left(\Phi_w\left(\theta_w^0, w\right)\right)+W^0+x\right], \right.\right. && (4.4)\\
& U\left[\pi\left(\theta_s^0, W^0+x, p^1, p_k^1, p_l^1\right)+W^0+x\right]\Big\}, \\
& \max_{0\le x\le \pi+W^0} U\left[\pi\left(\theta_s^0, W^0, p^0, p_k^0, p_l^0\right)+W^0-x\right]+\beta\max\left\{U\left[y\left(\theta_w^0\right)+W^0+x\right], \\
& \left.\left. U\left[\pi\left(\Phi_s\left(\theta_s^0, s\right), W^0+x, p^1, p_k^1, p_l^1\right)+W^0+x\right]\right\}\right\}
\end{aligned}
$$

where the second equality illustrates the functional relationships that generate the final-period values of the stock variables of occupational skill and wealth, and β is a discount factor in the $(0, 1)$ interval. Namely, in the initial period the agent faces a discrete choice over the immediate occupational return given by y or π, and the discounted future return obtained from the same activity choice in the second period.

4.3.1 Basic Properties of the Model

The value function in equation (4.2) formalizes the dynamic incentives in the occupational choice problem. First, the individual faces an initial "selection" incentive, influenced both by the initial returns generated by the values of θ_s^0 and θ_w^0 (and possibly the effect of binding credit constraints on the physical cap-

ital decision), and prospective second-period returns due to savings and human capital accumulation. θ_s^0 and θ_w^0 might be accumulated prior to formally entering the workforce through familial effects, education, or other life experiences. All things equal we expect that individuals with a relatively higher stock of ability in a given occupation to be more likely to self-select into that occupation. Formally, the condition for selection into self-employment is as follows,

$$\begin{aligned} & \max_{0 \leq x \leq \pi + W^0} U\left[\pi\left(\theta_s^0, W^0, p^0, p_k^0, p_l^0\right) + W^0 - x\right] + \beta V_1^s \\ \geq & \max_{0 \leq x \leq y + W^0} U\left[y\left(\theta_w^0\right) + W^0 - x\right] + \beta V_1^w, \end{aligned} \tag{4.5}$$

where to save on notation V_1^s and V_1^w denote the future utility derived from choosing the optimal occupation in the second period, given human capital accumulated by the choices of s and w, respectively, in the initial period. The complementary condition captures the incentive for selection into waged employment. Equation (4.5) can be used to characterize the subsets of the parameter space under which selection into each occupation is optimal.

Self-selecting into a given occupation can lead to the acquisition of relevant human capital that further shifts the choice margin between the two occupations. That is, human capital accumulation can lead to lock-in, in a given occupation. If, for example, the individual chooses self-employment in the first period, this increases the value of θ_s^1, which increases the value of second-period profit $\pi\left(\theta_s^1, W^1, p^1, p_k^1, p_l^1\right)$ and hence increases the propensity to select into self-employment in the second period. In fact, dynamic incentives might even generate dynamic selection effects, under which individuals are incentivized to enter self-employment today even for a lower static return, under the anticipation of greater returns in the future ([Foster & Rosenzweig, 1995] capture similar in-

tuition).

The human capital lock-in effect highlights the importance of the initial occupational choice, which is driven by the initial stocks of human capital, θ_s^0 and θ_w^0. If there is a reasonably high degree of correlation between the initial stocks of human capital, then it may be that the "highest-potential" entrepreneurs do not enter self-employment at all, because the opportunity cost to self-employment is high based on wage earnings possibilities. This insight, first discussed semi-formally in the economics literature in [Roy, 1951], points out that occupational selection will be driven by the distribution of skills in the population and how they are compensated in equilibrium. It could be that most high-ability individuals tend to enter waged employment, and human capital lock-in further reinforces that choice. On the other hand, low-skill individuals might receive relatively lower returns in wage employment, particularly if low-skill labor supply is abundant. This is consistent with the massive cohort of low-skill, self-employed individuals in developing countries, most of whose enterprises have low returns and grow little. In Chapter 5 I present a model of dynamic entrepreneurial human capital accumulation in which low-skill individuals can get trapped in a subsistence microentrepreneurship "poverty trap," not due to lack of financing but rather inadequate human capital.[9]

[9]This charactization is consistent with recent empirical evidence (e.g., [Carter & Olinto, 2003]; [de Mel *et al.*, 2008]; [Banerjee *et al.*, 2009]; [Karlan & Zinman, 2010]). Demand for capital ends up being relatively stronger amongst wealthier or higher-ability individuals and hence individuals end up more responsive to positive financial shocks.

4.3.2 The Effects of Exogenous Shocks to Occupational Choice Incentives

The potential for human capital lock-in to prevent the highest-potential entrepreneurs from entering self-employment is suggestive of the empirical strategy that will be employed in this chapter. I will seek a source of exogenous variation in selection incentives, orthogonal to individual ability, that leads relatively high-ability individuals to select into self-employment when they otherwise would not have done so. Such a shock can be due to any of the exogenous parameters of the model – to a price, to wealth, or to the earnings functions. The value function in equation (4.2) clarifies the effect of such changes, which are almost always unambiguous in the model. In this chapter I will focus on an exogenous, negative shock to the wage employment earnings function, $y\left(\theta_w^0\right)$, though it is not problematic if the effects of the shock are transmitted through additional parameters of the model. Due to an exogenous event, which we can think of as occurring prior to period 0, the margin of choice will shift for a number of individuals, and they will have a much greater incentive to select into self-employment, as it becomes more likely that equation (**??**) will see a tilt in incentives toward self-employment.

4.3.3 Testable Predictions and Alternative Theories

A direct prediction of the theory is that entrepreneurial experience should lead to entrepreneurial persistence, even after accounting for opportunity costs. This is the human capital lock-in effect that was discussed above. As θ_s^0 increases to θ_s^1, the individual should be more likely to again engage in self-employment

in period 1. This is true even if a negative wage shock caused an increase in self-employment, and then the wage returns to its previous level. The empirical version of this prediction will be developed in Section 4.5. However, while the finding that individuals who obtain self-employment experience are more likely to remain self-employed is consistent with and strongly suggestive of a theory of entrepreneurial human capital accumulation, such a finding is still not a conclusive basis to argue that entrepreneurship-specific human capital accumulation is the primary factor driving enterprise dynamics. There are at least two alternative theories that generate a similar prediction, which in contrast take the stock of entrepreneurial skill as *fixed* and generate dynamic effects through other channels.

In [Jovanovic, 1982], individuals are endowed with a fixed stock of entrepreneurial skill, which they are uncertain about and have prior beliefs over. In the context of the model developed herein, we can think of this as an entrepreneurial skill parameter θ_s that doesn't vary over time, but determines the distribution of stochastic realizations of the production function.[10] The individual holds subjective beliefs μ_{θ_s} over the distribution of θ_s, which is initially drawn from a normal distribution with known mean and variance. Since the individual does not know the exact value of her own θ_s, the initial belief is taken as the mean of the distribution μ_{θ_s}. Over time, as the firm operates, the agent draws observations on a stochastic production process, which allow for inferences on θ_s, with updating of beliefs through a standard Bayesian learning process. Hence there is a co-movement of beliefs and firm size – in expectation good entrepreneurs grow their firms as their beliefs about own-ability move upwards, while bad entrepreneurs shrink and eventually exit.

[10] In fact, in [Jovanovic, 1982] θ_s is a parameter that determines the distribution of shocks to the cost function.

In [Buera, 2009] and related models, skill is known but credit market constraints might prevent the optimal physical capital level from being attained, at least in the short run. Individuals know their entrepreneurial skill level, and indeed that knowledge may exactly induce them to save more ex ante in order to eventually be able to self-fund the startup or growth of an enterprise. In the context of the model herein, we can think of this as a case where the function $\Phi_s(\cdot,\cdot)$ is a constant function. For the model to be empirically relevant it is important that the credit constraint actually binds for a significant proportion of the population. Similarly to [Jovanovic, 1982], the theory predicts that physical capital increases over time for good entrepreneurs, as it converges to the level most compatible with the endowment of skill.

Hence both models allow for the possibility that a significant number of individuals who enter self-employment will be persistent and see an increase in inputs and earnings over time. They suggest that a significant number of enterprises will enter the market at a different scale from their long-run optimal scale, and that successful firms will converge to the long-run optimal size as dictated by the fixed stock of entrepreneurial skill. Of course, one can quickly see that the theory of entrepreneurial human capital accumulation will also predict increases in capital and labor inputs over time, to optimally complement the stock of entrepreneurial skill. However, what is critical is that the alternative theories suggest that increases in returns should be *input-driven,* in terms of labor and capital inputs. The theories do not allow for residual increases in profitability due to increases in the entrepreneurial and managerial abilities and business capital of the individual running the firm.

An additional test would then be to study the relationship between earn-

ings increases of the firm and increases in the size of the capital and labor stock of the firm. The testable prediction disentangling a model with dynamically-accumulating entrepreneurial human capital from the other theories would be the finding that the variation in earnings could not be explained by dynamic changes in inputs alone. Technically this requires some assumptions about the revenue function of firms, in order to discipline the relationship between inputs and revenues.

A final notable implication regards the *rate* of purported entrepreneurial human capital accumulation amongst entrants. It seems reasonable to posit that the learning function takes a concave shape, with diminishing returns to learning as more human capital is accumulated. This would mean that, all things equal, brand new entrants (those who had not previously run an enterprise) should learn at the highest rate, and hence be subject to the largest change in earnings and occupational choice incentives. We will also look for support for this final testable implication.

4.4 Design of the Study and Preliminary Evidence

4.4.1 Data

My primary dataset is the Indonesia Family Life Survey (IFLS).[11] The data was collected as a household panel survey in Indonesia, with data collection rounds in 1993, 1997-98, 2000-01 and 2007-08. The 1997-98 round directly proceeded the

[11] Various organizations and researchers have been involved in designing, collecting and funding the IFLS. For more details, see [Strauss *et al.*, 2009], [Strauss *et al.*, 2004], [Frankenberg & Thomas, 2000], and [Frankenberg & Karoly, 1995].

crisis. For the intervening years when the survey is not fielded, significant retrospective data are collected in the subsequent round. The dataset was designed to be representative of 83% of the Indonesian population in 1993, covering 13 of the higher-population provinces generally in the western parts of the country, with over-sampling of urban locations and locations outside Java island, the main economic hub. Data were collected at the individual, household, and community level, and these three sources can be matched together. More details on relevant parts of the dataset, including for enterprise activity, will be discussed in more detail below.

The original 1993 round of the survey (IFLS1) surveyed 7224 households. Subsequent rounds have involved re-sampling the original households, and then sampling all split-offs from the original households. Attrition has been relatively minor, at less than 10% between rounds, and overall 87.6% of the original households appear in all four rounds. Table A.1 presents the number of individuals,[12] households, household enterprises and communities appearing in each round of the survey. We see that the sample expands in each subsequent round, as splits from the original households are tracked and surveyed. In addition, the proportion of household members directly interviewed also increases across rounds.

There is significant geographic and size variation amongst the enterprises.[13] Though the largest firm representations are from Java, the economic and population center of the country, the bias is not overwhelming and a significant proportion of firms are observed from all of the main survey provinces. This is true

[12]Both adults and children (defined as those under age 15 at the time of the survey) are surveyed, though the childrens' module is less extensive.

[13]The distribution of enterprises is less even if we stratify by industry–the largest proportions of enterprises by far are in the sectors of restaurant/food, and sales:non-food, at around 30% each. The next two largest sectors are food processing, and services:transport.

even if we focus on firms with a relatively larger capital stock, above $1000 US (converted from Indonesian rupiah at the going exchange rate in a given survey year). It is notable that the slightly larger proportion of firms seems to be in rural areas. This fits with [Liedholm & Mead, 1999] and may be due to the fact that smaller firms are more likely to service demand in more remote areas. Also, we see that the sample contains a significant number of firms exceeding the sizes observed in the vast majority of studies on micro and small enterprises from developing countries, while firm-level surveys looking at such firms generally have little information on the primary entrepreneur. Given that conversion to US purchasing power parity implies a multiple of about 12, there are hundreds of enterprises with more than $25,000 US PPP equivalent in capital, and dozens with 10, 15 or more workers.

Table C.2 presents a summary of a number of community-level measures of market churning that will be useful in the background of the later analysis, as these variables are used as exogenous sources of variation in the individual propensity to enter and remain in self-employment.

4.4.2 Preliminary Evidence

In Chapter 5 I provide evidence on earning-experience profiles in the data, distinguishing across enterprise type cohorts. In particular, the cohorts of individuals running enterprises with no employees, only household/unpaid employees, and those actually hiring outside, paid employees can be distinguished. The empirical analysis employs panel data techniques that allow the fixed effect term to be interpreted as controlling for time-invariant ability. Those results sug-

gest a year-on-year value of entrepreneurial experience of 8-15% on average, depending on cohort. In Figure B.1 I non-parametrically plot experience-earnings (net profit) profiles across these three qualitative categories, using a Lowess tricube smoother. There we see that while all three groups enjoy an increase in earnings on average, the rate of increase is substantially higher for those running the enterprise we would expect to be most complex: firms with hired, wage workers. This bifurcation in returns is suggestive of the select group of individuals running more complex enterprises "pulling away" from the much larger group of individuals running enterprises in the other two categories. We would expect that significantly greater returns would enable significantly greater capital accumulation.

In Table C.1 I present summary statistics on the smaller population of individuals who enter self-employment during the financial crisis, a smaller sample. There are 684 such individuals who are eligible for the study due to entry during 1997, and 1355 eligible due to entry in 1998. In comparison to full-population summary statistics presented elsewhere in this thesis, we see that they are highly likely to be married, often quite well-educated, and more likely to be male. They also appear relatively younger, which could be a reflection of the role in seniority in worker separations during the crisis.

4.5 Identification Strategies and Empirical Specifications

The identification of EHC raises empirical challenges due to the selection processes highlighted in the model. The ideal experiment would randomly assign EHC to individuals, orthogonally to all other characteristics, and then ob-

serve the resulting enterprise performance trajectories. Clearly such an experiment would be infeasible for a number of reasons, including endogenous enterprise survival, and difficulties in assigning EHC. However, individuals with greater *ex ante* (unobserved) entrepreneurial ability are more likely to select into self-employment, and hence accumulate greater entrepreneurial human capital. Hence higher-ability self-employed individuals are likely to have better entrepreneurial performance (i.e., higher enterprise returns), while simultaneously having greater accumulated experience, due to endogenous survival effects.

Since the ideal experiment is not feasible in practice, I exploit a source of exogenous and unanticipated assignment into self-employment (experience) due to the 1997-98 Financial Crisis. Here the primary "treatment" group of interest is the subset of individuals that enter self-employment during 1998, the main year in which the effects of the crisis were felt in Indonesia. In particular, the interest is in individuals who were 'pushed' into self-employment, who would not have otherwise entered, which provides a source of a counterfactual to consider the effects of the quasi-random assignment of EHC.

Previous analyses of the effects of the crisis have shown that the crisis did not cause a significant drop in overall employment; however, it caused a significant shift in real wages, in some cases up to 40%, with effects particularly concentrated on relatively higher-earning, formal-sector wage workers ([Thomas *et al.*, 2000]). This exogenous shock is particularly useful for the purposes of this study, because it means that a significant number of relatively higher-ability individuals were 'pushed' into self-employment. Hence this natural experiment is quite appealing to test the theory of EHC, because other

sources of exogenous shocks such as rainfall might be expected to be concentrated on lower-income individuals who might have fewer alternatives to self-employment.

Based on this intuition, I construct two main tests of the interpretation of EHC as a natural experiment that assigns individuals to acquire entrepreneurial experience. First, I look at self-employment persistence. The theoretical model predicts that individuals who acquire human capital in a particular occupation should, all things equal, be more likely to remain in that occupation. In testing this implication I deal with the assumption of "all things equal" potentially not holding by using a number of regression controls, including in particular an out-of-sample estimate of the counterfactual wage that self-employed individuals would earn in wage employment. Since dynamic occupational persistence can be explained by a number of theories outside of EHC accumulation, I secondly look at earnings dynamics, to generate further evidence consistent with the proposed theory.

Building on this evidence, I provide evidence on the causal effect of entrepreneurial learning-by-doing on earnings. The empirical challenge that arises is that, of course, the crisis is not a perfect natural experiment. It does not necessarily randomize selection into self-employment (and subsequent acquisition of experience) orthogonally to unobserved ability, in particular. Those cohorts that enter self-employment, even during the crisis period, presumably include at least two groups: (i) those who enter self-employment as a survival response to the shock (due to having lost their job, etc.), or the 'pushed' group of interest, and (ii) those who enter self-employment voluntarily (perhaps because the disequilibrium process highlights a new profit-making opportunity), or because

they were already planning to enter self-employment independently from the crisis. I attempt to control for endogenous selection into self-employment in 1998 through a selection-on-observables-type strategy, which is plausibly exogenous to individual-level EHC.

Details behind these approaches are discussed in the remainder of this section, and estimation results are then presented in Section 4.6.1.

Self-employment Persistence

An important implication of the theoretical model is that the accumulation of EHC changes the occupational choice incentives of the individual. If in period 1 the individual chooses self-employment, s, perhaps due to a shock to the opportunity cost to self-employment (the wage $y\left(\theta_w^0\right)$), then the human capital accumulation function, $\Phi_s\left(\theta_s^0, s\right)$, implies that the stock of EHC increases from θ_s^0 to $\theta_s^1 > \theta_s^0$. Even if the opportunity cost of self-employment returns to near its previous level, the agent is more likely to find it optimal to remain in self-employment in subsequent periods. This trade-off is formalized in the model, in particular where we see that the second-period decision involves the static maximization problem,

$$\max\left\{U\left[y\left(\theta_w^1\right)+W^1\right], U\left[\pi\left(\theta_s^1, W^1, p^1, p_k^1, p_l^1\right)+W^1\right]\right\}. \tag{4.6}$$

Of course the outcome is not deterministic – since the choice is discrete it may be that the choice margin moves but still not enough to induce the agent to remain in self-employment once the wage recovers. However, across the population distribution we might expect to observe an effect.

Just looking at raw numbers, we see that 684 individuals newly shift into self-employment in 1998, about a 10% increase in the number of self-employed individuals. By the year 2000, 587 of these individuals are still self-employed (about 85%), even though the economy has already shown significant recovery from the crisis event. Even by the year 2008, about 78% remain in self-employment. This comes in stark contrast to the comparable figure from other the 10 years of the survey from which a 9-year persistence rate can be calculated, which averages 46%. However, these raw indications are subject to some key challenges in terms of identification. I discuss how I deal with these challenges in what follows.

Firstly, the descriptive evidence on self-employment persistence does not control for the expected wage, the opportunity cost to self-employment. It could be the case that wages don't recover for the types of individuals who enter self-employment during the crisis, and hence in fact the opportunity cost to self-employment remains low. In addition, as has been already discussed, the self-employment entry decision can be driven by unobservables. In order to account for these concerns, I carry out the following estimation procedure to attempt to provide more convincing evidence for the robustness of occupational persistence.

The intuition behind the procedure is to construct the (unobserved) counterfactual probability of being self-employed in absence of having entered self-employment during the crisis, and then compare that to two constructs of the realized propensity to be self-employed: (1) the empirical realization of self-employment propensity (the simple frequentist estimate), and (2) an estimated probit model on ex post occupational choice outcomes in the sample of individ-

uals who enter self-employment during the crisis. Hence it involves a within comparison of predicted entrepreneurial propensity based on pre-crisis choices, with ex post realized outcomes. I will then argue that results of sufficient magnitude overcome other explanations for the self-employment persistence such as, e.g., labor market frictions preventing re-integration into the formal wage sector.

The procedure works as follows. First, I empirically capture the occupational choice decision rule of individuals by estimating a probit self-employment selection equation of the form,

$$\Pr\left(y_{it} = 1|\mathbf{x}_{it}\right) = G\left(\mathbf{x}_{it}\beta + \gamma\omega_{it}\right), \tag{4.7}$$

where y represents the discrete occupational choice ($y_{it} = 1$ denotes self-employment, and $y_{it} = 0$ denotes wage employment), G is the standard normal density, $\mathbf{x}_{it}$ is a vector of regression controls such as age and age-squared, education (in years) and marital status, ω_{it} represents the wage, and β and γ represent regression coefficients.

I estimate the above model using two definitions of the population. First, just on the sub-sample of individuals who enter self-employment during the crisis, and secondly on the whole population. The former more directly captures the choice function of the specific individuals involved, though it might underestimate entrepreneurial propensity since these individuals are less likely to be self-employed pre-crisis. The latter better captures the determinants of entrepreneurial selection in the population, though it might induce estimates that are less applicable to the particular crisis-entrant sample.

Second, having used this model to estimate the occupational choice function, I then construct projected occupational selection propensities, $\hat{G}_{it}$ out of equation (4.7), by predicting out of sample using the estimators $\hat{\beta}$ and $\hat{\gamma}$. Since the wage, ω_{it} in equation (4.7), is not observed once individuals have entered self-employment, I employ the following wage equation in order to calculate the individual-specific projected wage, $\hat{\omega}_{it}$, as a measure of the opportunity cost of self-employment,

$$\omega_{it} = \mathbf{x}_{it}\delta + c_i + year_t + \varepsilon_{it}, \tag{4.8}$$

where $\mathbf{x}_{it}$ is a vector of regression controls such as age and age-squared, education (in years) and marital status, c_i is an individual-specific fixed effect term, and $year_t$ is a year effect.[14] I use a bootstrap approach to deal with the issue of using projected regressors as explanatory variables in a subsequent regression.

The out-of-sample prediction of self-employment propensity, $\hat{G}_{it}$, gives a counterfactual measure of self-employment propensity. I generate out-of-sample predictions of $\hat{G}_{it}$ from two different data samples, which I denote $P^{sub}_{Pre\rightarrow Post}$ and $P^{full}_{Pre\rightarrow Post}$, respectively. I denote the mean of the distribution of values of individual-specific self-employment propensities based on ex ante data only from the *subsample of crisis-period entrants* by $P^{sub}_{Pre\rightarrow Post}$. I denote the same object, estimated on *full ex ante population* data, by $P^{full}_{Pre\rightarrow Post}$.

Third, I construct ex post measures of self-employment propensity, from actual realizations in the data. I denote by $P^{sub}_{Post\ freq}$ the empirical realization of self-employment propensity (the simple frequentist estimate), and by $P^{sub}_{Post\ prob}$ occupational choice propensity estimates generated from a probit model on ex

[14] I do not include time-variant, location-specific variables as controls, since geographic identifiers are not always available for each observation, meaning sample size would be noticeably reduced.

post occupational choice outcomes in the sample of individuals who enter self-employment during the crisis.

I then test whether there is a statistically significant difference in the propensity to be self-employed, comparing the constructed counterfactuals, $P^{sub}_{Pre \rightarrow Post}$ and $P^{full}_{Pre \rightarrow Post}$ to the ex post realizations, $P^{sub}_{Post\ freq}$ and $P^{sub}_{Post\ prob}$. I apply t-tests to the differences in the means of the two distributions.

The Dynamics of Self-Employment Returns

The analysis of persistence provides a convincing source of evidence on the propensity to remain self-employed. Yet, it does not rule out some alternative hypotheses outside of the endogenous accumulation of EHC. The main competing theories of entrepreneurial dynamics take entrepreneurial ability as fixed, and then generate dynamics from learning about own-ability [Jovanovic, 1982], saving, or the like. To disentangle the proposed theory of EHC accumulation from a Jovanovic-type story, I study enterprise earnings dynamics.

In Jovanovic's model, individuals persist in self-employment because they turn out to be the 'good' entrepreneurs, through getting earnings draws and learning about own ability. In such a model we should not see entrepreneurial returns increase greatly relative to the overall economy, because optimal entrepreneurial inputs are available immediately at enterprise startup. Hence I study the dynamics of enterprise earnings and how they increase relative to the growth of the overall economy and counterfactual wages.

Selection-corrected earnings dynamics The final piece of evidence on earnings dynamics comes from taking years spent running an enterprise as a measure of learning-by-doing and entrepreneurial human capital acquisition. This is analogous to the literature on education as a source of human capital. Similarly to that literature, the main empirical problem in deriving causal estimates of the effects of human capital acquisition is one of selection: individuals are not randomly assigned to acquire entrepreneurial experience. I take the crisis to provide quasi-experimental variation in the incentives to enter self-employment and hence acquire entrepreneurial experience.

I calculate years of experience running enterprises in three different size categories – no employees, only household/unpaid employees in the enterprise, or those which hire permanent wage workers for an explicit wage. I then use an adaptation of the Heckman selection procedure to study the selection-corrected relationship between the experience measures and self-employment earnings (net profit).

The traditional Heckman model involved running a first-stage selection equation, then using it to generate an individual-level estimate of the propensity to select into one of the selection options, which is then fed into the second-stage equation as the inverse Mills ratio. I follow this approach, inserting an estimate of the propensity to enter self-employment which has already been presented above in Section 4.5, in equation (4.7). I take the appropriate version of $\hat{G}_{it}$ to give me the individual-level occupational selection propensity, then use it as a control in an earnings experience regression. As already discussed above, the first-stage selection model incorporates variables reasonably excluded in the second-stage earnings equation – primarily location-level measures of occupa-

tional churning. Hence the requirement of at least one non-intersection between the first- and second-stage equations in a traditional Heckman setup is satisfied. This approach accounts for individual-level variation in the propensity to enter self-employment in a given period, based on observables.

The second-stage earnings equation is given as follows:

$$y_i = \beta_0 + \exp_i\beta + x_i\gamma + \hat{G}_i\delta + \varepsilon_i, \tag{4.9}$$

where y_{it} represents reported self-employment earnings by individual i, β_0 is a constant, $\exp_i$ is a vector of individual-specific entrepreneurial experience counts, x_i is a vector of other controls (age, age^2, gender, education (in years), marital status), and $\hat{G}_i$ is the projected occupational selection value.

Given that this procedure introduces a generated regressor in the second-stage earnings equation through $\hat{G}_i$, in the second stage estimates I use a bootstrap procedure with 50 replications, to account for a potential non-standard error distribution rather than imposing normality on the model.

4.6 Estimation Results: Self-employment Persistence and Returns

In this section I present the empirical results on self-employment persistence, and self-employment earnings dynamics, respectively.

4.6.1 Self-employment Persistence

The results from this part of the analysis are summarized in Tables C.3 to C.5.

I present the results of the fixed effects wage equation analysis in Table C.3. The main goal of this equation is explanatory power, and that seems to be achieved with an R^2 of 0.49. The age effect is notable in implying a convex function, though the first-order coefficient is negative but not statistically-significant. In general Mincer equations will generate a concave age effect. Otherwise we find that the remaining regressors are almost always highly statistically-significant with reasonable coefficients.

Looking at entrepreneurial persistence, I report on analysis looking separately at the group of individuals who enter self-employment during 1998 (which might be affected by the onset of the crisis), and those who enter self-employment during 1999 (whose employment incentives would be expected to be heavily affected by the brunt of the crisis), in Tables C.4 and C.5, respectively. I find that the propensity of the individuals who enter during the crisis, which can be reasonably argued to be dominated by those involuntarily forced into self-employment, to remain in self-employment is remarkably high, even after controlling for the opportunity cost of self-employment, the expected wage. All changes in propensity are strongly statistically significant, by a standard t-test.

Individuals who are self-employed during the crisis are very likely to be self-employed even after the crisis – about a tripling of the propensity to be self-employed for the 1998 entrants, and anywhere from a four to ten times increase for those who enter in 1999. As we look at years further and further from the crisis, up until 2008, the propensity to remain in self-employment remains re-

markably strong. Namely, even after accounting for the expected wage, and hence the recovery of the economy, we still see a very strong increase in propensity to remain self-employed. I take this to suggest that the individuals who involuntarily enter self-employment during the crisis manage to accumulate a significant enough stock of EHC from that experience that they become much more likely to subsequently engage in self-employment. I argue that this increased propensity is far greater than would be predicted by any reasonable model that assumes away the endogenous accumulation of EHC. In particular, the effect seems to be so qualitatively large as to exceed any reasonable frictions that might inhibit back into the wage sector, such as job search frictions.

4.6.2 The Dynamics of Self-Employment Returns

If we look at the raw numbers, we see that in the year in which the main crisis cohort enters self-employment, 1999, their self-employment earnings are about 9% lower than the counterfactual expected wage. Note that this is perfectly reasonable in a model in which (i) individuals are unexpected forced to enter self-employment (due to the crisis), and/or (ii) they anticipate dynamic increases in earnings over time. By 10 years later the situation has flipped quite strongly – expected earnings are now 16% higher than the counterfactual expected wage. While the latter figure is biased somewhat by the natural attrition of some lower-performing entrepreneurs, the bias is limited by the low attrition that has already been discussed in this cohort.

This effect seems large in terms of levels, also. By the year 2000, 1998 entrants see a 40% increase in profitability, while 1999 entrants see a 20% increase. This is

substantial, and far exceeds the growth rate of the economy (as a control for time trend). For example, we see only about an 8% increase in expected wage, which provides a reasonable and context-relevant baseline comparison to control for economic growth overall. This trajectory in returns points to a human capital accumulation dynamic. In addition, it provides further evidence against labor market frictions in explaining the lack of switching – switching costs would have to be on the order of 20% of yearly income to justify not switching back into wage employment.

4.6.3 Selection-corrected earnings dynamics

The final piece of evidence on earnings dynamics comes from taking years spent running an enterprise as a measure of learning-by-doing and entrepreneurial human capital acquisition. As discussed, I employ a version of the Heckman selection model. As also noted, part of the first-stage analysis is taken from previous work on occupational persistence, and hence I do not present those first-stage results here, rather just focusing on the second-stage earnings equation.

Tables C.6 and C.7 provide final-stage selection-corrected evidence on returns, using entrants from the years 1998 and 1999, respectively. What we first notice is that selection bias, at least according to controls based on observables, does not appear to be an important problem, as the estimated coefficient on the inverse Mills ratio is not statistically distinguishable from zero in either regression.

Looking at the coefficients on the experience variables, we see that the shape

of returns in experience is intuitive, following a concave shape for all three types, with the exception of the single proprietor enterprises amongst 1999 entrants, with easily the highest returns for those running enterprises in the greater complexity category. Among the 1998 entrants, the estimated learning effect is positive for all three types, though seemingly less persistent for those running the simplest enterprises, where the positive learning effect diminishes after just over 4 years. By contrast, the learning effect persists for around 6 years for the other types. Keeping in mind that median enterprise experience is around 5-6 years in the population, it seems safe to conclude that there is a positive learning effect. Many of the coefficients for the 1998 entrants are not statistically significant. While it is tempting to rationalize this based on low sample size, we get much more precisely-estimated effects from the smaller 1999 entrant cohort.

Looking at the results from the 1999 cohort, there is interestingly a convex estimated experience effect for individuals running the simplest enterprises, which actually implies negative returns to experience for the first five plus years running an enterprise. While the initial estimated learning effect is positive in the other two enterprise types, it is dissipates quickly.

These results emerge after controlling for endogenous selection into self-employment, again with variables plausibly exogenous to individual EHC endowments. In other words, this evidence is about as close as we could reasonably expect to get to exogenously assigning experience to individuals.

4.7 Alternative Explanations and Further Evidence

While the analysis above is derived from the two strongest sources of evidence to support the interpretation of the natural experiment – occupational persistence and earnings dynamics – the evidence does not account for all alternative explanations. This section presents tests meant to account for and cast doubt on alternative explanations for the results.

4.7.1 Capital Stock Lock-in

One possible alternative explanation for enterprise persistence, other than the posited story of entrepreneurial human capital acquisition, is capital stock lock-in. Namely, that individuals who entered during the crisis took on greater sunk costs in their capital stock investments, which they might have been reluctant to abandon as the economy recovered. While this seems somewhat implausible as the crisis period was a time of great uncertainty that saw greater levels of investment from inexperienced entrepreneurs (both of which should lead to less significant investment), I carry out a test.

I test this by looking at capital stocks held by individuals who enter self-employment during the crisis period, and those who had entered at other times. What we would expect is that if capital stock lock-in were to explain enterprise persistence, the quantities of startup capital should be larger for firms that entered during the crisis years. Since data on startup capital is only provided in IFLS4, I am forced to focus on IFLS4 as the source of data. If we look at average startup capital in the full sample of firms, it is 6722876 Rph., while for firms

which started up in either 1998 or 1999 (following the main crisis years), the average starting capital is 5027424 Rph. While both figures are somewhat biased due to survival (of firms still active during the 2007-08 survey round), since they both are subject to the same type of bias this is less of a concern.

4.7.2 Optimal Industry Selection

Another possible explanation for the positive earnings dynamic among survivors after the crisis years is that the individuals who start such enterprises during the crisis might have optimally selected into higher-growth industries (due to skill or luck, or both). Hence their earnings increases would be better explained by the effects or riding a wave during the period of opportunities that a crisis brings about. The initial suggestion seems implausible in light of the fact that the crisis was a fairly long-lasting disruptive event, and there continued to be economic and political changes well beyond the initial onset which would have made it difficult for early entrants into self-employment to parse out the best opportunities.

4.7.3 Changes in Inputs

Another potential explanation for the positive earnings dynamic is that it is driven by inputs. Namely, whether it is because individuals don't know their optimal input mix initially (and need to learn) or because of market constraints preventing initial access to the optimal input mix for new entrants, firms will increase input usage over time, and hence naturally increase returns.

However, this assertion does not fit with what we know about changes in capital stock and labor stock amongst firms. Tables A.3 to A.5 summarizes the evidence on firms in the IFLS. In particular, we focus on observed changes in enterprise size from startup to present, the last 2 rows of the tables. The tables record transitions from startup size to current size for all of the enterprises in the sample that are operating in 2008.

We see that the propensity of household enterprises to significantly change in size is quite small, whether size is measured in terms of physical capital or labor. Only 14% of firms show any growth at all in labor, and for most the growth is minimal. Even as of the 95th percentile of the distribution, firms show no change in labor stock. The average change in labor employment is actually a small decrease. Looking at the data on capital stock, we again see minimal changes in firm size; even as of the 75th percentile of the capital stock growth distribution, we only have about a $135 USD change in physical capital since startup.

Taken together, this evidence indicates that firm growth is not driven by increases in inputs alone.

4.8 Conclusion

In this chapter I develop and test a microeconomic theory of entrepreneurial human capital (EHC) accumulation. The key channel for acquiring EHC is through direct learning-by-doing. The theory is tested through exploiting a natural experiment based on the 1997-98 Indonesian financial crisis, which provides a source of exogenous assignment into entrepreneurial activity. This is useful be-

cause it pins down some confounding factors in selection into self-employment. The setting is also particularly suitable because we observe a large cohort of formal-sector workers exogenously forced into self-employment, in the more-stable informal sector. Consistent with the theory of entrepreneurial human capital accumulation, entrepreneurial activity is remarkably and robustly persistent. Even after controls for opportunity costs the propensity to be self-employed amongst this cohort increase by 2 to 9 times. The selection-corrected dynamic increase in returns to self-employment exceeds what could be reasonably expected in the absence of human capital accumulation. Taken together, these results suggest the importance of modeling entrepreneurial dynamics in a way that incorporates the role of endogenous human capital accumulation.

These results have a number of implications for policies regarding entrepreneurship promotion in developing countries. First, they highlight the importance of the accumulation of entrepreneurial human capital in enterprise outcomes. The policy implication from a model in which ability is fixed is that it is the financier's job to identify the ex ante higher-ability types as soon as possible and provide them with the full complement of financing that is proportional to their stock of entrepreneurial skill. A theory of dynamic entrepreneurial learning, however, suggests a more incrementalist approach with greater attention to timing, mixing financing provision with other skill-building services.

Second, the results pertain to institutions for the transfer of entrepreneurial human capital. In most countries the primary institution for the formation of skills for the waged-sector is formal education, which can last twelve or more years. While some writers, notably Schultz, have suggested that education might be an important venue for the formation of entrepreneurial skill, such

a hypothesis is not well supported by the evidence in this chapter. Instead, the results suggest that entrepreneurial skills are more specific and require more focused and sustained exposure to enterprise activity itself.[15] Hence this suggests the potential for specialized institutions for the transfer of entrepreneurial human capital. In most developing countries, the existing institution seems to be the family unit, at least those households in which the parents have a significant stock of entrepreneurial human capital that can be transferred to their children.

There have attempts at various forms of entrepreneurial training, including recent tests in the economics literature based on RCT designs, but based on the results in this chapter it is not so surprising that the results from short-term training have been mixed at best. While many of the existing programs are focused on transferring low-level entrepreneurial skills (keeping records, basics of managing finances, etc.), it seems that high-level entrepreneurial skills (sales, marketing, risk judgment, product development, etc.) may be significantly more important, particularly for growth-oriented firms. It may be that a more intensive, sustained mix of direct experience and perhaps mentorship from more experienced and successful entrepreneurs is needed to enable the emergence of higher-potential entrepreneurs and the transfer of high-level entrepreneurial skills.

This chapter also raises a number of questions for future research. The most obvious one regards the identification of entrepreneurial human capital and its various components? What are the most important high-level entrepreneurial skills? Are they complementary to each other, or are certain skills critical at

[15]This is not to suggest that education is not useful in general, particularly for pushing up the overall level of human capital in the population. However, the evidence herein, based on within-population variation in education and EHC, suggests that EHC is a more important *relative* factor in enterprise outcomes.

certain stages? How can such skills be effectively transmitted? A key identification challenge faced by this chapter is that many of the results could be explained not by entrepreneurial human capital accumulation that is internal to the entrepreneur, but rather an external reputation-building process amongst customers and other business partners. Of course, reputation is heavily entangled with the underlying ability and performance of the entrepreneur in question. Hence future research might employ research strategies better suited to teasing out these internal and external effects.

CHAPTER 5

A DYNAMIC MODEL OF OCCUPATIONAL CHOICE AND ENTREPRENEURIAL HUMAN CAPITAL

5.1 Introduction

An important aspect of economic development is the specialization of economic roles. This includes the emergence of higher-ability entrepreneurs who accumulate resources according to their productive abilities. Such opportunity-oriented entrepreneurs can generate employment for people with inferior entrepreneurial skills who might otherwise be engaged in subsistence self-employment, and contribute to economic growth. I distinguish opportunity-oriented entrepreneurs, who enter self-employment to pursue an opportunity, from subsistence entrepreneurs who enter self-employment due to a lack of alternatives.

This description is consistent with cross-country data. Across less-developed economies more than half of the workforce is engaged in operating or working in microenterprises, which generate roughly half of GDP. A negligible proportion of such enterprises manage to grow beyond subsistence scale. In more developed economies the contribution of microenterprises to employment and GDP is closer to 15% and an active small and medium enterprise sector contributes close to half of GDP.[1] The apparent stasis of subsistence-oriented microenterprises along with broader poverty in developing countries has led to a significant interest in policies to assist existing enterprises and promote the

[1] I use the following enterprise size categories: microenterprise (1-4 workers), small enterprise (5-19 workers), medium enterprise (20-99 workers) large enterprise (100+ workers). These are the official cohort definitions in the source country of my primary dataset, Indonesia.

emergence of new ones.

Such policies have primarily focused on broad impediments to enterprise activity, including difficulties in obtaining financing. Yet recent evaluations of such policies in the economics literature have often generated puzzling or disappointing results relative to prior expectations.[2] In particular, positive treatment effects on business activity seem to be concentrated among higher-wealth households or existing business owners, indicating important underlying heterogeneity that might go beyond differences in innate ability. The evidence presented thus far in this thesis, particularly in chapters 3-4, provides further support for this way of thinking. This raises the natural questions: might accumulated entrepreneurial skill play a role in explaining the heterogeneity in outcomes, and more broadly, what are the factors behind the emergence of opportunity-oriented entrepreneurs?

In order to address these questions I propose and test a theory of entrepreneurial human capital (EHC) accumulation. EHC constitutes specialized, entrepreneurship-specific skills and knowledge, such as in selling, negotiating, product development, risk judgment ([Shane, 2003]) and entrepreneurial social capital. Above and beyond heterogeneous *ex ante* endowments of innate EHC, perhaps due to genetic inheritance or early upbringing, I hypothesize that EHC is significantly accumulated through *direct* exposure to entrepreneurial activ-

[2]E.g., [de Mel *et al.*, 2008], [Banerjee *et al.*, 2009], and [Karlan & Zinman, 2010], who report, for example, "The canonical case for microcredit...is not supported on average. Instead the impacts are diffuse, heterogeneous, and surprising...Our treatment effects are stronger for groups that are not typically targeted by microlenders: male and higher-income entrepreneurs." [Carter & Olinto, 2003], and [Field & Torero, 2006] find analogous evidence after policy changes meant to strengthen property rights and hence make it easier for the poor to collateralize their assets. In addition, [Bruhn *et al.*, 2010] summarize recent evaluations of training programs for the self-employed in developing countries. To date the effects of such programs, usually short-term in nature, are highly mixed, and in any case provide little guidance on the emergence of higher-potential entrepreneurs.

ity. I hypothesize that the dynamic accumulation of EHC is a crucial factor in entrepreneurial outcomes, and such entrepreneurial abilities are distinct from other, generalized forms of human capital such as education, life experience, and experience in waged employment. The key channels for acquiring EHC are learning-by-doing (i.e., running an enterprise) and direct familial exposure to enterprise activity. EHC cannot be transacted in the marketplace separately from the individual endowed with it, and public institutions for the transmission of EHC (such as the formal education system for wage work) are generally absent. Hence dynamic occupational selection incentives and family background play crucial roles in individuals' ability to accumulate EHC.

I build EHC accumulation onto the canonical dynamic, discrete-choice model of occupational choice with credit constraints, in which an individual assigns a single unit of labor to either working for a wage or running an enterprise, in each time period. EHC is accumulated through learning-by-doing from enterprise experience. Crucial to the learning model is distinguishing the *quality* of EHC acquired according to the *complexity* of the enterprise to which an individual is exposed. The idea is that individuals who run more complex enterprises relative to their current stock of EHC enjoy a period of increasing returns in learning due to experiencing greater complexity in the span of entrepreneurial functions. This creates an *endogenous* learning effect.

The model provides an explanation for the bifurcation in entrepreneurial outcomes between subsistence and opportunity-oriented self-employed. Individuals with low *ex ante* EHC optimally operate relatively small, simple enterprises, and hence accumulate little EHC. Higher-ability entrepreneurs, on the other hand, optimally run more complex enterprises and enjoy far greater EHC

accumulation, eventually potentially outstripping their access to financial capital. This formalizes an endogenous poverty trap in one's stock of EHC, in which low-skilled self-employed can get stuck in subsistence self-employment.

Through this modeling approach I am able to more plausibly capture heterogeneity in entrepreneurial behavior observed in the developing-country setting. This includes evidence on the dynamic relationship between wealth and enterprise activity that is captured in models such as [Buera, 2009], but primarily through the savings problem. The theory shows how EHC could constitute a crucial omitted variable in existing studies, generating incorrect inferences in favor of financial constraints for the subsistence entrepreneurs, while underemphasizing the role of financial constraints for the select group of higher-ability entrepreneurs.[3]

To my knowledge the theoretical model in this chapter is the first to capture the life-cycle accumulation of entrepreneurial ability in the presence of selection effects and market frictions that are particularly relevant in the developing country setting. The canonical model of entrepreneurship in economics takes the stock of entrepreneurial skill as fixed.[4] Hence the theory in this chapter provides alternative microeconomic foundations for entrepreneurial dynamics.[5]

[3]This potential bias, which may provide a partial explanation for apparent high returns to capital amongst some developing-country enterprises, is discussed in papers such as [Udry & Anagol, 2006] and [Naude, 2008].

[4]An early, equilibrium model of the division of workers between wage and self-employment, assigning to each individual an *ex ante* ability parameter from a fixed distribution, is [Lucas, 1978]. [Kihlstrom & Laffont, 1979] and [Kanbur, 1979] build equilibrium theories of entrepreneurship based on heterogeneity not in ability but risk preferences. [Evans & Jovanovic, 1989] develop and estimate a static model of occupational choice between wage and self-employment under credit constraints that extracts an estimate of the fixed entrepreneurial ability parameter. [de Mel *et al.*, 2008] use the same model to rationalize results on returns to capital amongst micro-firms in Sri Lanka.

[5]In the realm of dynamic models, one thread of the literature captures the correlation between wealth and entrepreneurial outcomes in an occupational-choice setting through the mechanism that higher-ability entrepreneurs will have higher savings ([Cagetti & Nardi, 2006]; [Buera, 2009]). These models are primarily motivated by explaining the pronounced inequal-

The approach is also in line with recent literature in fields such as psychology and management on the nature of dynamic entrepreneurial learning ([Cope, 2005]; [Politis, 2008]). Such literature has been motivated by the failures of primarily static, personality-based approaches to explain entrepreneurial behavior. Instead, this newer literature has particularly focused on learning through experience.

Within the economics literature the approach in this chapter is closest to the human capital-based theories of entrepreneurship of [Schultz, 1980]. However, a crucial distinction is the emphasis in this chapter on *direct* entrepreneurial experience as the key channel for acquiring EHC, as opposed to formal education, as emphasized by Schultz. Though the conceptual development of this chapter occurred independently, it also echoes recent proposals motivated by similar empirical literature (e.g., [Bruhn *et al.*, 2010], [Bloom *et al.*, 2010]), and provides empirical and numerical evidence on some of the tests suggested in that work.

The quantitative analysis of the model involves empirical analysis of earnings trajectories using panel data methods. I am able to construct rich measures of experience in running enterprises in three distinct complexity cohorts (no employees, only family/unpaid employees, employing waged employees), and relate those to earnings outcomes, while controlling for unobserved, time-invariant heterogeneity. I show that what I find in the data is consistent with the implications of the model.

This analysis sets the stage for a full numeric analysis of the model.

---

ity in the wealth distribution in developed economies. An additional class of models stem from the seminal work of [Jovanovic, 1982], in which entrepreneurs are modeled as Bayesian learners with an unknown, fixed stock of entrepreneurial ability that they learn about only by working as entrepreneurs. [Taveras, 2010] shows that an appropriately-calibrated version of this theory can provide an alternative explanation for a number of stylized facts about entrepreneurial activity that have been used as evidence for financing constraints in the context of the US economy.

The empirical analysis contributes a number of new insights about the dynamics of entrepreneurial skill formation and particularly the emergence of higher-ability, opportunity-oriented entrepreneurs in developing countries. Much of the literature to date has been descriptive or focused on uncovering broad correlates of self-employment selection and outcomes (e.g., [Liedholm & Mead, 1999], [Nafziger & Terrell, 1996], [Fields & Pfefferman, 2003], and [Barrett *et al.*, 2005]). I employ a uniquely-suitable panel dataset from Indonesia to conduct the empirical analysis.

The key results are as follows. First, endogenously-accumulated EHC matters to entrepreneurial outcomes above and beyond innate ability and generalized forms of human capital such as life experience and education. In addition, the estimates indicate that direct exposure to enterprise activity is an order of magnitude more important than overall life experience and education in explaining enterprise performance, implying a 5-12% increase in net profits for each year of experience. In addition, I show: (1) that distinguishing varying quality in acquired entrepreneurial experience is important, with greater value derived from experience with more complex enterprises, and (2) that there are increasing returns in higher-value entrepreneurial experience, which indicates a bifurcation in entrepreneurial dynamics between subsistence and opportunity entrepreneurs that is consistent with the theory. In addition, I provide evidence on two empirical questions raised in light of the model, finding that (1) firm-specific EHC seems to be much more important than generalized EHC in enterprise performance, and (2) that the effects of entrepreneurial learning-by-doing diffuse over years; positive marginal effects of learning-by-doing are evident for 7-12 or more years in the main specification.

The results from this analysis are largely new relative to the existing literature, particularly given the dearth of existing datasets providing detailed information on entrepreneurial activity beyond recording the broad occupational category of self-employment. Datasets connecting individual entrepreneurs to their firm have only recently begun to emerge in the US, for example, and hence still have a relatively short time-series dimension. I employ a dataset, the Indonesia Family Life Survey, that is unique in that the cross-sectional population is large enough to statistically distinguish subsistence entrepreneurs from plausibly opportunity-oriented entrepreneurs with 10, 15 or more employees. Comparable existing studies are largely based on individuals running firms with 1-2 employees.

In addition, using household data allows me to overcome the selection biases and lack of information on the individual entrepreneur inherent in firm-level data. The data also provide rich information on entrepreneurial histories within the panel and recall data spanning as far back as first employment. While most existing studies use rough approximations for experience such as age minus one's age when finishing school, I am able to construct far more detailed entrepreneurial experience profiles than in existing studies, stratified by plausible proxies for enterprise complexity. I use these measures of accumulated experience as proxies for the formation of entrepreneurial skill. Hence while the empirical analysis in this chapter is limited to providing statistically-meaningful results on entrepreneurs running firms with less than 20-25 employees, the results can be taken as suggestive for the nature of entrepreneurial activity more broadly, where similar analysis might not even be possible. In addition, while a number of studies point out the important distinction between subsistence and opportunity-oriented entrepreneurs (e.g.,

[Liedholm & Mead, 1999]; [Barrett *et al.*, 2005]; [Schoar, 2010]), this study develops and tests detailed micro-level mechanisms meant to account for how such heterogeneity arises.

This chapter proceeds as follows. First I outline the dynamic model and present some results in Section 5.2. I outline the panel regression approach and present the results of that analysis, along with a couple pieces of additional evidence, in Section 5.3. Section 5.4 concludes. Regression tables, proofs and other evidence are presented in Appendix D.

5.2 Dynamic Model of Entrepreneurial Capital Accumulation

This section provides a model of entrepreneurial human capital accumulation. The model builds on the standard framework for credit-constrained occupational choice in the literature by adding the possibility of accumulating entrepreneurial human capital through learning-by-doing from occupational experience, and familial effects. I insert a learning function for the accumulation of EHC in the model that formalizes my hypotheses about the nature of entrepreneurial learning.

This addition to the standard occupational choice model provides a contribution to the literature on entrepreneurship, though it echoes a large literature in labor economics on the role of experience and job- and occupation-specific human capital that dates back to [Becker, 1964]. Such work has seen remarkably little application to the process of entrepreneurship. In addition, the structure of the learning function is meant to capture human capital accumulation processes that seem more applicable to entrepreneurship. There have been a couple

of papers that attempt to formalize similar ideas, though they are distinct from the approach here.[Otani, 1996] models the accumulation of a specialized stock of entrepreneurial knowledge, though it occurs through a specialized managerial apprenticeship stage. [Iyigun & Owen, 1998] provide a brief analysis of the accumulation of entrepreneurial human capital in an overlapping-generations setting that incorporates the role of occupational choice, though it does not lend itself as well to studying the rich set of dynamic processes that I consider.

The learning process formalized in the model captures the idea that there are increasing returns to learning as entrepreneurs scale up their enterprise relative to their ex ante stock of ability. The idea is that because the entrepreneur typically ends up managing a number of complex enterprise processes simultaneously, this can lead to a kind of multiplier effect in these joint tasks. Hence an individual running a more sophisticated enterprise relative to their current knowledge and ability will accumulate significantly more learning than an individual running a significantly less sophisticated enterprise. I formalize this learning process, along with other standard entrepreneurial tasks such as production and investment, in a dynamic programming framework. I show that the model has a tractable, recursive representation, and then provide additional analysis of the model. I show how the model formalizes static and dynamic occupational selection incentives. In particular, the model captures the possibility that the entrepreneur might "strategically" invest in enterprise activity, perhaps taking lower static returns today in order to benefit from greater future returns due to EHC accumulation. The nature of the learning function also allows for the possibility of a "firm ladder" as seen in the data, in which successful entrepreneurs make successive, large investments over time in growing their enterprise activities, rather than making steady, incremental investments. Finally,

I show how the model can generate outcomes consistent with the relationships we see in the data between wealth and enterprise activity on the one hand, and income shocks and enterprise activity on the other. We see how a strong link between wealth and enterprise activity only emerges for the relatively wealthier subset of the population. In contradiction with the standard model of credit-constrained occupational choice, relatively wealthier individuals may be more responsive to positive income shocks in terms of enterprise activity.

The theoretical work in this chapter contributes to a long line of work in economics on the role of entrepreneurs in the economy. Some of the best-known work, by scholars such as Frank Knight and Joseph Schumpeter, develops theories taking the presence of entrepreneurs as given and then studies their role in the economy. While Schumpeter distinguished entrepreneurs from capitalists, Knight's approach sees entrepreneurs' primary role in making judgments under uncertainty, which at its essence involves the allocation of resources. This approach is closer to what I have in mind here, as it emphasizes important abilities of entrepreneurs that could be considered as elements of EHC, such as judgment and foresight, which cannot be directly transacted in the market.[6]

The closest existing work, though somewhat less known, is that of the Nobel-prize winner T.W. Schultz. Schultz's perspective on entrepreneurship explicitly focused on human-capital, allowing anyone to be defined as an entrepreneur who adjusts or reallocates economic resources in response to changes in economic circumstances [Schultz, 1980]. Schultz hypothesizes about the channels through which entrepreneurial ability is formed, focusing on formal education

[6] An additional, prominent strand of work in the Austrian tradition is that of Israel Kirzner. Kirzner's theory primarily defines entrepreneurs based on their alertness to profit opportunities. This approach does not lend itself to a model of an optimizing agent, though the approach fits with mine to the extent that such 'alertness' can be seen as a developed entrepreneurial skill.

though also acknowledging the role of experience.[7] However, the theory developed herein is conceptually distinct from Schultz's approach, both in allowing entrepreneurial actions to take place outside of a state of disequilibrium, and in emphasizing direct experience and exposure, rather than formal education, as the key factor in the accumulation of EHC.

5.2.1 Environment and Primitives

In each discrete time period individuals, denoted by i, allocate a single unit of labor to either waged employment or self-employment (aka., entrepreneurship). The agents in the model can also be thought of as households, which have unitary preferences and allocate a single unit of labor. I focus on individuals because most of the empirical analysis focuses on individuals, though there is some analysis with the household as the unit observation, with the head of the household taken as the relevant individual representative. Individuals may differ on three key dimensions. First, each individual possesses a stock of wealth, denoted by $W_i \geq 0$.

Second, individuals are endowed with stock variables representing the accumulation of human capital in each of the two occupational activities. Denote the individual's stock of wage experience by $x_i \in \{0, 1, 2, \ldots\}$, which increments in years of experience in waged employment. The return to waged employment is given by $w(x_i) = x_i\xi \geq 0$, where ξ is a mean one i.i.d. shock with positive support and finite variance σ^2_ξ. The parameter ξ will be important in modeling the role of shocks to the wage sector in occupational choice incentives. I take E to

[7] A succinct discussion of Schultz's perspective on entrepreneurship, including in relation to other prominent schools of thought, is provided in [Klein & Cook, 2006].

denote the expectation operator over the distribution of wages as a function of ξ.

Finally, the individual is endowed with a stock of EHC denoted by $e_i \in \mathbb{R}_+$. I will discuss the accumulation and evolution of EHC in detail shortly, after discussing the payoff functions. I model the labor allocation choice as a discrete choice between two occupations for simplicity, as a model of the choice over the primary occupation. The model can naturally be extended to a setting in which the individual holds multiple occupations simultaneously, for example allocating some time to running a business and some time to waged employment.

The revenue function for enterprise activity is given by,

$$f(e_i, k, l), \tag{5.1}$$

where l is the labor input in the business, k is the capital input, and f is a production function with the standard properties: f is strictly increasing in each of its arguments, f is concave, all of the cross-partials are strictly positive amongst the inputs, and the standard Inada conditions hold. The lack of an output price in the revenue function can be interpreted as normalizing the output price, $p = 1$. I assume that f is weakly DRS jointly in (k, l), but I allow the possibility of increasing returns in all three factors together once the entrepreneurial factor, e_i, is accounted for.[8] However, since e_i is not a choice variable of the agent, the

[8]Formally, I assume that for any $\lambda > 1$,

$$f(e_i, \lambda k, \lambda l) \leq \lambda f(e_i, k, l),$$

but it can be that,

$$f(\lambda e_i, \lambda k, \lambda l) \geq \lambda f(e_i, k, l).$$

within-period production function is always effectively DRS in (k, l).

The individual's net income if self-employment is chosen is given by the following return function

$$R(k, l; w, r; e_i, W_i) = f(e_i, k, l) + r(W_i - k) - wl, \tag{5.2}$$

where w is a market wage and r represents the cost of renting capital within the period.

I assume that the cost of capital must be paid up front and hence the individual may be bound by the following credit constraint,

$$0 \leq rk \leq \lambda(e_i, W_i), \tag{5.3}$$

where λ represents the individual's ability to obtain additional financing to purchase business capital as a function of EHC and wealth. In general $\lambda(e, W_i) \geq W_i$ for any value of e, since the individual can always employ her own wealth to increase the capital stock of the enterprise (assuming that W_i indexes liquid wealth). I assume that λ is increasing and concave in both arguments, and twice continuously differentiable. Between own internal resources and credit access, the individual may or may not be able to attain her return-maximizing level of capital stock, $k^*(w, r, e_i)$. Namely, it may be that $rk^* > \lambda(e_i, W_i)$. Such a constraint is common in the literature,[9] and is often interpreted as arising from credit market dysfunction, perhaps due to weak contracting institutions or asymmetric in-

[9] In papers such as [Evans & Jovanovic, 1989] this constraint takes the form λW, where λ is a parameter, and W represents household wealth. This also echoes the specification in [Feder, 1985], where credit supply is given by $S(V)$, where V is land owned by the household (the natural collateralizable asset in his rural agricultural setting) and S is a supply function.

formation on the part of lenders about borrowers. Having EHC enter this function is novel,[10] and pins down a separate role of EHC from its role in the production function. It may be that EHC generates greater access to capital based on reputation, social connections that generate access to credit, etc. Hence EHC and credit access can be positively correlated.

Notice that while I allow the individual to "sell" her stock of waged-employment-specific human capital, x_i, through the wage function $w(x_i)$, I do not allow for the same transaction for EHC. While EHC enters the return function for self-employment, $R(k, l; w, r; e_i, W_i)$, when the individual is running a business, there is not a separate occupational choice in which the individual directly "sells" their EHC for a wage, for example to work as a businessperson or manager on someone else's behalf. A similar assumption is commonly made in the relevant occupational choice literature, though usually only implicitly.

In fact, it constitutes an important "missing market" assumption that is crucial to showing that credit constraints can drive heterogeneous returns to physical capital (including values well above the market interest rate). This should not be surprising, as we know from standard neoclassical theory that one missing market is insufficient to generate market price idiosyncrasies. A similar point is made by [Feder, 1985]. Nonetheless, I maintain the assumption of non-tradability of EHC. I motivate this assumption by referring to labor market (i.e., contracting) frictions that make it difficult to family enterprises to hire and monitor outsiders as managers. I provide evidence that such market frictions matter

[10] As [Evans & Jovanovic, 1989] note, "In a more realistic model, λ would depend not only on W but also on observed characteristics...since the latter affect returns and hence the probability of repayment." After citing an empirical paper providing evidence that credit access indeed seems to depend on characteristics of the entrepreneur like business experience of the applicant, size of the firm and past credit record of the firm, they note that "...in future work it would be useful to explore whether the liquidity constraint and the interest rate depend on demographic characteristics."

in Table D.1, where I present an adaption of the test of [Benjamin, 1992] for labor market frictions in the agricultural setting, which is also motivated by asymmetric information concerns. A more detailed discussion of this test if provided in the Appendix.

Now I discuss the accumulation of EHC, e_i, in more detail. EHC can be acquired through two channels. First, the individual has an *ex ante* endowment of EHC upon entering the workforce, which is denoted by $\bar{e}_i$. $\bar{e}_i$ can be strictly greater than zero because it may be that the individual has some innate EHC at birth, or acquires EHC through her upbringing, particularly due to direct exposure to entrepreneurial activity in the family unit. This could be modeled by a familial EHC acquisition function $h(\tilde{e}_i) = \bar{e}_i$, where $\tilde{e}_i$ is the individual's stock of innate EHC and h is a positive, monotonic mapping from innate ability into the stock of EHC with which the individual enters the workforce. If the individual selects self-employment, she can additionally acquire EHC through learning by doing. For simplicity, I do not allow for the possibility of depreciation of the stock of EHC, though this would also constitute a natural extension to the model.

Under learning by doing, e_i evolves according to experience acquired through occupational choices. I model this by assuming that the person running an enterprise adds to her EHC stock according to the positive learning function $g(e_i, c) = e_i'$, where $c \geq 0$ is a measure of enterprise complexity. The idea is quite intuitive – there is relatively more to be learned by an individual who is running a business that is relatively more sophisticated and unfamiliar, due to the greater span of enterprise functions that the individual faces. I capture the nature of entrepreneurial learning through the following properties. First, the

degree to which an individual learns (the "quality" of learning) depends on the complexity of the enterprise to which they are exposed. A natural index for enterprise complexity in this setting is the size of the enterprise in a given period; hence I take the complexity to be a function of labor and capital employed in the period: $c(k, l)$. I assume that c is strictly increasing and CRS in its arguments, and twice-continuously differentiable.

Second, the degree of learning depends on the individual's existing stock of EHC. In particular, there exist diminishing learning returns to complexity such that operating an enterprise of complexity c leads to relatively more learning for individual with EHC stock e_i than the individual with e_i', where $e_i < e_i'$.

Third, there are increasing returns in learning, for the right ratio of enterprise complexity to the individual's current EHC. In mathematical terms, this means that the learning function is not specified in levels, but rather it takes as an argument a function that is roughly a ratio with c in the numerator and e_i in the denominator. To understand this latter property, we can think about an individual running a very simple enterprise not learning very much, because the enterprise's functions – sales, marketing, financing, negotiating, etc. – are all very simple and quickly mastered if they require any skill at all. This would be a case where c is small, so c/e is relatively small. However, for a relatively more complex enterprise one takes on a significantly more complex and comprehensive set of entrepreneurial functions. Hence there are initially increasing returns to learning over a certain range of the complexity/ability nexus. Namely, c/e will initially be small, but because the value of next period's e is increasing in c/e today, the value of the ratio will diminish as e "catches up" over time. This is the limit at which decreasing returns take over, perhaps because the indi-

vidual suffers from time constraints, learning capacity constraints, and the like. [Jovanovic & Nyarko, 1996] model the dynamic process of acquiring knowledge about the operation of a technology, and obtain an analogous convex learning effect.

To formalize these ideas I assume that g takes the following functional form,

$$g\left(e_i, c\left(k,l\right)\right) = \ell\left(\frac{c\left(k,l\right)}{1+e_i}\right)e_i = e_i' \geq 1, \tag{5.4}$$

where ℓ is a particular case of the learning function and e_i' is next-period's stock of EHC given current stock e_i. Here, I assume that ℓ is an increasing, monotonic, twice continuously-differentiable function with the properties that $\ell\left(0\right) = 1$, and $\ell'\left(0\right) = 0$. I assume that there exists a point $\underline{e} > 0$, such that ℓ is convex up to $\underline{e}$, then concave afterwards, i.e.,

$$\begin{aligned} \ell'' &> 0 \text{ if } 0 \leq e_i \leq \underline{e}, \\ \ell'' &< 0 \text{ if } e_i \geq \underline{e}. \end{aligned} \tag{5.5}$$

The mathematical assumptions provide ℓ with the following properties, which conform with the previous description of entrepreneurial learning:
(1) ℓ is increasing in c and decreasing in e_i, so that learning is increasing in enterprise complexity and decreasing in the existing stock of knowledge. The fact that the argument of ℓ is a ratio naturally captures the idea that learning is *relative* to existing EHC;
(2) if $c = 0$, then $\ell = 1$, which is essentially a normalization implying that running the simplest firm leads to no learning regardless of one's stock of e_i;

(3) ℓ is "not explosive" in small values of e_i; because of the normalization of the denominator by $(1 + e_i)$ the argument of ℓ cannot grow exponentially as the denominator goes to zero. This is meant to capture the idea that individuals should not accumulate arbitrary quantities of learning due to starting from a low initial point;
(4) the assumptions on the second-order shape of ℓ, given in equation (5.5) imply that there are increasing returns in learning over a certain range of $c(k, l) / (1 + e_i)$, but these effects eventually diminish. This captures the idea of multiplication in the challenge of being an entrepreneur as one's responsibilities increase across all tasks, but that there is eventually a limit to such learning.

As an alternative description, the function ℓ can then be displayed IN Figure D.1 in $c/(1 + e_i)$-$\left(e_i'/e_i\right)$ space.

Finally, I model individual preferences over money by the utility function U, where money is taken as the index of consumption. U has the standard mathematical properties – strictly increasing, strictly concave, twice continuously differentiable, and Inada conditions hold. For technical reasons I additionally assume that U is bounded.

5.2.2 Bellman Equation

Let V represent represent the individual's value function, which evaluates the value of the stream of benefits emanating from a given set of state variables,

(e_i, x_i, W_i, ξ),

$$
\begin{aligned}
& V(e_i, x_i, W_i, \xi) \qquad (5.6)\\
= \; & \max \left\{ \max_{s \in [0, w(x_i)+W_i]} U\left[W_i + w(x_i) - s\right] + \delta EV\left(e_i, x_i + 1, s, \xi'\right), \right.\\
& \left. \max_{\substack{s \in [0, R+W_i],\ l \geq 0,\\ k \in [0, \lambda(e_i, W_i)/r]}} U\left[W_i + R(k, l; w, r; e_i, W_i) - s\right] + \delta EV\left(\ell\left(c(k, l) / (1 + e_i)\right), x_i, s, \xi'\right) \right\},
\end{aligned}
$$

where s represents the intertemporal savings choice problem, and $\delta \in (0, 1)$ is the discount factor. Savings must be chosen from the feasible set at the end of a period – the stock of wealth, W_i, plus the intra-period income realization, which is $w(x_i)$ in the case of waged employment and R in the case of self-employment. This lays out the agent's dynamic choice problem. Given ability endowments e_i and x_i, wealth W_i and a realization of ξ from the distribution of wage shocks, the individual makes the discrete occupational choice between waged and self-employment, accumulating ability in one of the skill areas.

5.2.3 Analysis

Here I provide analysis of the dynamic model of EHC accumulation summarized in equation (5.6). I begin by stating some basic technical results about the model that I will analyze further in the Appendix, and then I state a number of predictions of the model that I will explore in the empirical work.

The basic technical properties of the model

Here I confirm that the mathematical assumptions on the model are sufficient to imply the existence of the recursive form, and additional properties of the model. The formal definition of the relevant mathematical objects and the formal proofs of these statements are provided in the Appendix.

The value function V exists and is unique. Furthermore, V is bounded and continuous, and a stationary optimal policy ζ^* exists.

Endogenous selection effects

The value function in equation (5.6) formalizes the dynamic incentives inherent in the occupational choice problem. First, the individual faces an initial "selection" incentive. Depending on the initial (possibly non-zero) endowments of e_i and x_i the individual makes the initial occupational choice, comparing the two discrete terms in the value function. All things equal we would expect that individuals with a relatively higher stock of ability in a given occupation to self-select into that occupation. e_i and x_i might be accumulated prior to formally entering the workforce through familial effects, education, or other life experiences. Self-selecting into a given occupation can then lead to the acquisition of relevant skills that further shift the choice margin between the two occupations.

It is important to keep this initial selection problem in mind while interpreting data on occupational trajectories. Notice also how the i.i.d. wage shock term, ξ, can influence the static decision between the two occupations. While the long-run terms, $\delta EV(\cdot)$, are unchanged by the present wage shock, the com-

parison

$$U\left[W_i + x_i\xi - s\right] \lesseqgtr U\left[W_i + R\left(k, l; w, r; e_i, W_i\right) - s\right] \tag{5.7}$$

is directly influenced by the realization of ξ. While one could extend the model with additional stochasticity, it is already apparent that ξ can serve to break the potential determinism in the model emanating from the initial values of W_i, x_i, and e_i.

The model also captures dynamic selection effects that are more subtle. First, there is a strategic investment incentive due to the possibility of accumulating x_i and e_i. While a simple static comparison between $w\left(x_i\right)$ and the highest attainable value of $R\left(k, l; w, r; e_i, W_i\right)$ (which could even be negative) might imply that waged employment is optimal, the dynamic considerations emanating from the possibility of accumulating e_i (and W_i) through self-employment might lead the individual to self-select into self-employment. Because of the nature of the learning process, this kind of strategic human capital investment behavior is possible for any initial value of e_i. Hence the individual may be willing to accept a static lower payoff for a number of periods in order to enjoy a greater long-run return. On the other hand, if the financial constraint is particularly binding the individual might have an incentive to pursue waged employment for a number of periods in order to build up savings, and then switch into self-employment when a sufficient investment in the enterprise can be made.

Firm ladder and occupational persistence

The possibility of a "firm ladder" is perhaps the most interesting dynamic behavior generated by the model. The motivation for such behavior was provided

in Chapter 3, where we saw that firms tend not to change in size significantly (at least in terms of number of employees). This suggests that if entrepreneurs are seen to operate larger and larger enterprises, this generally occurs through making a large investment in a new enterprise, rather than growing a given enterprise. This fits with the idea that in order to obtain the benefits of the convex learning function, individuals must pursue operating an enterprise at a sufficient complexity level. Hence the possibility of observing a policy in which the individual makes intermittent "large" enterprise investments, saving in the meantime. In addition, once individuals select into a given occupation, having made investments they are incentivized to persist in that firm type.[11] I formalize this intuition in the form of a proposition.

For e_i, $\underline{e}$, $1/\sigma_\xi^2$, and δ sufficiently large, ℓ'' sufficiently large on $\left(0, \underline{e}\right)$, and $\lambda\left(e_i, W_i\right)$ sufficiently small on $\mathbb{R}_+^2$, there exists an optimal policy ζ^* under which $s^* > 0$ in all states, with s^* increasing in e_i, and there exists a ratio of W_i/e_i such that $e'/e \gg 1$, whereas when such a ratio is not satisfied $e'/e \approx 1$.

This raises the possibility of observing individual enterprise trajectories in which the individual saves for a number of periods, then undertakes a significant enterprise investment (enjoying the significant increase in EHC from the convex learning function, perhaps for more than one period), then persists in a relatively similar enterprise size while continuing to save.

Such a potential optimal policy also implies that we would be more likely to observe the returns function to be convex in EHC for opportunity-oriented entrepreneurs.

[11] [Keane & Wolpin, 1997] present an analysis of labor force trajectories based on this insight.

The relationship between wealth and enterprise activity

The model can match the stylized observations discussed in Chapter 3 that (i) low-wealth households seem to show little relationship between wealth and enterprise activity, and (ii) as of some percentile of the wealth distribution, the correlation between wealth and enterprise activity becomes more pronounced. Low-wealth households generally have low stocks of EHC (else they would not have low-wealth). Hence if and when they start enterprises, those enterprises are optimally relatively small, due to the static incentives inherent in R – having a small stock of EHC, it is optimal for the capital and labor stocks employed in the enterprise to also be relatively small. For initial e_i sufficiently small this static profitability incentive overwhelms the long-run incentive to invest in the enterprise in order to increase EHC accumulation. But according to the structure of the learning process, employing relatively small k and l leads to little learning, and hence persistence of the small stock of EHC, regardless of financial situation.

On the other hand, for individuals with sufficient endowments of EHC, they are induced to (optimally) start relatively complex enterprises. This puts them higher on the learning curve, which allows them to more rapidly accumulate EHC. Hence, when possible they will own-save in order to fund the optimal capital stock to match their stock of EHC. However, because λ is concave and eventually bounded the the individual can become *relatively* financially-constrained.

The response to income/wealth shocks is increasing in wealth

This is the natural extension of the previous subsection, in which relatively poor, low-ability individuals can end up in an endogenous EHC poverty trap, whereas higher-ability individuals optimally begin to own-save out of it. Based on the convex learning function, the higher-ability, wealthier individual might actually end up more constrained by financing than a lower-ability, less-wealthy individual. This drives the pent up demand for capital amongst more wealthy or higher-ability individuals that has been verified in a number of empirical studies (e.g., [Carter & Olinto, 2003], [de Mel *et al.*, 2008], [Banerjee *et al.*, 2009], [Karlan & Zinman, 2010]). Higher-wealth individuals end up becoming *more* responsive to financial access shocks than do lower-wealth individuals, contrary to the conventional neoclassical model of occupational choice but consistent with much recent empirical evidence.

5.3 Identifying the Formation of and Returns to EHC

The focus of this section is on using panel data methods to identify the role of dynamically-accumulated EHC on entrepreneurial returns, accounting for the role of innate EHC and other time-invariant factors through individual fixed effects. The key variables used to identify accumulated EHC are rich measures of enterprise experience that distinguish between the complexity of the enterprise that the individual operates (no employees, only household/unpaid employees, or hiring waged employees). I primarily focus on testing the following hypotheses: (1) that endogenously-accumulated EHC matters to entrepreneurial outcomes above and beyond innate entrepreneurial ability, (2) that distinguish-

ing the value of varying quality in acquired entrepreneurial experience is important, (3) that endogenously-accumulated EHC matters above and beyond generalized forms of human capital such as life experience and education, and (4) that there is a bifurcation in entrepreneurial outcomes according to the quality of one's enterprise experience. In addition, I provide evidence on the question of whether EHC is primarily firm-specific, or accumulated as a general stock of entrepreneurial skill across specific firms that an individual operates, and the nature of the entrepreneurial learning curve.

These hypotheses drive the empirical specifications. I begin by outlining a general framework separating the roles of innate and endogenously-accumulated EHC. Then I outline how the identification operates in a panel setting. I show that because my main regressor of interest is a dynamically-accumulated factor, panel fixed effects are only a partial solution for identification problems emanating from selection. Then I provide a brief description of the data. I then provide regression analysis on the role of EHC. The main results are derived in a log-linear panel regression setup. I then conduct subsequent analysis that parses out the results using quantile regressions.

I then estimate an alternative, Cobb-Douglas model of EHC accumulation, which I interpret as providing an estimate of the "EHC production function." I use the results of this analysis to derive the implied distribution of endogenously-accumulated EHC in the population.

Finally, in order to deal with the identification challenges in identifying EHC, due to endogeneity of dynamically-accumulated EHC resulting from *ex ante* selection, I run the same panel regressions on selected cohorts. This is meant to further purge the role of unobserved factors in selection into entrepreneurship

and hence provide supportive evidence from smaller subsamples.

5.3.1 Framework for Distinguishing Innate and Accumulated EHC

The theory formalizes the idea that entrepreneurship-specific human capital is both (i) innate or accumulated through experiences during upbringing, and (ii) built through experience – direct exposure to enterprise activity through learning-by-doing (i.e., someone running an enterprise oneself) and learning from others (i.e., close exposure to how an enterprise is run, perhaps from family members (e.g., in a family business). So we can think of EHC, e_{it}, as being an unobserved stock variable composed of both fixed and evolving components for individual i,

$$e_{it} = (\overline{e}_i, e(exp_{it}, z_{it})), \tag{5.8}$$

where $\overline{e}_i$ is the fixed stock of EHC of person i acquired prior to entering the workforce, and $e(exp_{it}, z_{it})$ represents an EHC shifter function, capturing the way that one can augment their stock of EHC during their work life.[12] exp_{it} represents a vector capturing cumulative entrepreneurial experience as of period t, and z_{it} records other factors (e.g., education) that could influence the acquisi-

[12]This setup is related to the recent work by Heckman and Cunha on the technology of skill production (see, e.g., [Cunha *et al.*, 2006]). In that work they model and empirically estimate the production of occupational-relevant skills during upbringing, focusing on the role of critical stages in skill acquisition and complementarities between cognitive and non-cognitive skills. The analysis here takes many of the particular skills that Heckman and Cunha focus on as given, since those early skill formation processes are not easily measured here and of less interest anyway since they form more generalized human capital. In a standard Heckman-Cunha model we have a given stock of skills, and it is augmented through investments. The approach is similar here. In the data investments are identified as units of time spent in direct or at least close exposure to entrepreneurship. Time units are parsed by a quality index, which is essentially an index of enterprise size/complexity, which seems a good index of quality in these data.

tion of EHC.[13] I will take exp_{it} to be a vector that counts years of experience in directly running an enterprise, stratified in two directions. First, wherever the IFLS records occupational histories it distinguishes between experience running an enterprise with no employees, with only family or unpaid workers in the enterprise, or with outside, paid workers. While certainly not a perfect measure, I argue that distinguishing enterprise experience in this way is superior to other, more ad hoc ways of classifying the quality of self-employment experience. The hypothesis is that this distinction should make a difference–that years of experience in a more complex enterprise (i.e., which provides stable employment for hired workers) should have higher value than other forms of experience. Second, I stratify exp_{it} by years of experience in one's *current* enterprise in a given year, or whether experience was recorded in a *prior* enterprise. This allows me to distinguish between the value of firm-specific human capital and one's longer-term accumulated enterprise experience.

The distinction between innate, fixed EHC, $\overline{e}_i$, and that which is endogenously acquired, $e(exp_{it}, z_{it})$, is greatly facilitated by the availability of panel data. Indeed, the theory predicts that individuals with higher *ex ante* ability should be more likely to enter self-employment, all things equal. But this should then lead to further accumulation of EHC, according to the theory. This implies that *ex ante* ability should be correlated with endogenously-acquired EHC, or in other words, that there is unobserved selection bias in who accumulate EHC. If *ex ante* ability could not be controlled for, this would lead to upward bias in the estimate of the role of endogenously-acquired EHC. I show that under certain assumptions, the panel structure of the data allows me to control for *ex ante*,

[13] Of course, in theory z_{it} includes important unobservable factors such as learning effort. To the extent that such variables are unobservable I assume that they are distributed mean-zero and independent in the population so that they are captured by the noise parameter in a regression model.

fixed ability through an individual-specific fixed effect.

5.3.2 Identifying EHC in a Panel Regression Framework

The main approach here is to build on the large learning-by-experience literature in labor economics, and suppose that accumulated entrepreneurial ability is primarily generated by experience. To be concrete, let Y_{it} be the period-t observed value of some entrepreneurship-relevant outcome of interest, such as entry or enterprise returns. Then theoretically Y_{it} is a function $f(\bar{e}_i, e(exp_{it}, z_{it}), x_i, x_{it})$, with x_i and x_{it} representing other fixed and time-varying factors from enterprise inputs like labor and capital to other characteristics of an individual, household, industry or region.

In taking this model to the data, $\bar{e}_i$ and e_{it} are not directly observable, only some components of (x_i, x_{it}) are observable, and the form of f is not known. In order to make this empirically implementable I make a couple of assumptions:

1. Since the functional form of $e(exp_{it}, z_{it})$ is not known, I assume that $e(.)$ can be approximated through a parametric functional form.

2. While the form of f is unknown, I assume we can approximate it as follows. First suppose that f is additively separable in $\bar{e}_i$ and the unobservable components of x_i and x_{it}, which I denote by $\tilde{x}_i$ and $\tilde{x}_{it}$ (and with slight abuse of notation denote the observable parts by the original label), and affine in $\bar{e}_i, \tilde{x}_i$ and $\tilde{x}_{it}$; that is

$$Y_{it} = f(\bar{e}_i, e_i, x_i, x_{it}, \tilde{x}_i, \tilde{x}_{it}) = \bar{e}_i + \tilde{x}_i + \hat{f}(e(exp_{it}, z_{it}), x_i, x_{it}) + \tilde{x}_{it}, \tag{5.9}$$

so that I can control for $(\bar{e}_i + \tilde{x}_i)$ through an individual fixed effect. To control for $\tilde{x}_{it}$ (unobserved individual-specific time effects) I assume that $\tilde{x}_{it} = \bar{x}_{it} + \nu_{it}$, where $\bar{x}_{it}$ is homogenous across individuals and ν_{it} is i.i.d. noise, so I can control $\tilde{x}_{it}$ through a time period fixed effect.

Notice that this approach does not allow me to compare the role of *ex ante* EHC to endogenous EHC (i.e., $\bar{e}_i$ versus e_{it}) since the fixed effect $(\bar{e}_i + \tilde{x}_i)$ includes $\tilde{x}_i$, which in principle can be of any sign and magnitude. The most I can conclude from this setup is that e_{it} is statistically/economically significant after controlling for $\bar{e}_i$. In addition, notice that if key inputs in the profit function, such as labor and capital, are not accounted for, then $\tilde{x}_i$ controls for the average value of these inputs, while the deviations from the average value of the inputs enters the error term. According to the theory, this should imply some positive correlation between the error term and the measure of EHC.

In this approach I take (exp_{it}, z_{it}) to perfectly measure endogenous human capital, and assume that I know the functional form of $e(.)$. Hence I run the empirical version of equation (5.9) in one stage in a panel regression, with the following empirical specification,

$$Y_{it} = \bar{e}_i + \tilde{x}_i + \hat{f}(exp_{it}, z_{it}, x_i, x_{it}) + \tilde{x}_{it} + \mu_{it}, \tag{5.10}$$

where μ_{it} denotes all other residual noise, measurement and sampling error, etc.

Hence the regression equation becomes the following

$$Y_{it} = \gamma_i + g(exp_{it}, z_{it}, x_i, x_{it}, \beta) + \varepsilon_{it}, \tag{5.11}$$

where $\gamma_i = (\bar{e}_i + \tilde{x}_i)$ can be estimated as a regression fixed effect term, g will be a linear-in-estimators function, with β representing a vector of regression coefficients to be estimated, and $\varepsilon_{it} = \tilde{x}_{it} + \mu_{it}$ is the error term. In order to generate consistent estimates, I need to assume that $\tilde{x}_{it}$, representing unobserved, time-varying factors, is independent in i and t and mean-zero. Unfortunately, this implies heteroskedastic errors if unobserved factors are time-varying. In practice I am not able to include t as a regressor itself, due to collinearity with a number of other explanatory variables that follow similar time increments.

In practice I will consider a couple of alternative functional forms on the above. First, I consider the possibility that the right-hand-side is an exponential function; that is, that

$$Y_{it} = \exp\left(\gamma_i + g\left(exp_{it}, z_{it}, x_i, x_{it}, \beta\right) + \varepsilon_{it}\right). \tag{5.12}$$

Taking the natural log of both sides of this equation, I generated a log-linear estimation equation,

$$\ln Y_{it} = \gamma_i + g\left(exp_{it}, z_{it}, x_i, x_{it}, \beta\right) + \varepsilon_{it}. \tag{5.13}$$

Second, I consider a Cobb-Douglas functional form, allowing for the possibility of complementarities between the experience and other terms. I assume that the following functional form,

$$Y_{it} = e^{\gamma_i} e^{\alpha X_{it}} e^{\varepsilon_{it}} \left(1 + z_{it}\right)^{\beta_0} {}_{j=1}^{K} \left(1 + exp_{it}\right)^{\beta_j}, \tag{5.14}$$

where K is the number of experience measures employed, X_{it} represents addi-

tional linear controls, and α is the coefficient on the additional linear controls. Taking natural logs of both sides and adding some additional linear controls I obtain the following log-log regression equation,

$$\ln Y_{it} = \gamma_i + \beta_0 \ln(1 + z_{it}) + \sum_{j=1}^{K} \beta_j \ln(1 + exp_{it}) + \alpha' X_{it} + \varepsilon_{it}. \tag{5.15}$$

5.3.3 Description of the Data

The analysis in this section is conducted with the individual as the unit of observation. In the IFLS we have detailed information on year-by-year occupational profiles, with years between panel rounds covered by recall in the subsequent survey round. The analysis is restricted to individual-year observations on individuals who report running an enterprise as their primary occupation in the given year, with the top 5% and bottom 5% of the earnings distribution trimmed off.[14] Summary statistics on individuals in the panel who operated enterprises, and enterprise experience profiles, are provided in in Tables A.7 and A.6.

Table A.7 provides a summary of basic individual characteristics, parsed out by observations of individuals running one of the three enterprise types. The average age of self-employed individuals in the overall sample is 43, with a median of 53. This tends to fit with the profile of self-employed in more advanced economies, where self-employment tends to be chosen after a few years of wage work experience. However, we see that individuals running enterprises with no employees, or with waged employees, tend to be slightly younger on average than those running family enterprises. 62% of the self-employed are male

[14]The results have been checked for robustness against this sample selection criterion. The results do not show significant change under other sample selection rules.

and 60% are married, though in the parsed statistics we see that entrepreneurs running enterprises with waged employees are slightly more likely to be male, and to be married. The average number of years of education is just over two overall, though it is quite noticeably higher for those who run enterprises with waged employees. A simple summary of monthly net profits (enterprise earnings) in Indonesian rupiah is provided, pooled across individuals and years. Mean earnings are 260,000 Rph per month, or the equivalent of about $26 per month in US dollars with the rule of thumb of a 10,000 Rph = 1 USD exchange rate. Earnings for individuals running enterprises with waged employees are significantly higher.

A summary of statistics on enterprise experience profiles is provided in Table A.6. This table pools data across individual-year observations (hence any given person may be represented up to 20 times in this table). The first row of each of the three subsections summarizes experience in years for a given survey year in an individual's *current* enterprise. The average current number of years of experience running an enterprise with no employees is about 6, with a median value of 5. Years of experience in a current enterprise employing family/unpaid workers is just over 6, with a median value of 5. Years of experience running an enterprise employing waged workers is lower in the mean at 4.9, and with a lower median at 4.5. Notice that the sample of individuals running enterprises with waged workers is much smaller than the other two categories. The last three rows of each subsection of the table summarize total enterprise experience profiles – years of experience running one's current enterprise, *plus* any prior experience running other enterprises. The respective means and medians across the three enterprise categories (no employees, only family/unpaid, waged) are as follows: (10.77, 7), (11.15,7), and (7.2,5). In general this indicates

that individuals tend to accumulate somewhat less experience running more complex enterprises (those that hire permanent, waged workers).

5.3.4 Results

Tables of results are provided in Appendix D.

Log-linear Panel Regressions: Main results

The main panel regression appears in Table D.2, controlling for unobserved heterogeneity through fixed effects, with the relationship between accumulated EHC and outcomes (net profit) given in log-linear form. Accumulated EHC matters above and beyond one's innate stock of ability – experience-based proxies for accumulated EHC have statistically and economically significant effects on the key measure of performance, net profit. The preferred specification is in column (6), which contains the largest set of controls. An extra year of experience, depending on enterprise type, leads to a 5-12% increase in profits in the first-order term. The shape of returns in EHC is concave for those running enterprises with no employees, or only family/unpaid employees, though the coefficients on the quadratic term are about an order of magnitude smaller than those on the linear term. These estimates also indicate that the learning effect persists for a number of years, at least 6.5 and up to 12 years. This suggests that the entrepreneurial learning-by-doing persists for a significant amount of time, and hence that the skill accumulation process is quite long-lasting.

For those in the highest enterprise category, in the preferred specification we

see that the relationship of current experience to returns is positive and convex, in columns 1, 2, 5 and 6 of Table D.2. Matching this result to the model, the hypothesis that individuals running "more complex" enterprises would be in a higher-returns region of the learning curve, perhaps even enjoying increasing returns to EHC, fits with the evidence in the data.

Additional evidence is provided through a Cobb-Douglas log-log specification in Table D.3, which I interpret as capturing the "production function" for EHC, if we interpret net enterprise returns to represent an index of entrepreneurial skill (after accounting for other observables). Consistent with the prior analysis, the direct enterprise experience variables are statistically and economically significant. Again, the total career experience terms seem to have a negative relationship with returns, particularly in the preferred specification in column (6).

Accounting for Unobservables: Selected cohorts based on age and experience

In this section I carry out a final piece of analysis that is meant to deal more rigorously with the role of the endogeneity of entry. As noted, measures of experience do not constitute perfect measures of EHC, because individuals endogenously select into self-employment, and hence EHC accumulation, at least partly as a function of (unobserved) *ex ante* ability. In order to purge some of these effects, in this subsection I re-run the analysis on a couple of key subsamples: a group of individuals who form a balanced panel by all having been self-employed for 8-12 years (Table D.5), and a cohort of individuals who enter self-employment in their 20s (Table D.6). The idea is that such sample selection better accounts for unobserved heterogeneity that might bias the results,

by looking at the role of experience in a more homogenous sample.

What I find is that these results provide some support for the previous results. The experience variables are sometimes less likely to be statistically significant, due to the smaller sample sizes involved. The shape of returns in total experience generally seems to be concave for individuals running less complex enterprises (categories 1 and 2), though by contrast returns from experience in the current enterprise are more likely to have a convex shape (positive second derivative). This could be reflective of the fact that the selected sub-samples are drawn from individuals that have better learned their optimal enterprise-type match. The shape of the learning process is more likely to have a convex shape for individuals running enterprises with outside wage workers, as before. It is notable that the variables on total experience are more likely to have a positive sign in these regressions, whereas in the previous analysis they often entered with negative signs. This may again be because the samples selected in this part of the analysis are more likely to select on higher-potential entrepreneurial types (longer survivors or those who enter self-employment at a younger age). We also see that the variable for age enters with larger magnitude than in the previous analysis.

5.3.5 Summary

What I find in this section is broadly fitting with the predictions of the theory.[15] First, accumulated EHC seems to matter above and beyond one's innate

[15]The results have been further checked for robustness, such as to sample selection. In addition, quantile regression estimates of these same results, stratifying across the earnings distribution, provide further confirmation for the main results. These results were available in the working paper version of this chapter.

stock of ability. Proxies for accumulated EHC have statistically and economically significant effects on the key measure of performance, net profit, even after controlling for the role of innate ability and other non-observable factors through panel fixed effects. Second, as hypothesized, the quality of enterprise experience seems to matter. Third, endogenously-accumulated EHC matters above and beyond generalized forms of human capital such as life experience and education.[16] The coefficients on the variables measuring accumulated entrepreneurial experience are generally significantly larger than those for age and education. Fourth, it seems that firm-specific EHC (tenure) is relatively more important than one's total accumulated EHC in outcomes. This is supportive of arguments in the literature on entrepreneurship on how entrepreneurial experience and the firm are fundamentally intertwined and provides alternative support for idea that EHC is not easily tradable in the market. Finally, I find that the "convex" relationship between EHC and returns, particularly in more complex enterprises, seems to be supported in the data. The fits with the nature of the learning specification in the model.

It should additionally be noted that these patterns are only partially supported by additional analysis focusing on what might be described as more committed entrepreneurs – those with relatively more enterprise experience, or those who started running a business earlier in life. Even individuals in these sub-groups running relatively less complex enterprises tend to look more like those individuals running more complex enterprises in the full sample.

[16]This is not to argue by any means that education is of little value for entrepreneurship. Education may very well be the crucial variable explaining *levels* of entrepreneurial performance across countries, for example. However, in a relatively homogenous population it appears that *relative* entrepreneurial experience plays a more important role than education in outcomes.

5.3.6 Additional Evidence

The Role of EHC versus Financial Capital

A natural way to bring the analysis back to the original motivation is to look at the question: what matters relatively more for enterprise outcomes, entrepreneurial human capital or financial capital? In this subsection I test this question by looking at a different unit of observation–the household. This is because it is difficult to robustly allocate household assets to each individual with the household. Instead, I look at the choice to startup and invest in enterprises at the level of the household, measuring the EHC stock of the household by the EHC measures of the household head. My approach is to regress the propensity to engage in self-employment on various community and household-level characteristics (most critically, household wealth) in a probit framework, and then in a second stage add in measures of accumulated EHC of the household head through experience measures as used extensively in this chapter.[17]

As baseline explanatory variables, I employ: age and age-squared of the household head, measures of available household labor (adult male, adult female, child), wealth in 2000 and the square, and various measures of financial access–dummy variables for the presence of a number of financial institutions, including Bank Rakyat Indonesia, People's Credit Bank, and a Village Credit Union, a dummy variable for whether or not the location's province is in Java, and a dummy for whether the enterprise is in an urban area or not. The EHC

[17] A natural bridge between the flexible non-parametric approach and the linearized approach is a semi-parametric model in which key variables (e.g., wealth) can be specified to have a flexible relationship with the propensity to start an enterprise. I have conducted such analysis (not reported), analogous to the parametric evidence below, using a spline-based approach in the mgcv package in R, and find qualitatively similar results to the non-parametric evidence, inserting additional linear controls on the startup decision.

variables are calculated as described above: enterprise experience of the household head in current enterprise activity, household head's experience in prior enterprises with no employees, household/unpaid employees and waged employees, respectively, then the same distinction regarding the household head's first occupation (dummy for whether the head ran a business in each of the three categories). At this time I simply report coefficients and not marginal effects; the purpose here is mainly to illustrate the importance of the EHC measures for the decision to start an enterprise.

The results are reported in Table D.7. To be clear, these are cross-sectional estimates. What we see is that the experience measures are generally statistically-significant, with signs fitting with the theory. While the direct first-order coefficient on current experience is negative, this is probably a mechanical relation due to the fact that the household head most often runs the household enterprise (and hence in already running an enterprise, would be less likely to start running a *new* enterprise, which is what the dependent variable measures). More interesting are the coefficients on the prior experience variables–we see that the largest effect seems to be from prior experience running an enterprise with waged employees. The marginal insignificance of enterprise experience in one's first occupation is probably mainly due to the low propensity of people to have their first occupation be in running an enterprise in general. Most importantly, the coefficients on the wealth and wealth^2 terms are statistically significant only in the first-order term, and only marginally significant in the first specification. In addition, I find that the inclusion of a term capturing the lag of number of enterprises owned, which I interpret as a measure of accumulated entrepreneurial capital within the household, leads household wealth to go insignificant in its in influence on the propensity to start a new enterprise.

In light of my model, which incorporates the role of both financial constraints and EHC, I interpret this as fitting with the prediction that EHC and wealth should be correlated. There we see that EHC is the fundamental driver of both wealth and enterprise outcomes, and hence would be expected to significantly "wash out" the role of wealth in enterprise behavior once included. This (positive) relation between wealth and EHC obtains because higher-ability individuals have greater incentive to save, and they obtain higher returns due to their higher ability.

However, what these results also point to is significant omitted-variable bias involved from running financial measures and omitting the role of EHC. We will over-estimate the role of financing in enterprise activity and outcomes if we omit the role of unobserved entrepreneurial ability, which according to the theory actually drives both wealth and entrepreneurial dynamics.

As a form of robustness check I have conducted similar analysis looking at enterprise returns, and also controlling for the physical capital stock of the enterprise. These further results are consistent with the results I have enclosed herein.

5.4 Conclusion

In this paper I develop and test a microeconomic theory of life-cycle entrepreneurial human capital (EHC) accumulation. I hypothesize that endogenously-accumulated EHC is an important factor in entrepreneurial performance. The key channels for acquiring EHC are through direct learning-by-doing and familial transmission. I further posit that the quality of EHC accu-

mulation matters – namely, the complexity of the enterprise to which one is exposed affects how rapidly EHC is accumulated. Given that EHC is not tradable in the market, dynamic occupational selection incentives and market frictions can inhibit its accumulation. Building on complementary work in fields such as management and psychology that emphasize the role of dynamic experience effects in the accumulation of EHC, I formalize the theory in a natural, dynamic setup. I show that the model captures the possibility of an endogenous EHC poverty trap, whereby low-ability self-employed start up relatively simple enterprises and hence accumulate little EHC.

The theory is tested through exploiting rich information on occupational histories in the data, which allows the accumulation of entrepreneurial human capital to be modeled while controlling for innate ability, and an alternative empirical strategy exploiting significant occupational churning during the 1997-98 East Asian crisis. I present additional evidence on familial effects, the connection between financing and latent EHC, and the correlation of a number of common behavioral measures to entrepreneurial activity. I show that endogenously-accumulated entrepreneurial human capital is a statistically and economically-important factor in enterprise outcomes, above and beyond innate ability and generalized forms of human capital. Hence I show that the omission of EHC constitutes an important omitted variables bias in studies regarding the sources and performance of entrepreneurs. The results support the argument that the crucial constraint to most microenterprises is not financing, but rather entrepreneurial ability. Furthermore, even for higher-potential entrepreneurs, the results highlight the nature of the entrepreneurial learning curve, which seems to require years of direct experience with and exposure to enterprise activity before one reaches significantly diminishing returns.

These results have a number of implications for policies regarding enterprise and entrepreneurship promotion in developing countries. First, they highlight the importance of the accumulation of entrepreneurial skill in enterprise outcomes, which suggests greater care be taken in determining the optimal timing of financing access. The evidence from this paper, and from field tests of financing provision, seem to suggest that relatively few individuals are ready to make good use of financial transfers for enterprise activity. The policy implication from a model in which ability is fixed is that it is the financier's job to identify the *ex ante* higher-ability types as soon as possible and provide them with financing. The results in this paper suggest that there is a life cycle process of entrepreneurial development, requiring a mix of learning, increased access to financing, and perhaps even the freedom to fail. This is not to say that policy approaches such as microfinance are not useful in general, but that we should not be surprised if relatively few individuals see much more than incremental improvement to their enterprise outcomes due to increased access to financing. The results highlight the need for financing policies to be much more focused on identifying entrepreneurs in the "take-off" stage, once they have accumulated sufficient entrepreneurial skill and are ready for significant scale-up in the scope of their enterprise. On this see, e.g., Beck (2007) and Beck et al. (2008).

Second, the results pertain to institutions for the transfer of entrepreneurial skill. In most countries the primary institution for the formation of skills for the waged-sector is formal education, which can last twelve or more years. While some writers, notably Schultz, have suggested that education might be an important venue for the formation of entrepreneurial skill, such a hypothesis is not supported by the evidence in this paper. Instead, the results suggest that entrepreneurial skills are more specific and require more focused and sustained

exposure to enterprise activity.[18] Hence this suggests the potential for specialized institutions for the transfer of entrepreneurial skill. In most developing countries, the existing institution seems to be the family unit, at least those households in which the parents have a significant stock of entrepreneurial skill that can be transferred to their children. While there have been broad attempts at various forms of entrepreneurial training, based on the results in this paper it is not so surprising that the results from short-term training have been fairly mixed. It may be that a more intensive, sustained mix of direct experience and perhaps mentorship from more experienced and successful entrepreneurs is needed to enable the emergence of higher-potential entrepreneurs.

[18]This is not to suggest that education is not useful in general, particularly for pushing up the overall level of human capital in the population. However, the evidence herein, based on within-population variation in education and EHC, suggests that EHC is a more important *relative* factor in enterprise outcomes.

APPENDIX A

APPENDIX TO CHAPTER 2

A.1 Overview of Data Available in the IFLS

I summarize the main variables available in the IFLS, at the individual, household, and community level.

A.1.1 Individual-level data

The individual survey component of the IFLS includes the following booklets:[1]

- Book IIIA contains detailed information on education (including on language and literacy, detailed school educational history, record of standardized test scores at various levels, education expenditure), subjective well-being (including qualitative rankings relative to society and one's and one's family's own needs), household assets (details on homeownership, land, animals, vehicles, appliances and household durables, savings, receivables, household), non-labor income (from government, insurance, prizes), marital history, household decision-making, pregnancy summary, migration activity, employment and labor activity (including details of work history regarding first job, present primary and secondary job, and detailed retrospective information on prior employment going back to previous round of the survey), retirement, elicitation of risk and

[1] In this subsection I summarize the questionnaires from IFLS4, which generally has the richest data, though much of these variables are contained in IFLS3, and a good amount is in IFLS2 and IFLS1.

time preferences, expectations and trust (i.e., community social norms, religiosity).

- Book IIIB contains detailed information on health, such as smoking behavior, detailed questions about current and recent health condition, chronic medical conditions, mental health, cognitive capacity (i.e., ability to recall date, memorize lists), acute morbidity, health insurance, self-treatment, outpatient care, frequency of various food consumption, inpatient care, participation in community activities, detailed information about non-coresident parents, siblings and children, giving and receiving transfers (of some value, including loans and borrowing),[2] and expectations for children.
- Book IV is specifically for an ever-married woman age 15-49, asking about marital history, pregnancy, non-coresident children, and expectations.
- Book V is for children less than 15 years old, asking about the child's education, acute morbidity, out-patient care, food frequency, inpatient care, and information on the parent.
- There is a Proxy Book, which appears to be a shorter version of the overall individual survey, perhaps for situations in which there is not time to administer all of the questionnaires.

A.1.2 Household-level data (including enterprise data)

The household survey component of the IFLS includes the following booklets:

[2]There were detailed questions about borrowing behavior in IFLS3, but that may have been largely removed for IFLS4.

- Book T is a tracking book that briefly records basic information about the household (mostly on location). There are additional books that also regard the tracking purpose.
- Book K is a control book, which records the household roster and a number of basic characteristics of household members.
- Book I is answered by the female head of the household. It asks about details of consumption (and expenditure), assistance received from the government (for food), and knowledge of health and family planning services.
- Book II is about the "household economy." It asks about features of the household (status of the house (owned, rented, etc), water sources and use, sewage and garbage, household language, details of assistance received from the community, a number of pages of information on household farm businesses, a number of pages on non-farm enterprise, household assets (with breakdown into numerous categories)), household non-labor income, and shocks and hardships to the household in the past 5 years.
- Books US I and US II take health measurements on all of the household members.

A.1.3 Community and facility-level data

The IFLS involves data collection in 312 communities. The community survey component of the IFLS includes the following booklets:

- Book I gets at transportation and infrastructure (including nearest transport terminals, nearest market, nearest telephone, nearest bank or finan-

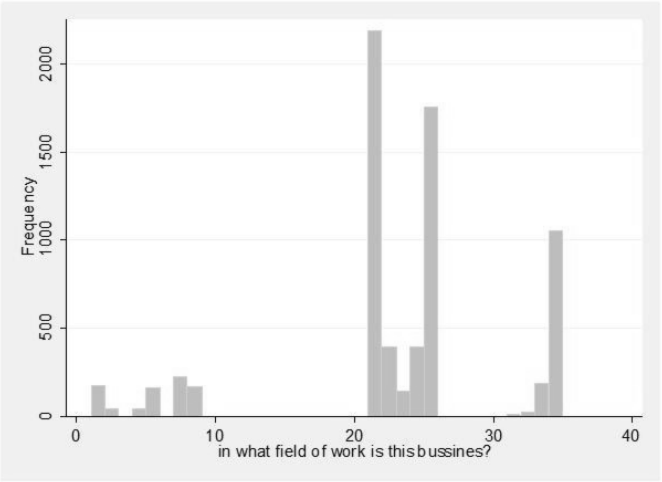
cial institution, public motor vehicle service, road quality), availability and quality of electricity (including on providers), water sources and sanitation, agriculture and industry (including main crops in the community, types of irrigation, information campaigns, current wages for various forms of farm laborers, factories and their outputs in the community, public works projects), history and climate of the community (including important events such as infrastructure construction projects), natural disasters, savings and borrowing (i.e., details on various financial institutions in the village), history of the presence of schools, history of the presence of health facilities, citizen participation in the community, subjective well-being across the village according to the informant, poverty alleviation programs, perceptions of public services and infrastructure, government and decentralization, and trust within the community.

- Book OL is based on direct observations of characteristics of the community such as air quality, security, shopping, etc, and also environmental conditions, land certification, housing, types of employment, and village finances.
- There are additional booklets including: Book PKK for women's groups, Book ADAT for traditional law and community customs, a Book for Community Health Centers, a Book for private health practices, a Book for Integrated Community Health Posts, a Book for traditional health practitioners, a Book on Public Perception on Government Programs and Public Services, a Book on community and facility people characteristics, and finally 3 Books on various prices in the village.

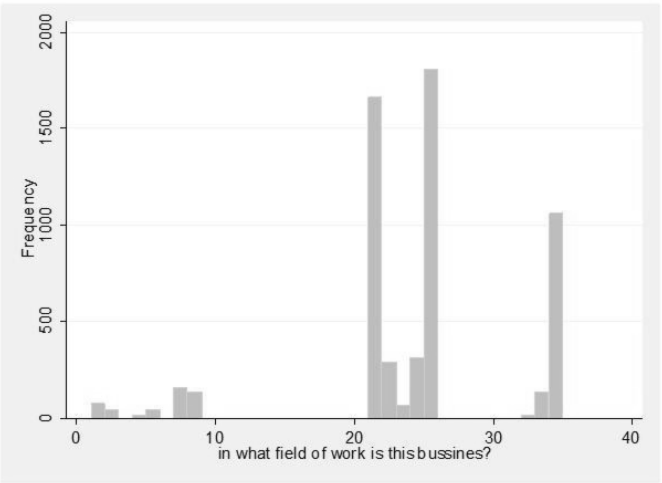
A.2 Figures

Figure A.1: Map of Indonesia, highlighting IFLS provinces

Note: The map of Indonesia highlights the provinces in which the Indonesia Family Life Survey has been conducted, in a darker shade.
Note: Image from ⟨http://www.rand.org/labor/FLS/IFLS/⟨.



(a) Field of work of business, 2000



(b) Field of work of business, 2008

Figure A.2: The distribution of firms by industry, 2000 and 2008

Note: The IFLS uses the following industry codes: 1=Ag, Forest, Fish, 2=Mining, Quarrying, 4=Elec, Gas, Water, 5=Construction, 7=Transport and Comm, 8=Fin, Ins, Real Estate, 21=Restaurant, Food Sales, 22=Industry: food processing, 23=Industry: clothing, 24=Industry: other, 25=Sales: non-food, 31=Serv: govt, 32=Serv: teacher, 33=Serv: professional, 34=Serv: Transport, 35=Serv: Other, 95=Other.

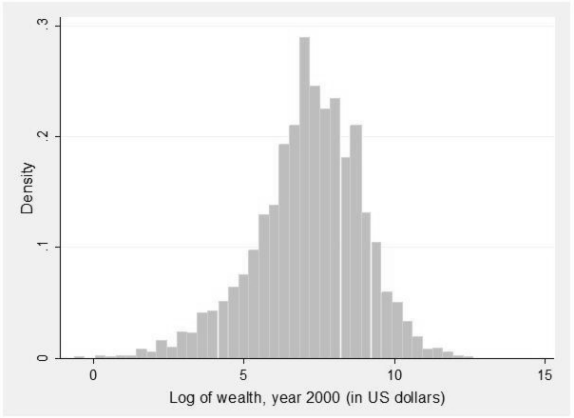
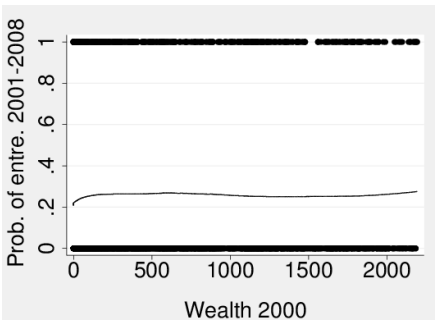
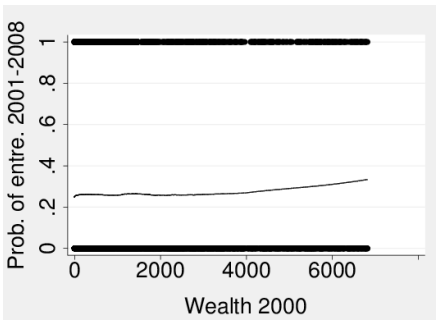


Figure A.3: Household wealth distribution, in logs, year 2000

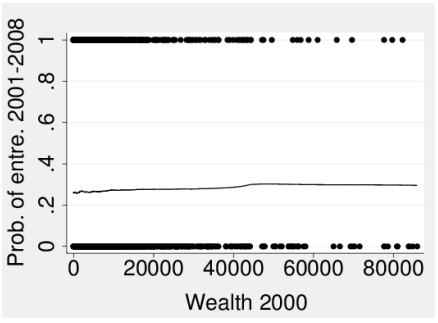
Note: Wealth converted into US dollars in the year 2000. Wealth taken as a sum of reported wealth values in a broad range of asset categories, as described elsewhere in the Chapter.



(a) 0-50th percentile



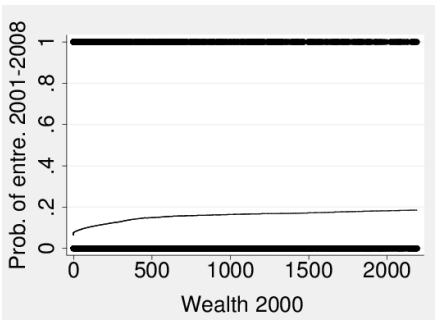
(b) 0-75th percentile



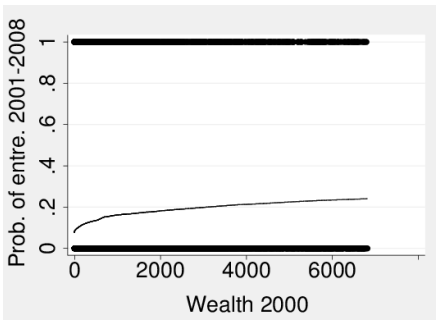
(c) 0-99th percentile

Figure A.4: The relationship between wealth and the propensity to engage in enterprise activity, urban households parsed by wealth percentile.

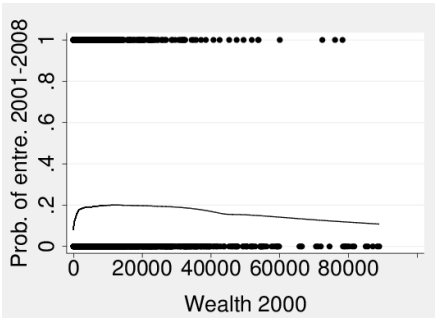
Note: In each sub-figure, the x-axis records household wealth (in terms of US dollars in 2008) in year 2000, while the y-axis records the propensity to engage in enterprise activity in 2001-2008. Non-parametric graphs generated by a tri-cube smoother, which overweights local observations in placing fitted curve. In all cases a bandwidth of 0.8 is used.



(a) 0-50th percentile



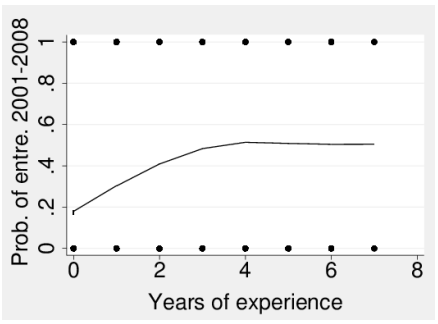
(b) 0-75th percentile



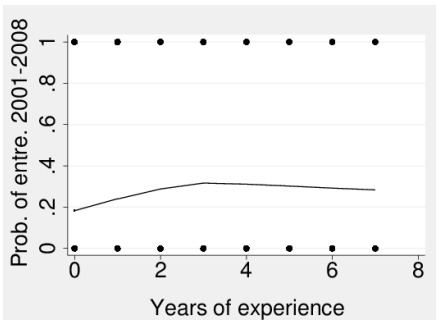
(c) 0-99th percentile

Figure A.5: The relationship between wealth and the propensity to engage in enterprise activity, rural households parsed by wealth percentile.

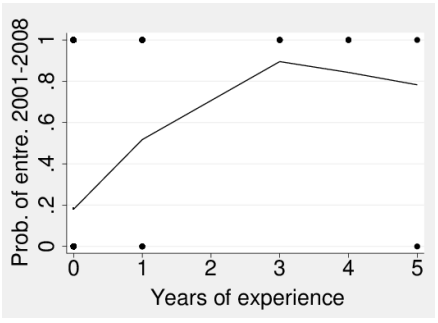
Note: In each sub-figure, the x-axis records household wealth (in terms of US dollars in 2008) in year 2000, while the y-axis records the propensity to engage in enterprise activity in 2001-2008. Non-parametric graphs generated by a tri-cube smoother, which overweights local observations in placing fitted curve. In all cases a bandwidth of 0.8 is used.



(a) Self-employed



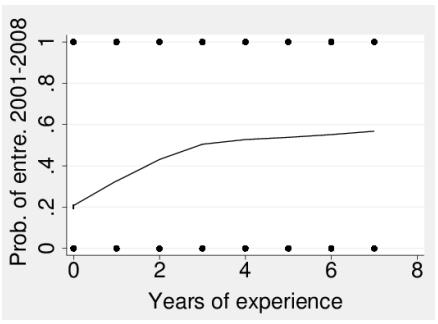
(b) Unpaid/family



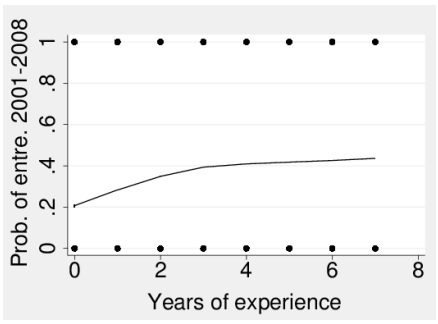
(c) Wage employees

Figure A.6: The relationship between years of experience and the propensity to engage in enterprise activity, 0-50th percentile of wealth distribution

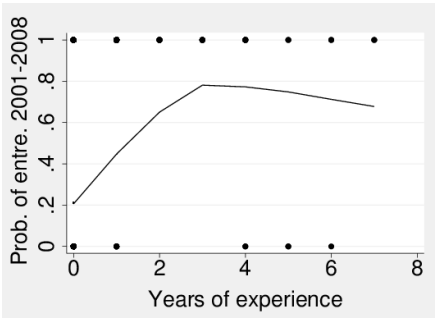
Note: In each sub-figure, the x-axis records years of experience of the household head in a given enterprise activity, while the y-axis records the propensity to engage in enterprise activity. Non-parametric graphs generated by a tri-cube smoother, which overweights local observations in placing fitted curve. In all cases a bandwidth of 0.8 is used. Self-employed refers to enterprises with no employees, Unpaid/family refers to enterprises with only unpaid/family employees, while Wage employees refers to enterprises that may have waged employee(s).



(a) Self-employed



(b) Unpaid/family



(c) Wage employees

Figure A.7: The relationship between years of experience and the propensity to engage in enterprise activity, 0-99th percentile of wealth distribution

Note: In each sub-figure, the x-axis records years of experience of the household head in a given enterprise activity, while the y-axis records the propensity to engage in enterprise activity. Non-parametric graphs generated by a tri-cube smoother, which overweights local observations in placing fitted curve. In all cases a bandwidth of 0.8 is used. Self-employed refers to enterprises with no employees, Unpaid/family refers to enterprises with only unpaid/family employees, while Wage employees refers to enterprises that may have waged employee(s).

A.3 Tables

Table A.1: Summary Statistics on IFLS Rounds

Survey Round	Year	Individuals	Households enterprises	Household	Communities
IFLS4	2007-08	44103 (50580)	13536	6186	313
IFLS3	2000	38433 (43649)	10435	5452	311
IFLS2	1997	22019 (33081)	7619	2625*	313
IFLS1	1993-94	22019 (33081)	7224	2439*	312
	Overall	66784	(unique)		

Table A.2: Geographic distribution of firms in the IFLS

Provinces	IFLS3 All	IFLS3 >= 5 Empl.	IFLS3 >= $1000	IFLS4 All	IFLS4 >= 5 Empl.	IFLS4 >= $1000
13 core provinces						
North Sumatra	344	27	44	374	39	78
West Sumatra	261	35	40	348	40	90
South Sumatra	234	26	18	235	23	50
Lampung	181	11	12	155	10	45
DKI Jakarta	603	46	79	507	33	126
West Java	1,029	55	103	779	54	137
Central Java	884	63	118	882	49	151
Yogyakarta	413	50	90	396	30	102
East Java	1,006	64	100	892	50	157
Bali	296	22	75	340	35	105
West Nusa Tenggara	456	29	18	303	25	46
South Kalimantan	288	11	28	269	11	50
South Sulawesi	308	18	38	311	30	75
Subtotal	6303	457	763	5791	429	1212
Tracking provinces						
Riau	52	3	6	51	1	16
Bangka-Belitung	45	5	7	50	5	17
Banten	204	8	25	126	13	24
West Kalimantan				1		
Central Kalimantan	1			3		2
East Kalimantan	11		5	9	3	5
North Sulawesi	1					
Subtotal	314	16	43	240	22	64
Total	6617	473	806	6031	451	1276

Note: The Table records counts of non-farm enterprises by province for the two most recent IFLS rounds, IFLS3 (2000) and IFLS4 (2007-08). I distinguish the 13 original survey target provinces from provinces that were added to the survey through tracking of household split-offs. For each round I report total numbers, along with the subset of firms with more than 5 employees (including the entrepreneur, summing over paid and unpaid employees), and firms with more than 1000 USD equivalent in reported business capital stock (all values converted to 2005 dollars).

Table A.3: Summary statistics on firms, 2008, firms with no employees

	Enterprises with no employees						
	N	**Mean**	**SD**	**P25**	**P50**	**P75**	**P95**
Bus owned by household	2711	1.0	0.1	1	1	1	1
Pct owned by household	41	38.2	22.1	25.0	50.0	50.0	50.0
Bus. operated out. home	2711	0.8	0.4	1	1	1	1
Applied for permit	2711	0.0	0.2	0	0	0	0
Permit issued	100	1.0	0.0	1	1	1	1
Cost obtain permit	100	13014.2	35310.1	2.7	11.9	78.4	1081000.0
Unpaid labor startup	2711	0.2	0.5	0.0	0.0	0.0	1.0
Wage labor startup	2711	0.1	1.6	0.0	0.0	0.0	0.0
Total labor startup	2711	1.3	1.7	1.0	1.0	1.0	2.0
Startup capital	2251	409.9	2816.6	10.8	54.1	216.3	1406.0
Current unpaid labor	2711	0.0	0.0	0.0	0.0	0.0	0.0
Current wage labor	2711	0.0	0.0	0.0	0.0	0.0	0.0
Current total labor	2711	0.9	0.2	1.0	1.0	1.0	1.0
Current land assets	2711	324.2	3720.3	0.0	0.0	0.0	216.3
Current building assets	2711	182.6	1329.4	0.0	0.0	0.0	432.6
Current 4-wheel vehicle	2711	122.4	885.3	0.0	0.0	0.0	324.5
Current other vehicles	2711	101.2	301.4	0.0	0.0	0.0	757.1
Curr. other non-farm eq.	2711	84.3	414.0	0.0	8.7	43.3	324.5
Current total capital	2711	814.7	4467.1	5.4	37.9	346.1	2379.4
Unpaid labor shutdown	151	0.6	0.9	0.0	0.0	1.0	2.0
Wage labor shutdown	151	0.3	1.1	0.0	0.0	0.0	2.0
Total labor shutdown	151	2.0	1.3	1.0	2.0	2.0	4.0
Net profit	2637	679.2	1958.3	146.0	389.4	778.7	1946.8
Total revenue	48	448.5	515.5	108.2	324.5	648.9	1081.6
Total expense	43	251.4	266.8	108.2	108.2	324.5	648.9
Ent. products consumed	2660	92.9	254.9	0.0	13.0	75.2	389.4
Ent returns used by HH	2649	448.6	681.3	86.5	259.6	584.0	1349.8
Ent returns left over	2643	119.9	727.4	0.0	0.0	54.1	540.8
Total procure. of goods	643	188.0	556.0	5.4	21.6	86.5	1081.6
Total sales	88	410.1	711.2	13.5	64.9	405.6	2163.1
Total shared profit	54	304.8	394.2	2.7	64.9	584.0	1092.4
Unit returns to capital (%)	2259	983.6	45179.7	0.9	4.5	20.6	133.3
Unit returns to labor (USD)	2496	693.8	2005.1	155.7	389.4	778.7	1952.2
Net ch. labor since start	2711	-0.2	1.7	0.0	0.0	0.0	0.0
Net ch. capital since start	2251	450.1	5100.0	-21.6	0.0	135.2	1676.4

Note: Monetary values converted to 2008 US dollars. Dummy variables have decimal values removed.
Note: The three enterprise categories are mutually exclusive; in 2008 there are 6186 firms reported by IFLS households.

Table A.4: Summary statistics on firms, 2008, firms with only family/unpaid employees

	Enterprises with family/unpaid employees						
	N	**Mean**	**SD**	**P25**	**P50**	**P75**	**P95**
Bus owned by household	2326	1.0	0.2	1	1	1	1
Pct owned by household	56	39.8	23.4	25.0	50.0	50.0	90.0
Bus. operated out. home	2326	0.7	0.5	0	1	1	1
Applied for permit	2326	0.1	0.2	0	0	0	1
Permit issued	135	1.0	0.7	1	1	1	1
Cost obtain permit	135	12853.3	35075.4	3.0	21.6	54.1	1081000.0
Unpaid labor startup	2326	1.3	0.9	1.0	1.0	2.0	3.0
Wage labor startup	2326	0.1	1.1	0.0	0.0	0.0	0.0
Total labor startup	2326	2.4	1.4	2.0	2.0	3.0	4.0
Startup capital	2091	432.1	3544.3	10.8	48.7	216.3	1297.9
Current unpaid labor	2326	1.5	0.9	1.0	1.0	2.0	3.0
Current wage labor	2326	0.0	0.0	0.0	0.0	0.0	0.0
Current total labor	2326	2.5	0.9	2.0	2.0	3.0	4.0
Current land assets	2326	234.3	1763.5	0.0	0.0	0.0	540.8
Current building assets	2326	298.4	1617.5	0.0	0.0	32.5	1081.6
Current 4-wheel vehicle	2326	120.1	831.2	0.0	0.0	0.0	324.5
Current other vehicles	2326	84.3	273.5	0.0	0.0	0.0	757.1
Curr. other non-farm eq.	2326	130.9	401.1	5.4	21.6	108.2	540.8
Current total capital	2326	867.9	3057.9	13.0	84.4	562.4	3785.4
Unpaid labor shutdown							
Wage labor shutdown							
Total labor shutdown							
Net profit	2270	739.8	1102.9	162.2	389.4	908.5	2258.3
Total revenue	45	428.3	468.6	108.2	324.5	648.9	1297.9
Total expense	43	278.1	384.9	108.2	108.2	324.5	1081.6
Ent. products consumed	2283	137.9	264.0	3.8	38.9	129.8	648.9
Ent returns used by HH	2283	457.7	652.3	97.3	259.6	594.9	1557.4
Ent returns left over	2282	119.8	551.5	0.0	0.0	64.9	540.8
Total procure. of goods	704	275.1	1618.1	5.4	21.6	108.2	973.4
Total sales	98	2304.0	11956.4	8.1	70.3	757.1	5191.4
Total shared profit	33	286.0	581.9	0.0	54.1	216.3	1622.3
Unit returns to capital (%)	2089	20.6	64.9	0.8	3.2	15.0	84.0
Unit returns to labor (USD)	2270	309.4	434.3	64.9	194.7	389.4	973.4
Net ch. labor since start	2326	0.1	1.4	0.0	0.0	0.0	2.0
Net ch. capital since start	2091	447.4	4192.6	-14.1	11.9	324.5	2920.2

Note: Monetary values converted to 2008 US dollars. Dummy variables have decimal values removed.
Note: The three enterprise categories are mutually exclusive; in 2008 there are 6186 firms reported by IFLS households.

Table A.5: Summary statistics on firms, 2008, firms with waged employees

	Enterprises with waged employees						
	N	**Mean**	**SD**	**P25**	**P50**	**P75**	**P95**
Bus owned by household	1149	0.9	0.2	1	1	1	1
Pct owned by household	72	44.6	17.8	33.0	50.0	50.0	75.0
Bus. operated out. home	1149	0.8	0.4	1	1	1	1
Applied for permit	1149	0.3	0.5	0	0	1	1
Permit issued	340	1.0	0.2	1	1	1	1
Cost obtain permit	340	8763.2	29239.7	10.8	54.1	216.3	1081000.0
Unpaid labor startup	1149	0.7	1.2	0.0	0.0	1.0	2.0
Wage labor startup	1149	2.2	6.3	0.0	1.0	2.0	6.0
Total labor startup	1149	3.8	6.4	2.0	3.0	4.0	9.0
Startup capital	1024	2021.4	7684.3	54.1	324.5	1622.3	6489.3
Current unpaid labor	1149	0.6	1.2	0.0	0.0	1.0	2.0
Current wage labor	1149	3.5	8.9	1.0	2.0	3.0	10.0
Current total labor	1149	5.1	8.9	3.0	3.0	5.0	12.0
Current land assets	1149	1519.9	8200.2	0.0	0.0	0.0	5407.7
Current building assets	1149	1534.9	7019.0	0.0	0.0	216.3	6489.3
Current 4-wheel vehicle	1149	1556.1	6032.4	0.0	0.0	0.0	8652.4
Current other vehicles	1149	277.3	768.7	0.0	0.0	32.5	1406.0
Curr. other non-farm eq.	1149	898.2	4066.3	10.8	108.2	540.8	3244.7
Current total capital	1149	5786.5	15666.2	119.0	1081.6	4326.2	26005.8
Unpaid labor shutdown							
Wage labor shutdown							
Total labor shutdown							
Net profit	1108	2749.0	6432.8	519.1	1297.9	2595.7	10123.3
Total revenue	30	1087.5	969.5	648.9	1081.6	1081.6	3893.6
Total expense	31	646.6	457.0	216.3	648.9	1081.6	1297.9
Ent. products consumed	1133	223.4	692.7	0.0	32.5	173.1	986.4
Ent returns used by HH	1129	1194.8	1838.5	259.6	648.9	1297.9	3893.6
Ent returns left over	1110	782.9	3263.5	0.0	64.9	540.8	2595.7
Total procure. of goods	457	1696.5	8365.4	27.0	108.2	540.8	6489.3
Total sales	72	4579.2	14121.5	39.2	200.1	2974.3	27038.7
Total shared profit	40	1442.8	4516.8	41.1	384.0	1081.6	3839.5
Unit returns to capital (%)	1037	13.2	100.7	0.3	1.0	4.5	60.0
Unit returns to labor (USD)	1108	614.3	1121.6	144.2	324.5	648.9	1946.8
Net ch. labor since start	1149	1.3	7.9	0.0	0.0	2.0	6.0
Net ch. capital since start	1024	3848.7	14910.6	-45.4	216.3	2109.0	22496.2

Note: Monetary values converted to 2008 US dollars. Dummy variables have decimal values removed.
Note: The three enterprise categories are mutually exclusive; in 2008 there are 6186 firms reported by IFLS households.

Table A.6: Summary statistics on enterprise experience

Individuals running enterprises with no employees

	N	Mean	SD	P25	Median	P75	P95
Exp. 1 curr. ent.	17622	6.01	3.88	3.5	5	8	13.5
Total exp. 1	19683	10.77	9.69	4	7	14	31
Total exp. 2	2389	7.36	6.32	5	6	6	19
Total exp. 3	91	5.39	5.11	3	5	6	8.5

Individual running enterprises with only household/unpaid empl.

	N	Mean	SD	P25	Median	P75	P95
Exp. 2 curr. ent.	14348	6.28	4.11	3.5	5	8.5	14
Total exp. 1	2915	7.09	6.36	4	6	7	17
Total exp. 2	15738	11.15	9.98	4	7	14	33
Total exp. 3	101	5.27	3.48	4	5	6	13

Individual running enterprises with waged employees

	N	Mean	SD	P25	Median	P75	P95
Exp. 3 curr. ent.	625	4.86	2.55	3	4.5	6	10.5
Total exp. 1	169	7.02	5.86	3	6	8.5	15
Total exp. 2	160	8.37	5.89	6	6	12	15.5
Total exp. 3	676	7.2	6.55	3.5	5	8	22

Note: Exp and exp refer to 'experience'. 'curr' refers to current enterprise–namely enterprise individual is running in a given survey year. 'Total' refers to total cumulative experience, whether in current or prior enterprise. Use enterprise category labels: 1=no employees, 2=only household/unpaid labor, 3=hiring outside labor.
Note: each observation is a person-year observation from the panel (1988-2008), where enterprise experience is recorded only if it is reported as the primary occupation.

Table A.7: Summary statistics on individuals who are self-employed in the panel

Individuals running enterprises with no employees

	N	Mean	SD	P25	Median	P75	P95
Age	34538	41.94	14.1	31	40	52	67
Education (years)	34538	2.01	4.57	0	0	0	12
Gender (male=1)	34538	0.59	0.49				
Marriage (married=1)	34538	0.57	0.49				
Net profit	34538	257999	322620	15	151057	389864	970874

Individual running enterprises with only household/unpaid empl.

	N	Mean	SD	P25	Median	P75	P95
Age	25961	45.11	13.9	34	44	55	69
Education (years)	25971	2.03	4.35	0	0	0	12
Gender (male=1)	25971	0.68	0.47				
Marriage (married=1)	25971	0.58	0.49				
Net profit	25971	238910	339253	6	88582	354610	984246

Individual running enterprises with waged employees

	N	Mean	SD	P25	Median	P75	P95
Age	1529	41.07	12.77	32	39	49	66
Education (years)	1534	4.75	6.28	0	0	12	19
Gender (male=1)	1534	0.7	0.46				
Marriage (married=1)	1534	0.65	0.48				
Net profit	1534	495473	482890	14	381723	831255	1461988

Note: The table only summarizes statistics from individual-year observations in which the individual is active in the given self-employment category. Monetary values are expressed in Indonesian Rupiah, deflated relative to 2005 value. In 2008 10,000 Rph. 1 USD. Net profits are monthly returns.

Table A.8: Variance-covariance matrix for key enterprise variables, 2008

	Enterprise family owned (0/1)	Operate outside home (0/1)	Have business permit (0/1)	H-hold labor startup	Wage labor startup	Total labor startup	Startup capital	H-hold labor current	Wage labor current	Total labor current	Capital value: land	Capital value: buildings	Capital value: 4-wheel vehicles	Capital value: other vehicles
Enterprise family-owned (0/1)	1.00													
Operate outside home (0/1)	-0.04	1.00												
Have business permit (0/1)	0.05	0.07	1.00											
H-hold labor startup	-0.15	-0.15	0.02	1.00										
Wage labor startup	0.00	0.12	-0.02	-0.12	1.00									
Total labor startup	-0.03	0.09	-0.02	0.08	0.98	1.00								
Startup capital	0.03	0.10	-0.01	-0.04	0.31	0.30	1.00							
H-hold labor current	-0.02	-0.08	0.00	0.61	-0.08	0.05	0.10	1.00						
Wage labor current	-0.04	0.02	-0.03	-0.10	0.53	0.51	0.17	-0.13	1.00					
Total labor current	-0.04	0.01	-0.03	-0.02	0.53	0.52	0.19	-0.01	0.99	1.00				
Capital value: land	-0.16	-0.01	-0.05	0.09	0.12	0.14	0.18	0.02	0.21	0.21	1.00			
Capital value: buildings	0.04	0.12	-0.01	0.00	0.24	0.24	0.37	0.03	0.16	0.16	0.44	1.00		
Capital value: 4-wheel vehicles	0.00	0.13	-0.01	-0.03	0.41	0.41	0.41	0.01	0.30	0.31	0.12	0.22	1.00	
Capital value: other vehicles	0.03	0.02	-0.03	-0.07	0.13	0.12	0.11	-0.06	0.16	0.16	-0.02	0.00	0.13	1.00
Capital value: Other farm equip	0.04	0.08	-0.01	-0.11	0.27	0.25	0.23	-0.10	0.27	0.26	0.11	0.28	0.00	0.06
Total current capital	-0.05	0.12	-0.04	-0.02	0.42	0.42	0.46	-0.03	0.40	0.40	0.70	0.68	0.53	0.11
Net profit	0.02	0.05	-0.02	0.11	0.44	0.47	0.50	0.03	0.49	0.50	0.58	0.49	0.49	0.13
Total product consumed	0.08	-0.09	-0.03	0.01	0.06	0.06	0.14	0.21	0.04	0.06	0.01	0.03	0.14	0.15
Money used by h-hold from ent.	0.00	0.10	-0.02	-0.01	0.31	0.31	0.08	0.00	0.54	0.54	0.40	0.12	0.12	0.16
Money left over from business	0.03	0.09	-0.01	0.08	0.40	0.42	0.55	0.06	0.48	0.49	0.53	0.56	0.53	0.14
Total procurement of goods	0.04	0.10	-0.01	-0.06	0.37	0.36	0.60	-0.04	0.31	0.31	0.30	0.58	0.51	0.10
Avg. return to capital	0.02	0.03	0.02	-0.06	-0.05	-0.06	-0.03	-0.07	-0.04	-0.05	-0.03	-0.04	-0.03	-0.04
Avg. return to labor	0.03	-0.01	-0.02	0.00	-0.04	-0.04	0.19	-0.08	-0.04	-0.05	0.44	0.21	0.07	0.01
Net change labor since startup	-0.02	-0.05	-0.03	-0.08	-0.10	-0.12	0.00	-0.04	0.78	0.78	0.15	0.02	0.06	0.10
Net change capital since startup	-0.08	0.05	-0.03	0.01	0.21	0.21	-0.29	-0.11	0.29	0.28	0.61	0.44	0.25	0.04
Urban location (0/1)	0.03	-0.11	-0.02	-0.04	0.02	0.01	-0.06	-0.10	-0.01	-0.02	0.03	-0.06	-0.04	-0.08
Java location (0/1)	0.04	-0.15	-0.03	-0.03	-0.07	-0.07	-0.01	-0.05	-0.05	-0.05	0.03	-0.07	0.02	-0.06
Age of household head	0.00	0.07	0.00	-0.02	-0.05	-0.05	-0.01	-0.03	-0.04	-0.04	-0.01	-0.01	-0.01	0.00

Table A.9: Variance-covariance matrix for key enterprise variables, 2008, cont.

	Capital value: Other farm equip	Total current capital	Net profit	Total product consumed	Money used by h-hold from ent.	Money left over from business	Total procurement of goods	Avg. return to capital	Avg. return to labor	Net change labor since startup	Net change capital since startup	Urban location (0/1)	Java location (0/1)	Age of household head
Enterprise family-owned (0/1)														
Operate outside home (0/1)														
Have business permit (0/1)														
H-hold labor startup														
Wage labor startup														
Total labor startup														
Startup capital														
H-hold labor current														
Wage labor current														
Total labor current														
Capital value: land														
Capital value: buildings														
Capital value: 4-wheel vehicles														
Capital value: other vehicles														
Capital value: Other farm equip	1.00													
Total current capital	0.57	1.00												
Net profit	0.14	0.68	1.00											
Total product consumed	-0.03	0.06	0.12	1.00										
Money used by h-hold from ent.	0.19	0.37	0.48	0.10	1.00									
Money left over from business	0.21	0.72	0.93	0.11	0.32	1.00								
Total procurement of goods	0.40	0.68	0.66	0.04	0.10	0.79	1.00							
Avg. return to capital	-0.03	-0.05	-0.02	-0.03	-0.02	-0.03	-0.02	1.00						
Avg. return to labor	0.12	0.35	0.56	0.12	0.31	0.45	0.26	0.03	1.00					
Net change labor since startup	0.12	0.16	0.24	0.03	0.41	0.27	0.10	-0.01	-0.02	1.00				
Net change capital since startup	0.44	0.72	0.34	-0.04	0.34	0.34	0.27	-0.03	0.23	0.18	1.00			
Urban location (0/1)	-0.01	-0.02	-0.05	0.04	0.02	-0.09	-0.11	-0.05	-0.02	-0.03	0.02	1.00		
Java location (0/1)	-0.08	-0.03	-0.05	0.07	-0.08	-0.05	-0.07	-0.04	-0.02	-0.01	-0.02	0.44	1.00	
Age of household head	-0.05	-0.04	-0.01	-0.04	-0.03	-0.01	0.00	0.00	0.00	-0.01	-0.03	0.13	-0.05	1.00

Table A.10: Variance-covariance matrix for key enterprise variables, 2000

	Enterprise family owned (0/1)	Operate outside home (0/1)	Have business permit (0/1)	H-hold labor startup	Wage labor startup	Total labor startup	Startup capital	H-hold labor current	Wage labor current	Total labor current	Capital value: land	Capital value: buildings	Capital value: 4-wheel vehicles	Capital value: other vehicles
Enterprise family-owned (0/1)	1.00													
Operate outside home (0/1)	-0.04	1.00												
Have business permit (0/1)	0.05	0.07	1.00											
H-hold labor startup	-0.15	-0.15	0.02	1.00										
Wage labor startup	0.00	0.12	-0.02	-0.12	1.00									
Total labor startup	-0.03	0.09	-0.02	0.08	0.98	1.00								
Startup capital	0.03	0.10	-0.01	-0.04	0.31	0.30	1.00							
H-hold labor current	-0.02	-0.08	0.00	0.61	-0.08	0.05	0.10	1.00						
Wage labor current	-0.04	0.02	-0.03	-0.10	0.53	0.51	0.17	-0.13	1.00					
Total labor current	-0.04	0.01	-0.03	-0.02	0.53	0.52	0.19	-0.01	0.99	1.00				
Capital value: land	-0.16	-0.01	-0.05	0.09	0.12	0.14	0.18	0.02	0.21	0.21	1.00			
Capital value: buildings	0.04	0.12	-0.01	0.00	0.24	0.24	0.37	0.03	0.16	0.16	0.44	1.00		
Capital value: 4-wheel vehicles	0.00	0.13	-0.01	-0.03	0.41	0.41	0.41	0.01	0.30	0.31	0.12	0.22	1.00	
Capital value: other vehicles	0.03	0.02	-0.03	-0.07	0.13	0.12	0.11	-0.06	0.16	0.16	-0.02	0.00	0.13	1.00
Capital value: Other farm equip	0.04	0.08	-0.01	-0.11	0.27	0.25	0.23	-0.10	0.27	0.26	0.11	0.28	0.00	0.06
Total current capital	-0.05	0.12	-0.04	-0.02	0.42	0.42	0.46	-0.03	0.40	0.40	0.70	0.68	0.53	0.11
Net profit	0.02	0.05	-0.02	0.11	0.44	0.47	0.50	0.03	0.49	0.50	0.58	0.49	0.49	0.13
Total product consumed	0.08	-0.09	-0.03	0.01	0.06	0.06	0.14	0.21	0.04	0.06	0.01	0.03	0.14	0.15
Money used by h-hold from ent.	0.00	0.10	-0.02	-0.01	0.31	0.31	0.08	0.00	0.54	0.54	0.40	0.12	0.12	0.16
Money left over from business	0.03	0.09	-0.01	0.08	0.40	0.42	0.55	0.06	0.48	0.49	0.53	0.56	0.53	0.14
Total procurement of goods	0.04	0.10	-0.01	-0.06	0.37	0.36	0.60	-0.04	0.31	0.31	0.30	0.58	0.51	0.10
Avg. return to capital	0.02	0.03	0.02	-0.06	-0.05	-0.06	-0.03	-0.07	-0.04	-0.05	-0.03	-0.04	-0.03	-0.04
Avg. return to labor	0.03	-0.01	-0.02	0.00	-0.04	-0.04	0.19	-0.08	-0.04	-0.05	0.44	0.21	0.07	0.01
Net change labor since startup	-0.02	-0.05	-0.03	-0.08	-0.10	-0.12	0.00	-0.04	0.78	0.78	0.15	0.02	0.06	0.10
Net change capital since startup	-0.08	0.05	-0.03	0.01	0.21	0.21	-0.29	-0.11	0.29	0.28	0.61	0.44	0.25	0.04
Urban location (0/1)	0.03	-0.11	-0.02	-0.04	0.02	0.01	-0.06	-0.10	-0.01	-0.02	0.03	-0.06	-0.04	-0.08
Java location (0/1)	0.04	-0.15	-0.03	-0.03	-0.07	-0.07	-0.01	-0.05	-0.05	-0.05	0.03	-0.07	0.02	-0.06
Age of household head	0.00	0.07	0.00	-0.02	-0.05	-0.05	-0.01	-0.03	-0.04	-0.04	-0.01	-0.01	-0.01	0.00

Table A.11: Variance-covariance matrix for key enterprise variables, 2000, cont.

	Capital value: Other farm equip	Total current capital	Net profit	Total product consumed	Money used by h-hold from ent.	Money left over from business	Total procurement of goods	Avg. return to capital	Avg. return to labor	Net change labor since startup	Net change capital since startup	Urban location (0/1)	Java location (0/1)	Age of household head
Enterprise family-owned (0/1)														
Operate outside home (0/1)														
Have business permit (0/1)														
H-hold labor startup														
Wage labor startup														
Total labor startup														
Startup capital														
H-hold labor current														
Wage labor current														
Total labor current														
Capital value: land														
Capital value: buildings														
Capital value: 4-wheel vehicles														
Capital value: other vehicles														
Capital value: Other farm equip	1.00													
Total current capital	0.57	1.00												
Net profit	0.14	0.68	1.00											
Total product consumed	-0.03	0.06	0.12	1.00										
Money used by h-hold from ent.	0.19	0.37	0.48	0.10	1.00									
Money left over from business	0.21	0.72	0.93	0.11	0.32	1.00								
Total procurement of goods	0.40	0.68	0.66	0.04	0.10	0.79	1.00							
Avg. return to capital	-0.03	-0.05	-0.02	-0.03	-0.02	-0.03	-0.02	1.00						
Avg. return to labor	0.12	0.35	0.56	0.12	0.31	0.45	0.26	0.03	1.00					
Net change labor since startup	0.12	0.16	0.24	0.03	0.41	0.27	0.10	-0.01	-0.02	1.00				
Net change capital since startup	0.44	0.72	0.34	-0.04	0.34	0.34	0.27	-0.03	0.23	0.18	1.00			
Urban location (0/1)	-0.01	-0.02	-0.05	0.04	0.02	-0.09	-0.11	-0.05	-0.02	-0.03	0.02	1.00		
Java location (0/1)	-0.08	-0.03	-0.05	0.07	-0.08	-0.05	-0.07	-0.04	-0.02	-0.01	-0.02	0.44	1.00	
Age of household head	-0.05	-0.04	-0.01	-0.04	-0.03	-0.01	0.00	0.00	0.00	-0.01	-0.03	0.13	-0.05	1.00

Table A.12: Variance-covariance matrix for household variables and enterprise activity

	Started first-ever business (0/1)	Started new business (0/1)	Self-employed (0/1)	Net business profit	Startup capital: ent. 1	Current capital: ent. 1	Household wealth (lag)	Household wealth (liquid)	Household wealth (illiquid)	Number of children	Number of adults (male)
Started first-ever business (0/1)	1.00										
Started new business (0/1)	0.80	1.00									
Self-employed (0/1)	0.81	0.99	1.00								
Net business profit	-0.04	-0.02	-0.02	1.00							
Startup capital: ent. 1	0.06	0.07	0.07	0.47	1.00						
Current capital: ent. 1	-0.05	-0.02	-0.02	0.50	0.37	1.00					
Household wealth (lag)	-0.03	0.01	0.01	0.21	0.17	0.32	1.00				
Household wealth (liquid)	-0.03	0.00	0.00	0.18	0.21	0.31	0.75	1.00			
Household wealth (illiquid)	-0.02	0.01	0.01	0.19	0.12	0.27	0.95	0.51	1.00		
Number of children	-0.05	-0.03	-0.03	0.03	0.00	0.02	-0.01	0.01	-0.02	1.00	
Number of adults (male)	-0.17	-0.10	-0.11	0.00	0.00	0.03	0.12	0.11	0.11	0.04	1.00
Number of adults (female)	-0.16	-0.10	-0.10	0.01	0.03	0.07	0.19	0.17	0.17	0.09	0.47
Age of household head	-0.04	-0.06	-0.06	-0.04	-0.02	-0.03	-0.06	-0.05	-0.06	0.00	0.05
Education of household head	0.16	0.14	0.14	0.13	0.13	0.16	0.20	0.15	0.19	-0.01	-0.18
Gender of household head	-0.01	-0.02	-0.03	0.03	0.02	0.01	0.04	0.02	0.04	0.06	0.14
Married (0/1)	-0.01	-0.01	-0.01	0.03	0.03	0.03	0.02	0.03	0.02	0.15	0.06
Head total years SE experience	-0.34	-0.33	-0.33	-0.03	-0.05	0.00	0.03	0.02	0.03	-0.01	0.28
Head exp ent no employees	0.16	0.18	0.18	-0.02	0.02	0.00	-0.04	-0.02	-0.04	-0.03	-0.09
Head exp ent hhold employees	0.11	0.14	0.13	0.00	0.00	0.01	0.02	0.04	0.00	0.00	0.04
Head exp ent wage employees	0.08	0.07	0.07	0.09	0.11	0.06	0.04	0.05	0.03	0.00	-0.01
Urban (0/1)	-0.09	-0.05	-0.05	0.00	-0.01	0.03	0.10	0.04	0.11	0.02	0.14
Java location (0/1)	-0.15	-0.12	-0.12	-0.05	-0.03	-0.03	0.01	-0.06	0.04	-0.06	0.10

Table A.13: Variance-covariance matrix for household variables and enterprise activity, cont.

	Number of adults (female)	Age of household head	Education of household head	Gender of household head	Married (0/1)	Head total years SE experience	Head exp ent no employees	Head exp ent hhold employees	Head exp ent wage employees	Urban (0/1)	Java location (0/1)
Started first-ever business (0/1)											
Started new business (0/1)											
Self-employed (0/1)											
Net business profit											
Startup capital: ent. 1											
Current capital: ent. 1											
Household wealth (lag)											
Household wealth (liquid)											
Household wealth (illiquid)											
Number of children											
Number of adults (male)											
Number of adults (female)	1.00										
Age of household head	0.06	1.00									
Education of household head	-0.14	-0.18	1.00								
Gender of household head	-0.04	-0.14	0.07	1.00							
Married (0/1)	0.02	-0.15	0.08	0.51	1.00						
Head total years SE experience	0.27	0.10	-0.21	0.00	-0.05	1.00					
Head exp ent no employees	-0.10	0.01	-0.02	0.00	0.00	-0.27	1.00				
Head exp ent hhold employees	-0.02	-0.03	-0.02	0.03	0.02	-0.20	-0.01	1.00			
Head exp ent wage employees	-0.03	0.01	0.07	0.02	0.02	-0.09	-0.01	-0.01	1.00		
Urban (0/1)	0.16	-0.01	0.00	0.00	0.00	0.04	-0.05	-0.07	-0.03	1.00	
Java location (0/1)	0.07	0.04	-0.17	-0.06	-0.04	0.16	-0.08	-0.07	-0.03	0.36	1.00

APPENDIX B
APPENDIX TO CHAPTER 3

B.1 Figures

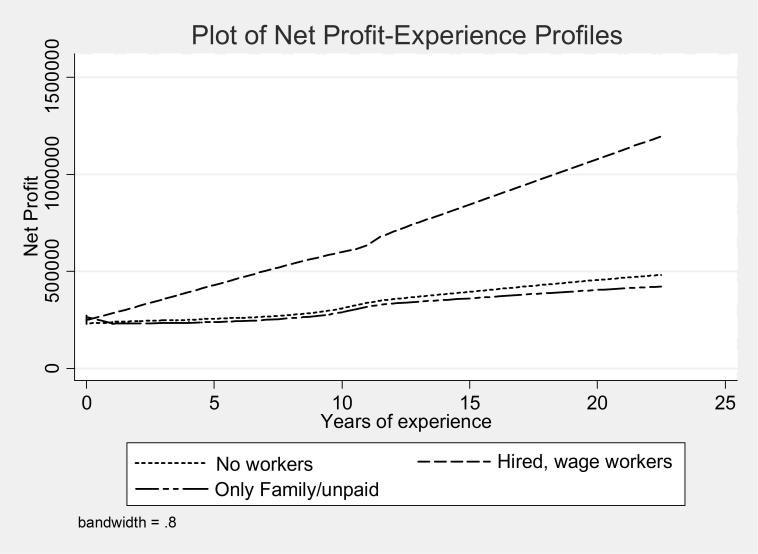
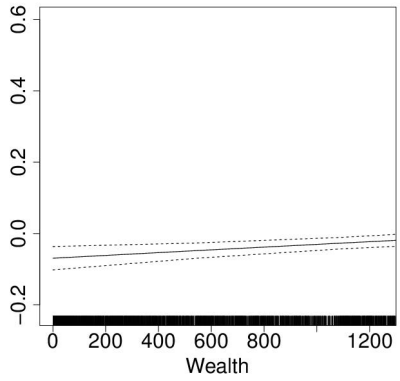
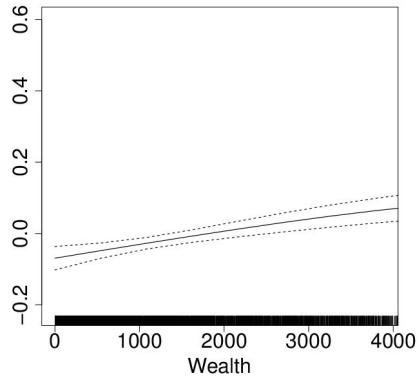


Figure B.1: Plot of net profit against experience in three enterprise types

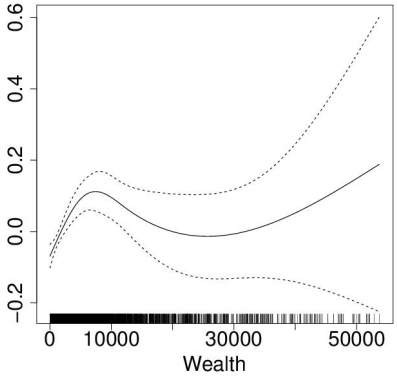
Note: The Figure records net profits of enterprises that startup in the three employment categories ((i) no employees, (ii) only family/unpaid workers, (iii) having waged employees) against years of experience of the individual entrepreneur running the enterprise.



(a) 0-50th percentile



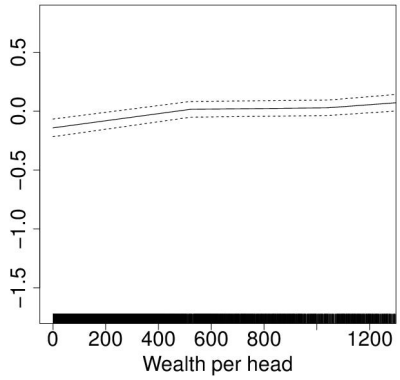
(b) 0-75th percentile



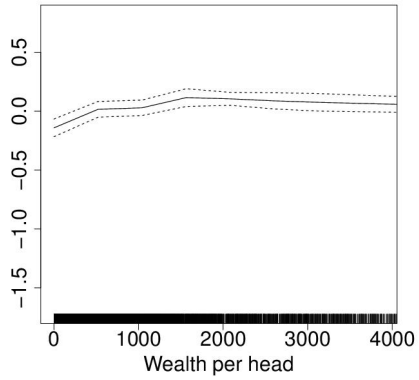
(c) Full distribution

Figure B.2: The relationship between (lagged) wealth and the propensity to engage in enterprise activity

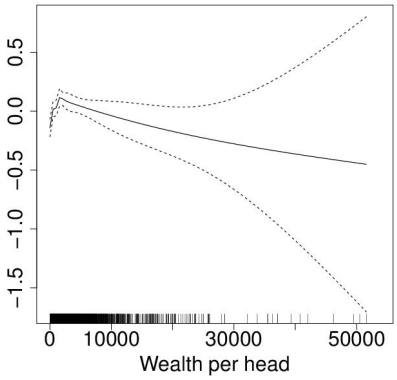
Note: In each figure the x-axis (household wealth) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of the influence of wealth on the propensity to engage in enterprise activity.



(a) 0-50th percentile



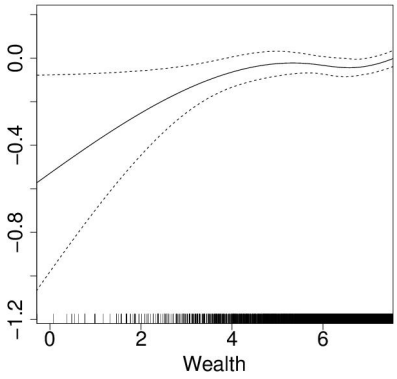
(b) 0-75th percentile



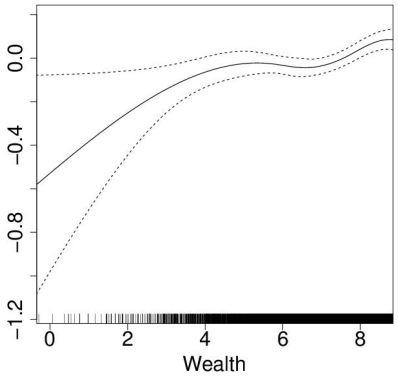
(c) Full distribution

Figure B.3: The relationship between (lagged) wealth per head and the propensity to engage in enterprise activity

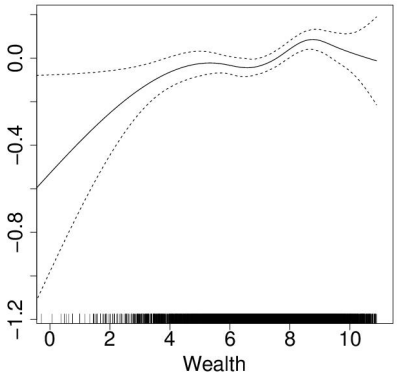
Note: In each figure the x-axis (household wealth per head, using total adult labor) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of the influence of wealth per head on the propensity to engage in enterprise activity.



(a) 0-50th percentile



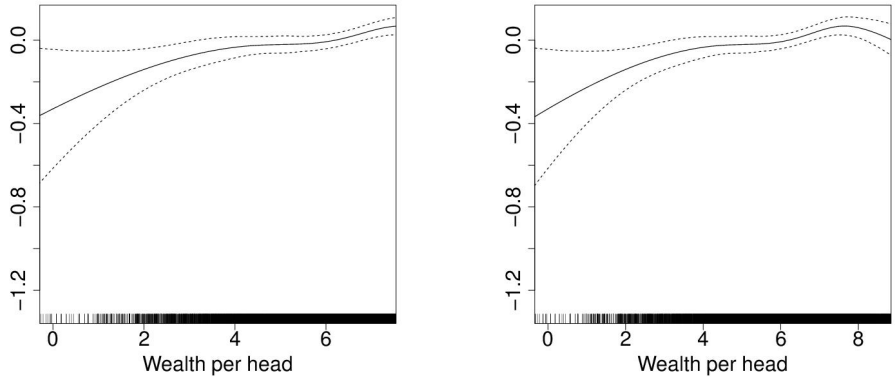
(b) 0-75th percentile



(c) Full distribution

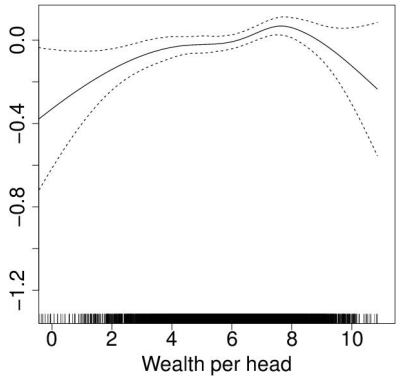
Figure B.4: The relationship between (lagged) wealth in logs and the propensity to engage in enterprise activity

Note: In each figure the x-axis (household wealth) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of the influence of wealth on the propensity to engage in enterprise activity.



(a) 0-50th percentile

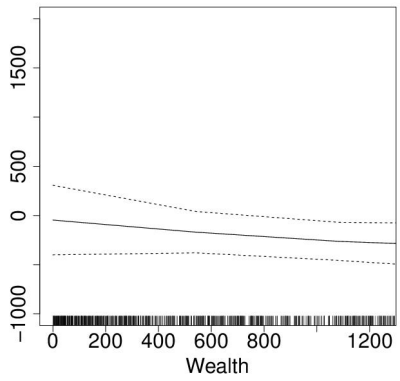
(b) 0-75th percentile



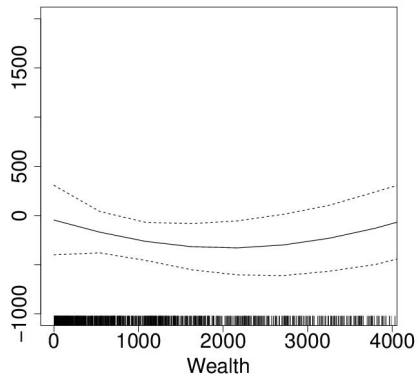
(c) Full distribution

Figure B.5: The relationship between (lagged) wealth per head in logs and the propensity to engage in enterprise activity

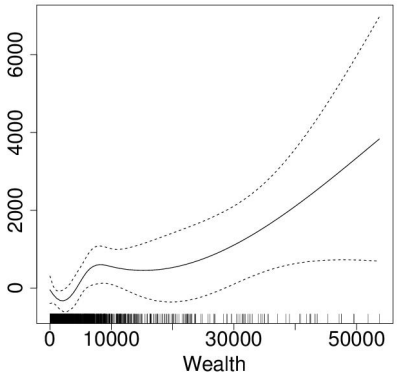
Note: In each figure the x-axis (household wealth per head, using total adult labor) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of the influence of wealth per head on the propensity to engage in enterprise activity.



(a) 0-50th percentile



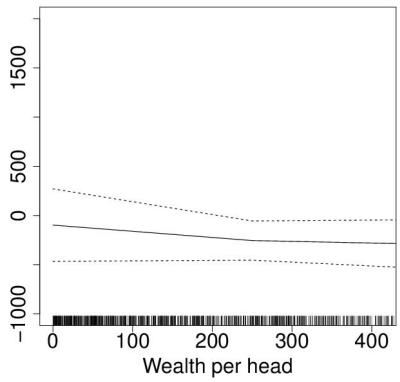
(b) 0-75th percentile



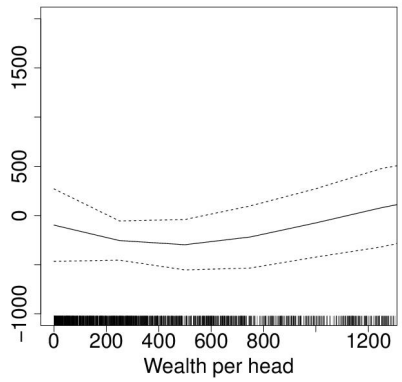
(c) Full distribution

Figure B.6: The relationship between (lagged) wealth and startup capital

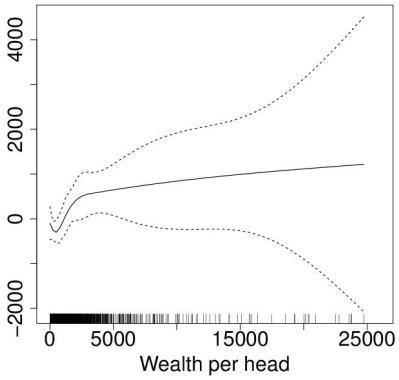
Note: In each figure the x-axis (household wealth) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of dollars, again, representing startup capital.



(a) 0-50th percentile



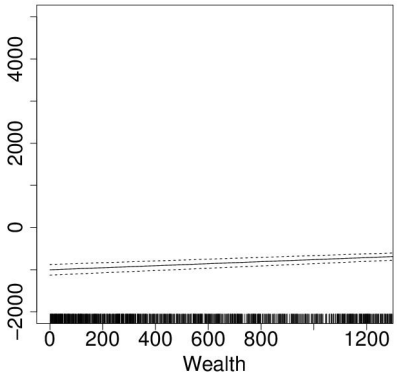
(b) 0-75th percentile



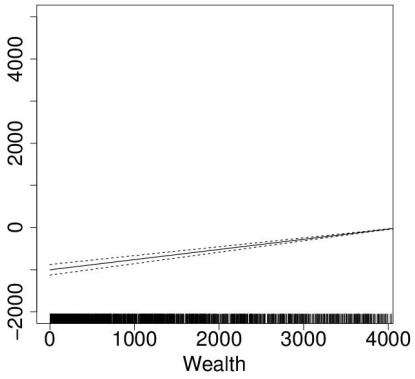
(c) Full distribution

Figure B.7: The relationship between (lagged) wealth per head and startup capital

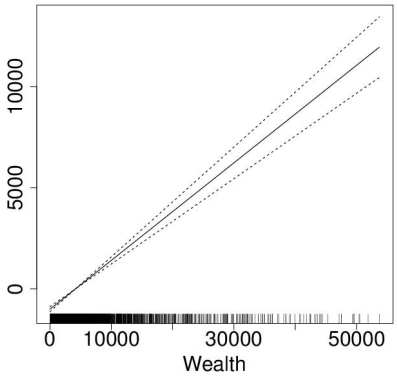
Note: In each figure the x-axis (household wealth per head, using total adult labor) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of dollars, again, representing startup capital.



(a) 0-50th percentile



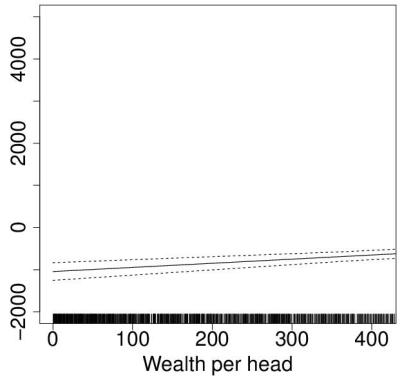
(b) 0-75th percentile



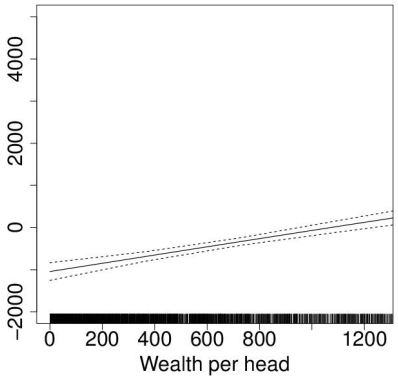
(c) Full distribution

Figure B.8: The relationship between (lagged) wealth and current capital

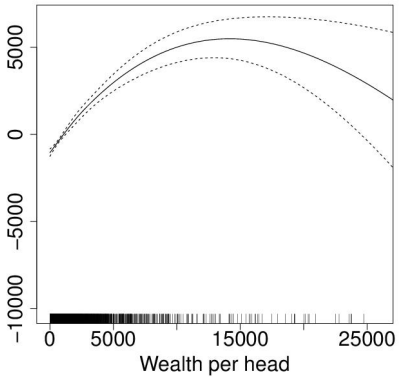
Note: In each figure the x-axis (household wealth) is scaled using year 2000 US dollars, while the y-axis is scaled in terms of dollars, again, representing current capital.



(a) 0-50th percentile



(b) 0-75th percentile



(c) Full distribution

Figure B.9: The relationship between (lagged) wealth per head and current capital

Note: In each figure the x-axis (household wealth per head, using total adult labor) is scaled in terms of year 2000 US dollars, while the y-axis is scaled in terms of dollars, again, representing current capital.

B.2 Tables

Table B.1: Percentiles of wage- and self-employment returns, IFLS, 2008

	10th	**25th**	**50th**	**90th**	**95th**	**99th**
Net Profit	0	6	32	162	270	540
Wage	10	21	48	162	216	302

Note: Values converted to 2008 USD terms.

Table B.2: Wealth and propensity to start enterprise

	Estimate	Std. Error	z value	Pr	
Intercept	-1.4289	0.1045	-13.67	0.0000	***
Age Head	0.0171	0.0043	3.97	0.0001	***
Age Head^2	-0.0002	0.0000	-4.62	0.0000	***
Education head	0.0106	0.0023	4.65	0.0000	***
Number adult females	0.0132	0.0115	1.14	0.2538	
Number adult males	0.0221	0.0110	2.01	0.0448	*
Number children	0.0638	0.0097	6.57	0.0000	***
Bank BRI	0.0147	0.0664	0.22	0.8253	
Ppl Credit Bank	-0.0630	0.0437	-1.44	0.1493	
Village Credit Union	0.0323	0.0388	0.83	0.4057	
Village Unit Coop.	-0.0368	0.0374	-0.98	0.3258	
Other Formal Coop.	-0.0617	0.0347	-1.78	0.0752	.
State Priv. Bank	-0.0234	0.0459	-0.51	0.6109	
BMT	0.0193	0.0426	0.45	0.6499	
Urban dummy	0.2430	0.0353	6.89	0.0000	***
Province dummies	YES				
Observations	13422				
R-squared (adj.)	0.0199				
UBRE	0.0173				
Significance of smooth terms:					
edf	248.00				
Ref. df.	3727.00				
Chi. Sq.	23.86				
p	0.0001				***

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.3: Wealth per head and propensity to start enterprise

	Estimate	**Std. Error**	**z value**	**P**	
Intercept	-1.2520	0.1097	-11.41	<2e-16	***
Age Head	0.0100	0.0045	2.23	0.0259	*
Age Head^2	-0.0001	0.0000	-2.96	0.0030	**
Education head	0.0106	0.0023	4.63	0.0000	***
Number adult females	0.0185	0.0115	1.61	0.1074	
Number adult males	0.0269	0.0111	2.42	0.0156	*
Number children	0.0648	0.0098	6.65	0.0000	***
Bank BRI	0.0122	0.0664	0.18	0.8544	
Ppl Credit Bank	-0.0654	0.0437	-1.50	0.1343	
Village Credit Union	0.0359	0.0389	0.92	0.3558	
Village Unit Coop.	-0.0415	0.0374	-1.11	0.2676	
Other Formal Coop.	-0.0645	0.0347	-1.86	0.0631	.
State Priv. Bank	-0.0233	0.0459	-0.51	0.6119	
BMT	0.0188	0.0426	0.44	0.6586	
Urban dummy	0.2519	0.0353	7.14	0.0000	***
Province dummies	YES				
Observations	13246				
R-squared (adj.)	0.0188				
UBRE	0.0255				
Significance of smooth terms:					
edf	7.77				
Ref. df.	8.32				
Chi. Sq.	29.36				
p	0.0003				***

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.4: Wealth (in logs) and propensity to start enterprise

	Estimate	Std. Error	z value	Pr	
Intercept	-1.4377	0.1062	-13.54	<2e-16	***
Age Head	0.0182	0.0044	4.13	0.0000	***
Age Head^2	-0.0002	0.0000	-4.82	0.0000	***
Education head	0.0106	0.0023	4.59	0.0000	***
Number adult females	0.0179	0.0118	1.52	0.1286	
Number adult males	0.0235	0.0113	2.08	0.0375	*
Number children	0.0590	0.0099	5.99	0.0000	***
Bank BRI	0.0084	0.0670	0.13	0.8999	
Ppl Credit Bank	-0.0612	0.0442	-1.39	0.1656	
Village Credit Union	0.0293	0.0391	0.75	0.4543	
Village Unit Coop.	-0.0442	0.0377	-1.17	0.2406	
Other Formal Coop.	-0.0593	0.0350	-1.70	0.0896	.
State Priv. Bank	-0.0245	0.0462	-0.53	0.5953	
BMT	0.0219	0.0429	0.51	0.6095	
Urban dummy	0.2445	0.0356	6.87	0.0000	***
Province dummies	YES				
Observations	12983				
R-squared (adj.)	0.0194				
UBRE	0.0250				
Significance of smooth terms:					
edf	4.5720				
Ref. df.	5.4470				
Chi. Sq.	20.6700				
p	0.0000				**

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.5: Wealth per head (in logs) and propensity to start enterprise

	Estimate	**Std. Error**	**z value**	**P**	
Intercept	-1.2660	0.1115	-11.36	<2e-16	***
Age Head	0.0113	0.0046	2.46	0.0138	*
Age Head^2	-0.0001	0.0000	-3.22	0.0013	**
Education head	0.0107	0.0023	4.60	0.0000	***
Number adult females	0.0221	0.0117	1.89	0.0584	.
Number adult males	0.0279	0.0114	2.45	0.0142	*
Number children	0.0606	0.0099	6.13	0.0000	***
Bank BRI	0.0061	0.0669	0.09	0.9273	
Ppl Credit Bank	-0.0626	0.0441	-1.42	0.1559	
Village Credit Union	0.0321	0.0391	0.82	0.4118	
Village Unit Coop.	-0.0480	0.0377	-1.27	0.2033	
Other Formal Coop.	-0.0615	0.0350	-1.76	0.0786	.
State Priv. Bank	-0.0247	0.0462	-0.53	0.5930	
BMT	0.0213	0.0429	0.50	0.6193	
Urban dummy	0.2523	0.0355	7.11	0.0000	***
Province dummies	YES				
Observations	12809				
R-squared (adj.)	0.0179				
UBRE	0.0334				
Significance of smooth terms:					
edf	4.2390				
Ref. df.	5.1370				
Chi. Sq.	17.2900				
p	0.0044				**

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.6: Wealth and startup capital

	Estimate	Std. Error	t value	Pr(>t)	
Intercept	1244.1296	936.7108	1.33	0.1843	
Age Head	-36.4415	38.0083	-0.96	0.3378	
Age Head^2	0.2464	0.3409	0.72	0.4698	
Education head	64.8418	18.3574	3.53	0.0004	***
Number adult females	113.7766	99.6981	1.14	0.2539	
Number adult males	234.2174	99.4955	2.35	0.0187	*
Number children	-50.7645	84.0007	-0.60	0.5457	
Bank BRI	-43.9789	547.5146	-0.08	0.9360	
Ppl Credit Bank	-521.2357	358.5536	-1.45	0.1462	
Village Credit Union	-55.4262	314.6766	-0.18	0.8602	
Village Unit Coop.	-325.8137	304.7241	-1.07	0.2851	
Other Formal Coop.	107.3569	283.3204	0.38	0.7048	
State Priv. Bank	-397.8533	373.3023	-1.07	0.2866	
BMT	396.8833	340.5205	1.17	0.2439	
Urban dummy	149.1028	282.2210	0.53	0.5973	
Province dummies	YES				
Observations	2356				
R-squared (adj.)	0.0170				
GVC-score	23684000				
Significance of smooth terms:					
edf	4.0010				
Ref. df.	4.5170				
F	3.7080				
p	0.0035				**

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.7: Wealth per head and startup capital

	Estimate	Std. Error	t value	Pr(>t)	
Intercept	983.8943	937.2849	1.05	0.2940	
Age Head	-33.3477	38.1955	-0.87	0.3827	
Age Head^2	0.2223	0.3428	0.65	0.5166	
Education head	64.5312	18.4332	3.50	0.0005	***
Number adult females	158.6860	99.6565	1.59	0.1114	
Number adult males	286.3323	99.9409	2.87	0.0042	**
Number children	-50.7850	84.1816	-0.60	0.5464	
Bank BRI	-9.9225	549.0813	-0.02	0.9856	
Ppl Credit Bank	-551.0652	358.7398	-1.54	0.1246	
Village Credit Union	-50.6642	315.0150	-0.16	0.8722	
Village Unit Coop.	-329.3952	305.3299	-1.08	0.2808	
Other Formal Coop.	109.5967	283.9963	0.39	0.6996	
State Priv. Bank	-421.4492	373.8881	-1.13	0.2598	
BMT	402.6586	340.6742	1.18	0.2373	
Urban dummy	171.4895	282.3445	0.61	0.5437	
Province dummies	YES				
Observations	2353				
R-squared (adj.)	0.0138				
GVC-score	23785000				
Significance of smooth terms:					
edf	4.2590				
Ref. df.	4.7670				
F	2.0420				
p	0.0733				.

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.8: Wealth and current capital

	Estimate	Std. Error	t value	Pr(>t)	
Intercept	320.5724	905.2269	0.35	0.7233	
Age Head	32.6240	35.1632	0.93	0.3536	
Age Head^2	-0.2662	0.3082	-0.86	0.3878	
Education head	89.2555	17.0866	5.22	0.0000	***
Number adult females	16.3150	86.1882	0.19	0.8499	
Number adult males	67.7213	81.3886	0.83	0.4054	
Number children	175.8264	71.6203	2.46	0.0141	*
Bank BRI	-948.7519	489.7230	-1.94	0.0528	.
Ppl Credit Bank	-156.9668	316.5865	-0.50	0.6201	
Village Credit Union	-360.6211	272.7296	-1.32	0.1861	
Village Unit Coop.	-13.7800	267.9794	-0.05	0.9590	
Other Formal Coop.	188.0882	247.6476	0.76	0.4476	
State Priv. Bank	151.3047	329.4934	0.46	0.6461	
BMT	-21.5436	294.7214	-0.07	0.9417	
Urban dummy	70.2687	246.8394	0.29	0.7759	
Province dummies	YES				
Observations	4870				
R-squared (adj.)	0.0763				
GVC-score	40635000				
Significance of smooth terms:					
edf	0.9999				
Ref. df.	1.0000				
F	254.20				
p	0.0000				***

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.9: Wealth per head and current capital

	Estimate	Std. Error	t value	Pr(>t)	
Intercept	-759.7484	913.6841	-0.83	0.4057	
Age Head	42.4202	35.5974	1.19	0.2335	
Age Head^2	-0.3407	0.3118	-1.09	0.2746	
Education head	88.9570	17.4200	5.11	0.0000	***
Number adult females	238.1930	86.5871	2.75	0.0060	**
Number adult males	259.2176	82.6373	3.14	0.0017	**
Number children	162.0976	72.1922	2.25	0.0248	*
Bank BRI	-969.2448	493.6695	-1.96	0.0497	*
Ppl Credit Bank	-229.7799	319.0855	-0.72	0.4715	
Village Credit Union	-330.8110	274.9154	-1.20	0.2289	
Village Unit Coop.	-23.7473	270.2024	-0.09	0.9300	
Other Formal Coop.	205.5699	249.6711	0.82	0.4103	
State Priv. Bank	146.5955	332.2187	0.44	0.6590	
BMT	4.7869	297.1347	0.02	0.9872	
Urban dummy	116.9098	248.8866	0.47	0.6386	
Province dummies	YES				
Observations	4866				
R-squared (adj.)	0.0623				
GVC-score	41296000				
Significance of smooth terms:					
edf	2.4710				
Ref. df.	2.8380				
F	61.2100				
p	0.0000				***

*** p <0.001, ** p <0.01, * p <0.05, . p <0.1

Table B.10: Probit: Effects of income flows on post-2000 startup

VARIABLES	(1)	(2) M. effects	(3)	(4) M. effects	(5)	(6) M. effects
Gov. trans. pre-2000	-0.2325	-0.0602	-0.1478	-0.0401	-0.1716	-0.0458
	(0.4236)	(0.0979)	(0.4289)	(0.1087)	(0.4345)	(0.1070)
Non-gov. trans. pre-2000	0.5468	0.1868	0.5674	0.1955	0.4536	0.1516
	(0.5333)	(0.2055)	(0.5368)	(0.2084)	(0.5400)	(0.2022)
Uncond. cash trans. pre-2000	1.4615**	0.5335**	1.4475*	0.5292**	1.4450*	0.5282**
	(0.7447)	(0.2411)	(0.7551)	(0.2454)	(0.7677)	(0.2509)
Exog. trans. pre-2000	-0.1297*	-0.0354*	-0.1110	-0.0307	-0.1792**	-0.0480**
	(0.0774)	(0.0200)	(0.0780)	(0.0207)	(0.0788)	(0.0196)
Gov. trans. pre-2000*wealth	0.0003	0.0001	0.0003	0.0001	0.0002	0.0001
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
Gov. trans. pre-2000*wealth^2	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-gov. trans. pre-2000*wealth	-0.0004	-0.0001	-0.0004	-0.0001	-0.0004	-0.0001
	(0.0010)	(0.0003)	(0.0011)	(0.0003)	(0.0011)	(0.0003)
Non-gov. trans. pre-2000*wealth^2	-0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
UC. cash trans. pre-2000*wealth	-0.0013	-0.0004	-0.0013	-0.0004	-0.0014	-0.0004
	(0.0011)	(0.0003)	(0.0011)	(0.0003)	(0.0011)	(0.0003)
UC. cash trans. pre-2000*wealth^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. pre-2000*wealth	0.0000**	0.0000**	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. pre-2000*wealth^2	-0.0000*	-0.0000*	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000			0.0000***	0.0000***	0.0000*	0.0000*
			(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000^2			-0.0000**	-0.0000**	-0.0000	-0.0000
			(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age hhold head			0.0177***	0.0051***	0.0185***	0.0053***
			(0.0060)	(0.0017)	(0.0060)	(0.0017)
Age hhold head^2			-0.0002***	-0.0001***	-0.0002***	-0.0001***
			(0.0001)	(0.0000)	(0.0001)	(0.0000)
Education hhold head			0.0054***	0.0016***	0.0063***	0.0018***
			(0.0019)	(0.0006)	(0.0020)	(0.0006)
Java dummy					0.0070	0.0020
					(0.0470)	(0.0135)
Urban dummy					0.2382***	0.0708***
					(0.0328)	(0.0101)
Location effects					YES	YES
Constant	-0.8126***		-0.9063***		-0.9978***	
	(0.0124)		(0.0339)		(0.0384)	
Observations	13818	13818	13536	13536	13536	13536
Chi-squared	17.287	17.287	113.61	113.61	230.67	230.67

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.11: Probit: Effects of income flows on post-2000 startup, Part 1

VARIABLES	(1)	(2) M. effects	(3)	(4) M. effects	(5)	(6) M. effects
Gov. trans. pre-2000	-0.2297	-0.0588	-0.1462	-0.0392	-0.1705	-0.0450
	(0.4226)	(0.0966)	(0.4273)	(0.1072)	(0.4336)	(0.1056)
Non-gov. trans. pre-2000	0.5244	0.1769	0.5419	0.1844	0.4253	0.1400
	(0.5383)	(0.2053)	(0.5422)	(0.2084)	(0.5470)	(0.2017)
Uncond. cash trans. pre-2000	1.4782**	0.5383**	1.4623*	0.5336**	1.4595*	0.5324**
	(0.7441)	(0.2404)	(0.7540)	(0.2447)	(0.7668)	(0.2503)
Exog. trans. pre-2000	-0.1447*	-0.0388**	-0.1276	-0.0347*	-0.1946**	-0.0512***
	(0.0781)	(0.0197)	(0.0787)	(0.0203)	(0.0795)	(0.0192)
Gov. trans. pre-2000*wealth	0.0003	0.0001	0.0003	0.0001	0.0002	0.0001
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
Gov. trans. pre-2000*wealth^2	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-gov. trans. pre-2000*wealth	-0.0003	-0.0001	-0.0003	-0.0001	-0.0004	-0.0001
	(0.0010)	(0.0003)	(0.0011)	(0.0003)	(0.0011)	(0.0003)
Non-gov. trans. pre-2000*wealth^2	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
UC. cash trans. pre-2000*wealth	-0.0013	-0.0004	-0.0013	-0.0004	-0.0014	-0.0004
	(0.0011)	(0.0003)	(0.0011)	(0.0003)	(0.0011)	(0.0003)
UC. cash trans. pre-2000*wealth^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. pre-2000*wealth	0.0000**	0.0000**	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. pre-2000*wealth^2	-0.0000*	-0.0000*	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Gov. trans. '07-'08	0.7750	0.2742	0.8336	0.2985	0.7934	0.2819
	(0.5441)	(0.2160)	(0.5521)	(0.2196)	(0.5565)	(0.2212)
Non-gov. trans. '07-'08	0.2330*	0.0720*	0.2391*	0.0745*	0.2474**	0.0769*
	(0.1248)	(0.0416)	(0.1253)	(0.0421)	(0.1260)	(0.0424)
Uncond. cash trans. '07-'08	-0.8970	-0.1634***	-0.8619	-0.1617***	-0.8155	-0.1558**
	(0.5689)	(0.0531)	(0.5769)	(0.0579)	(0.5811)	(0.0622)
Exog. trans. '07-'08	0.3054**	0.0966**	0.2696*	0.0848*	0.2510*	0.0782*
	(0.1386)	(0.0481)	(0.1390)	(0.0475)	(0.1392)	(0.0469)
Gov. trans. '07-'08*wealth	-0.0003	-0.0001	-0.0003	-0.0001	-0.0002	-0.0001
	(0.0007)	(0.0002)	(0.0008)	(0.0002)	(0.0008)	(0.0002)
Non-gov. trans. '07-'08*wealth	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Uncond. cash trans. '07-'08*wealth	0.0003	0.0001	0.0003	0.0001	0.0002	0.0001
	(0.0007)	(0.0002)	(0.0008)	(0.0002)	(0.0008)	(0.0002)
Exog. trans. '07-'08*wealth	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Gov. trans. '07-'08*wealth^2	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-gov. trans. '07-'08*wealth^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Uncond. cash trans. '07-'08*wealth^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. '07-'08*wealth^2	-0.0000	-0.0000	0.0000	0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000			0.0000***	0.0000***	0.0000*	0.0000*
			(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000^2			-0.0000**	-0.0000**	-0.0000	-0.0000
			(0.0000)	(0.0000)	(0.0000)	(0.0000)

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.12: Probit: Effects of income flows on post-2000 startup, Part 2

Age hhold head			YES	YES	YES	YES
Education hhold head			YES	YES	YES	YES
Java dummy			YES	YES	YES	YES
Urban dummy			YES	YES	YES	YES
Location			YES	YES	YES	YES
Constant	-0.8321***		-0.9218***		-1.0141***	
	(0.0127)		(0.0342)		(0.0387)	
Observations	13818	13818	13536	13536	13536	13536
Chi-squared	42.698	42.698	129.49	129.49	240.85	240.85

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.13: Probit: Effects of income flows on post-2000 startup

VARIABLES	(3)	(4) M.effect	(5)	(6) M.effect
Exog. trans. ent. 1	1.1637***	0.4234***	1.1738***	0.4266***
	(0.3078)	(0.1161)	(0.3067)	(0.1156)
Exog. trans. ent. 2	-1.2155***	-0.1994***	-1.2858***	-0.2025***
	(0.3005)	(0.0206)	(0.2990)	(0.0184)
Exog. trans. ent. 1*wealth	-0.0001**	-0.0000**	-0.0001**	-0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. ent. 1*wealth^2	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. ent. 2*wealth	0.0001**	0.0000**	0.0001**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. ent. 2*wealth^2	-0.0000*	-0.0000*	-0.0000*	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. '07-'08	0.2676*	0.0840*	0.2489*	0.0774*
	(0.1391)	(0.0474)	(0.1393)	(0.0469)
Exog. trans. '07-'08*wealth	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Exog. trans. '07-'08*wealth^2	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000	0.0000***	0.0000***	0.0000**	0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Wealth 2000^2	-0.0000**	-0.0000**	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age hhold head	0.0164***	0.0047***	0.0173***	0.0049***
	(0.0060)	(0.0017)	(0.0061)	(0.0017)
Age hhold head^2	-0.0002***	-0.0001***	-0.0002***	-0.0001***
	(0.0001)	(0.0000)	(0.0001)	(0.0000)
Education hhold head	0.0051***	0.0014***	0.0060***	0.0017***
	(0.0019)	(0.0006)	(0.0020)	(0.0006)
Java dummy			0.0199	0.0056
			(0.0472)	(0.0134)
Urban dummy			0.2337***	0.0687***
			(0.0329)	(0.0100)
Location dummy			YES	YES
Constant	-0.9134***		-1.0078***	
	(0.0340)		(0.0385)	
Observations	13536	13536	13536	13536
Chi-squared	124.04	124.04	238.04	238.04

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.14: Probit: Exogenous shocks

	(3)	(4)	(5)	(6)
VARIABLES		M. effects		M. effects
Exog. trans. ent. 1	1.1871***	0.4338***	1.1983***	0.4373***
	-0.3074	-0.1147	-0.3062	-0.1141
Exog. trans. ent. 2	-1.2361***	-0.2045***	-1.3029***	-0.2071***
	-0.3000	-0.0205	-0.2984	-0.0183
Exog. trans. ent. 1*wealth	-0.0001**	-0.0000**	-0.0001**	-0.0000**
	0.0000	0.0000	0.0000	0.0000
Exog. trans. ent. 1*wealth^2	0.0000	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0000	0.0000
Exog. trans. ent. 2*wealth	0.0001***	0.0000***	0.0001**	0.0000**
	0.0000	0.0000	0.0000	0.0000
Exog. trans. ent. 2*wealth^2	-0.0000**	-0.0000**	-0.0000*	-0.0000*
	0.0000	0.0000	0.0000	0.0000
Age hhold head	0.0179***	0.0052***	0.0184***	0.0053***
	-0.0060	-0.0017	-0.0060	-0.0017
Age hhold head^2	-0.0002***	-0.0001***	-0.0002***	-0.0001***
	-0.0001	0.0000	-0.0001	0.0000
Education hhold head	0.0063***	0.0018***	0.0068***	0.0020***
	-0.0019	-0.0005	-0.0019	-0.0006
Java dummy			0.0058	0.0017
			-0.0470	-0.0135
Urban dummy			0.2457***	0.0731***
			-0.0326	-0.0100
Location controls			YES	YES
Observations	13536	13536	13536	13536
Chi-squared	115.53	115.53	240.25	240.25

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.15: Effects of income flows on post-2000 startup capital

VARIABLES	(1)	(2)	(3)
Gov. trans. pre-2000	-569.0904	93.7502	111.3354
	(1,845.1878)	(1,824.3425)	(1,829.4691)
Non-gov. trans. pre-2000	-539.7361	-179.6838	-136.5810
	(2,083.1740)	(2,058.5838)	(2,062.8880)
Uncond. cash trans. pre-2000	-51.5571	-403.0299	-334.5148
	(4,166.1678)	(4,116.3869)	(4,123.6399)
Exog. trans. pre-2000	-204.0281	108.0891	164.0060
	(381.9870)	(379.1201)	(382.3471)
Gov. trans. pre-2000*wealth	0.0946	-0.2305	-0.1529
	(1.2003)	(1.1866)	(1.1917)
Gov. trans. pre-2000*wealth^2	-0.0000	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)
Non-gov. trans. pre-2000*wealth	-0.0699	-0.0270	0.0550
	(2.3172)	(2.2893)	(2.2921)
Non-gov. trans. pre-2000*wealth^2	0.0000	-0.0000	-0.0000
	(0.0001)	(0.0001)	(0.0001)
Uncond. cash trans. pre-2000*wealth	0.0230	0.0714	-0.1943
	(8.9067)	(8.7987)	(8.8070)
Uncond. cash trans. pre-2000*wealth^2	-0.0000	0.0000	0.0001
	(0.0026)	(0.0026)	(0.0026)
Exog. trans. pre-2000*wealth	0.0420	-0.0273	-0.0288
	(0.0406)	(0.0411)	(0.0411)
Exog. trans. pre-2000*wealth^2	-0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Wealth 2000		0.0621***	0.0639***
		(0.0092)	(0.0093)
Wealth 2000^2		-0.0000***	-0.0000***
		(0.0000)	(0.0000)
Age hhold head		24.9057	19.2492
		(29.0486)	(29.2669)
Age hhold head^2		-0.2509	-0.1929
		(0.2879)	(0.2901)
Education hhold head		49.8499***	47.5468***
		(9.9066)	(10.0689)
Java dummy			-475.1471
			(313.5889)
Urban dummy			-12.3024
			(150.4702)
Constant	649.3994***	19.3900	236.7383
	(60.9343)	(165.3383)	(190.8061)
Location			YES
Observations	4303	4303	4303
R-squared	0.001	0.026	0.029

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.16: Effects of exogenous income flows on post-2000 startup capital

VARIABLES	(1)	(2)	(3)
Exog. trans. ent. 1	119.4913	164.0438	169.0928
	(493.6262)	(491.0313)	(493.6417)
Exog. trans. ent. 1*wealth	0.0190	0.0117	0.0102
	(0.0498)	(0.0496)	(0.0497)
Exog. trans. ent. 1*wealth^2	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Age hhold head		29.3749	21.9457
		(29.1925)	(29.4181)
Age hhold head^2		-0.3029	-0.2261
		(0.2893)	(0.2916)
Education hhold head		63.8283***	60.9842***
		(9.7786)	(9.9629)
Java dummy			-449.6798
			(315.4871)
Urban dummy			87.3300
			(150.2264)
Constant	648.3066***	207.0042	433.2426**
	(60.0066)	(163.7399)	(189.7490)
Location			YES
Observations	4303	4303	4303
R-squared	0.000	0.012	0.014

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.17: Effects of income flows on post-2000 startup capital, quantile reg.

	(1)	(2)	(3)
VARIABLES	25th perc.	50th perc.	75th perc.
Gov. trans. pre-2000	-1.5658	13.2634	9.9258
	(10.2957)	(34.8987)	(102.0795)
Non-gov. trans. pre-2000	11.4076	21.0190	110.7548
	(15.9063)	(38.8980)	(162.8232)
Uncond. cash trans. pre-2000	9.6019	-12.1742	-38.8697
	(13.9381)	(39.4937)	(147.5160)
Exog. trans. pre-2000	13.7657***	2.1961	53.4300
	(3.0836)	(8.1149)	(34.7475)
Gov. trans. pre-2000*wealth	0.0213***	0.0050	-0.0227
	(0.0050)	(0.0218)	(0.0633)
Gov. trans. pre-2000*wealth^2	-0.0000***	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Non-gov. trans. pre-2000*wealth	-0.0108	-0.0266	-0.1139
	(0.0114)	(0.0464)	(0.1086)
Non-gov. trans. pre-2000*wealth^2	0.0000***	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Uncond. cash trans. pre-2000*wealth	-0.0741***	0.0308	0.3698
	(0.0246)	(0.1342)	(0.2652)
Uncond. cash trans. pre-2000*wealth^2	0.0000***	-0.0000	-0.0001*
	(0.0000)	(0.0000)	(0.0001)
Exog. trans. pre-2000*wealth	-0.0049***	-0.0014	-0.0146***
	(0.0003)	(0.0008)	(0.0035)
Exog. trans. pre-2000*wealth^2	0.0000***	0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Wealth 2000	0.0023***	0.0108***	0.0445***
	(0.0001)	(0.0002)	(0.0009)
Wealth 2000^2	-0.0000***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)
Age hhold head	0.6856***	2.0373***	9.3668***
	(0.2353)	(0.6306)	(2.7296)
Age hhold head^2	-0.0072***	-0.0213***	-0.0968***
	(0.0023)	(0.0063)	(0.0271)
Education hhold head	1.2517***	6.0077***	22.5216***
	(0.0784)	(0.2176)	(0.9710)
Java dummy	-0.2643	-14.5499**	-72.4292***
	(2.3598)	(6.3679)	(27.5023)
Urban dummy	-3.8289***	-10.4580***	-31.0491**
	(1.2016)	(3.2462)	(14.0529)
Location	YES	YES	YES
Constant	6.5214***	25.4176***	115.5045***
	(1.5399)	(4.1154)	(17.4912)
Observations	4303	4303	4303

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.18: Effects of exogenous income flows on post-2000 startup capital, quantile reg.

	(1)	(2)	(3)
VARIABLES	25th perc.	50th perc.	75th perc.
Exog. trans. ent. 1	13.0740***	17.0518	152.9946***
	(3.5903)	(12.5829)	(43.5792)
Exog. trans. ent. 1*wealth	0.0003	0.0071***	0.0089**
	(0.0003)	(0.0012)	(0.0040)
Exog. trans. ent. 1*wealth^2	0.0000***	0.0000***	0.0000*
	(0.0000)	(0.0000)	(0.0000)
Age hhold head	1.0730***	3.7174***	13.7805***
	(0.2070)	(0.7593)	(2.7751)
Age hhold head^2	-0.0113***	-0.0388***	-0.1453***
	(0.0021)	(0.0075)	(0.0275)
Education hhold head	1.3760***	8.3490***	31.0752***
	(0.0675)	(0.2566)	(0.9755)
Java dummy	-2.5170	-20.4398**	-62.3638**
	(2.2252)	(8.1385)	(29.7314)
Urban dummy	-3.1795***	-4.7920	-8.2275
	(1.0586)	(3.8702)	(13.9139)
Constant	9.2668***	33.5599***	189.3299***
	(1.3536)	(4.9023)	(17.6483)
Location	YES	YES	YES
Observations	4303	4303	4303

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.19: Effects from Unconditional Cash Transfers (UCT), without selection controls

	Without Selection Controls			
VARIABLES	Coef.	Std. Err.	z	P >|z|
Constant	-0.9971	0.0383	-26.05	0.000
Wealth 2000 USD (/million)	3.2517	1.9703	1.65	0.099
Wealth 2000 USD^2(/*million*)	0.0000	0.0000	-1.62	0.104
Age hhold head	0.0180	0.0060	2.98	0.003
Age hhold head^2	-0.0002	0.0001	-3.55	0.000
Education hhold head	0.0065	0.0020	3.31	0.001
Java dummy	0.0084	0.0470	0.18	0.859
Urban dummy	0.2352	0.0327	7.20	0.000
Province dummy	YES			
Variables controlling for selection into UCT				
Yearly rent on house (*1000)				
House electricity (dummy)				
Health card (dummy)				
Health program				
Ent. Exp. Hhold Head				
Number of children				
Number of adult female				
Number of adult male				
Water source				
Toilet type				
Fuel source				
Household assets				
Household head occupation				
Observations	13536			
Chi-sq.	213.64			
Prob >Chi-sq.	0.0000			
Log likelihood	-6902.01			
Pseudo R^2	0.02			

Table B.20: Effects from Unconditional Cash Transfers (UCT), with selection controls

	With Selection Controls			
VARIABLES	Coef.	Std. Err.	z	P >|z|
Constant	4.5100	175.6875	0.03	0.980
Wealth 2000 USD (/million)	3.5538	3.0702	1.16	0.247
Wealth 2000 USD^2(/*million*)	0.0000	0.0000	-1.07	0.283
Age hhold head	-0.0011	0.0083	-0.14	0.890
Age hhold head^2	0.0000	0.0001	-0.13	0.900
Education hhold head	0.0025	0.0032	0.77	0.439
Java dummy	-0.0875	0.0661	-1.32	0.186
Urban dummy	0.2531	0.0465	5.44	0.000
Province dummy	YES			
Variables controlling for selection into UCT				
Yearly rent on house (*1000)	0.0000	0.0000	2.08	0.037
House electricity (dummy)	0.1569	0.1021	1.54	0.124
Health card (dummy)	0.0123	0.0483	0.25	0.799
Health program	0.0927	0.0588	1.58	0.115
Ent. Exp. Hhold Head	-0.0254	0.0016	-15.67	0.000
Number of children	0.0396	0.0138	2.86	0.004
Number of adult female	0.0416	0.0176	2.37	0.018
Number of adult male	0.0385	0.0165	2.34	0.019
Water source	YES			
Toilet type	YES			
Fuel source	YES			
Household assets	YES			
Household head occupation	YES			
Observations	8003			
Chi-sq.	1189.55			
Prob >Chi-sq.	0.0000			
Log likelihood	-3851.24			
Pseudo R^2	0.13			

Table B.21: Parent Cross Tab

	0	1	2	3	Total
0	9,286	898	682	419	11,285
	82.29	7.96	6.04	3.71	100
	90.52	85.93	86.55	84.99	89.68
1	515	75	62	44	696
	73.99	10.78	8.91	6.32	100
	5.02	7.18	7.87	8.92	5.53
2	394	64	36	28	522
	75.48	12.26	6.9	5.36	100
	3.84	6.12	4.57	5.68	4.15
3	63	8	8	2	81
	77.78	9.88	9.88	2.47	100
	0.61	0.77	1.02	0.41	0.64
Total	10,258	1,045	788	493	12,584
	81.52	8.3	6.26	3.92	100
	100	100	100	100	100

Note: Values in first column are enterprise father ran (0 is father was not self-employed). Values in top row same, for mother. Within each box, first value is total number of children falling under observation, second is cross-tab frequency for father, third is cross-tab frequency for mother.

Table B.22: Parent Effect

VARIABLES	(1)	(2) M. effects
Father self-empl no employees	0.1677**	0.0328**
	(0.0688)	(0.0146)
Father self-empl hhold/unpaid employees	0.2891***	0.0602***
	(0.0630)	(0.0150)
Father self-empl waged employees	0.2429	0.0501
	(0.1922)	(0.0452)
Mother self-empl no employees	0.2050***	0.0409**
	(0.0725)	(0.0160)
Mother self-empl hhold/unpaid employees	0.2162***	0.0435***
	(0.0751)	(0.0168)
Mother self-empl waged employees	0.0362	0.0066
	(0.3181)	(0.0593)
Both parents self-employed	-0.1563	-0.0254*
	(0.1030)	(0.0152)
Constant	-1.3120***	
	(0.0171)	
Observations	12584	12584

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table B.23: Parent Return Effect

VARIABLES	(1)
Father self-empl no employees	-10.1411
	(14.6023)
Father self-empl hhold/unpaid employees	-19.4735
	(12.7695)
Father self-empl waged employees	127.3198***
	(43.6096)
Mother self-empl no employees	3.5335
	(15.9017)
Mother self-empl hhold/unpaid employees	9.0751
	(15.3417)
Mother self-empl waged employees	-12.1912
	(68.6272)
Both parents self-employed	-13.1895
	(21.6882)
Constant	78.8867***
	(3.8770)
Observations	1253
R-squared	0.011
F	1.9758

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

APPENDIX C
APPENDIX TO CHAPTER 4

C.1 Tables

Table C.1: Summary statistics on individual entrants

1998 Entrants	N	Mean	SD	P25	Median	P75	P95	P99
Age	684	33.81	12.78	25	30	40	60	73
Marriage (married=1)	684	0.87	0.34	1	1	1	1	1
Gender (male=1)	684	0.71	0.46	0	1	1	1	1
Education (years)	684	5.58	6.04	0	0.5	12	15	19

1999 Entrants	N	Mean	SD	P25	Median	P75	P95	P99
Age	1355	27.67	10.72	20	25	33	49	59
Marriage (married=1)	1355	0.87	0.33	1	1	1	1	1
Gender (male=1)	1355	0.68	0.46	0	1	1	1	1
Education (years)	1355	7.14	6.15	0	9	12	19	19

Table C.2: Summary statistics on community-level sources of variation

Year 1998 Variable	N	Mean	SD	P25	Median	P75	P95	P99
Avg. change formal 1998	573	-0.35	1.03	-1.00	-0.75	0.00	2.00	3.00
Growth employment 1998	631	1.09	2.23	-0.25	0.33	1.75	5.50	9.00
Comm unemployment rate 1998	625	0.13	0.18	0.00	0.00	0.22	0.50	0.73
Year 1999 Variable	**N**	**Mean**	**SD**	**P25**	**Median**	**P75**	**P95**	**P99**
Avg. change formal 1999	526	0.03	0.19	0.00	0.00	0.00	0.50	1.00
Growth employment 1999	829	0.12	0.23	0.00	0.00	0.17	0.50	1.00
Comm unemployment rate 1999	1245	0.29	0.33	0.00	0.17	0.50	1.00	1.00

Table C.3: Panel wage regression

VARIABLES	(1)
Age	-729.6043
	(970.2945)
Age ^{2}	219.4695***
	(11.0449)
Education (years)	17,743.6683***
	(328.4670)
Marriage (married=1)	233,807.5485***
	(5,274.8745)
1989 year dummy	-15,202.2139**
	(6,129.3431)
1990 year dummy	-29,255.7173***
	(6,002.8286)
1991 year dummy	-43,050.8011***
	(5,913.6615)
1992 year dummy	-7,592.2594
	(5,578.9516)
1993 year dummy	308,289.2601***
	(12,759.3060)
1994 year dummy	471,856.2062***
	(5,852.7692)
1995 year dummy	423,124.5305***
	(5,578.6388)
1996 year dummy	365,138.9674***
	(5,215.6677)
1997 year dummy	334,528.8544***
	(5,050.2212)
1998 year dummy	151,353.3406***
	(4,807.7177)
1999 year dummy	142,859.3440***
	(4,393.3809)
2000 year dummy	156,652.8447***
	(4,177.0906)
Constant	-152094.1954***
	(20,436.4606)
Observations	75843
R-squared	0.490
R-squared adjusted	0.3515
F	3583.5

Note: Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: Observations restricted to 4.6 Rph. <= wage <= 1872075 Rph.

Table C.4: Change in propensity to stay in self-employment after Asian Financial Crisis, 1998 entrants

Using 1998 New Entrants Into Self-Employment

Using Year-by-Year Choice Estimates

Year	Obs.	Mean	Std. Dev.	t-test	
1999	477	0.37	0.27	29.75	***
2000	477	0.33	0.28	26.19	***
2001	477	0.22	0.21	23.18	***
2002	477	0.22	0.23	20.76	***
2003	477	0.22	0.23	20.71	***
2004	477	0.24	0.23	22.89	***
2005	477	0.24	0.22	23.98	***
2006	477	0.24	0.22	23.23	***
2007	477	0.26	0.25	22.44	***
2008	477	0.32	0.25	27.67	***

Using Pooled Choice Estimates

Year	Obs.	Mean	Std. Dev.	t-test	
1999	477	0.34	0.26	28.64	***
2000	477	0.36	0.26	30.07	***

Using 1999, 2000 Choice Estimates

Year	Obs.	Mean	Std. Dev.	t-test	
1999	477	0.37	0.27	29.90	***
2000	477	0.33	0.28	26.22	***

Using Pooled Estimates Pooled Across 1999-2008

Year	Obs.	Mean	Std. Dev.	t-test	
1999-2008	954	0.35	0.26	154.16	***

Using 1999-2000 Pooled Across 1999-2000

Year	Obs.	Mean	Std. Dev.	t-test	
1999-2000	954	0.35	0.27	39.58	***

Note: Reports on the mean change in propensity to remain in self-employment for those who enter self-employment during 1998, with various methods of measuring propensity to persist and couterfactual. t-tests reported testing for differences in means of distributions.
*** p <0.01, ** p <0.05, * p <0.1

Table C.5: Change in propensity to stay in self-employment after Asian Financial Crisis, 1999 entrants

Using 1999 New Entrants Into Self-Employment

Using Year-by-Year Choice Estimates

Year	Obs.	Mean	Std. Dev.	t-test	
2000	514	0.90	0.11	191.95	***
2001	514	0.42	0.15	63.44	***
2002	514	0.44	0.17	57.48	***
2003	514	0.45	0.16	62.64	***
2004	514	0.48	0.17	64.91	***
2005	514	0.52	0.17	68.29	***
2006	514	0.57	0.18	73.41	***
2007	514	0.66	0.18	83.30	***
2008	514	0.75	0.16	108.56	***

Using Pooled Choice Estimates

Year	Obs.	Mean	Std. Dev.	t-test	
2000	514	0.89	0.09	217.80	***

Using Pooled Estimates Pooled Across 2000-2008

Year	Obs.	Mean	Std. Dev.	t-test	
2000-2008	514	0.89	0.09	307.73	***

Note: Reports on the mean change in propensity to remain in self-employment for those who enter self-employment during 1999, with various methods of measuring propensity to persist and couterfactual. t-tests reported testing for differences in means of distributions.
*** p <0.01, ** p <0.05, * p <0.1

Table C.6: Returns to experience, individuals entering self-employment during 1998

	Coefficient		Bootstrap Std. Err.	z	P >|z|
Exp. 1 curr. ent.	122053.6	**	60520.25	2.02	0.04
Exp. 1 curr. ent. ^{2}	-29698.09	*	16522.57	-1.80	0.07
Exp. 2 curr. ent.	111070.5		82077.02	1.35	0.18
Exp. 2 curr. ent. ^{2}	-19199.98		24561.68	-0.78	0.43
Exp. 3 curr. ent.	351829.4	*	190108.40	1.85	0.06
Exp. 3 curr. ent. ^{2}	-66646.76		52523.57	-1.27	0.20
Age	14752.11		11359.57	1.30	0.19
Age ^{2}	-176.0679		134.49	-1.31	0.19
Gender	319052.6	***	53910.98	5.92	0.00
Education (years)	19810.85	***	3808.41	5.20	0.00
Marriage (married=1)	-1841.24		75149.90	-0.02	0.98
Selection correction, 1998 entry	-137276.9		116344.20	-1.18	0.24
Constant	-17804.76		220855.60	-0.08	0.94
Observations	1220				
Bootstrap replications	50				
Wald Chi-sq.	110.86				
Prob >Chi-sq.	0				
R-squared	0.0564				

*** p <0.01, ** p <0.05, * p <0.1
Note: Estimation by OLS with bootstraped standard errors due to projected regressor (selection correction).

Table C.7: Returns to Experience, Individuals Entering Self-Employment During 1999

	Coefficient		Std. Err.	t	P >|t|
Exp. 1 curr. ent.	-158.00	***	55.99	-2.82	0.005
Exp. 1 curr. ent. ^{2}	31.31	**	14.55	2.15	0.032
Exp. 2 curr. ent.	316.40	**	158.59	2	0.046
Exp. 2 curr. ent. ^{2}	-198.14	**	89.32	-2.22	0.027
Exp. 3 curr. ent.	1623.77	***	532.92	3.05	0.002
Exp. 3 curr. ent. ^{2}	-651.30	**	299.85	-2.17	0.03
Age	5.76		39.35	0.15	0.884
Age ^{2}	-0.17		0.36	-0.47	0.64
Gender	145.40	**	67.60	2.15	0.032
Education (years)	26.05	**	11.24	2.32	0.021
Marriage (married=1)	-24.97		56.97	-0.44	0.661
Selection correction, 1999 entry	-1213.65		1974.65	-0.61	0.539
Constant	431.39		1113.93	0.39	0.699
Observations	599				
F	9.13				
Prob >Chi-sq.	0.00				
R-squared	0.1575				
Adj R-squared	0.1402				

*** p <0.01, ** p <0.05, * p <0.1
Note: Estimation by OLS with bootstraped standard errors due to projected regressor (selection correction).

APPENDIX D
APPENDIX TO CHAPTER 5

D.1 Additional Material

D.1.1 Evidence for labor market frictions in EHC transmission

Here I provide evidence that, if working in family enterprises is an important source of entrepreneurship-relevant human capital, that labor market frictions preventing people from outside the family from working in a household enterprise could inhibit such human capital transfer. The idea that labor market frictions and weak contracting make household members much more likely to work in household enterprises is not a new one in development economics, with the fundamental hypothesis being known as the "separation hypothesis".[1] Benjamin [Benjamin, 1992] is a seminal paper providing a test of the separation hypothesis, with the basic idea being that if the household faces no frictions in hiring outside labor, then the size of a household enterprise should not be determined in any way by available household labor. The vast majority of applications of such tests have been in the context of agricultural enterprises; to my knowledge this is one of the first applications of such a test in the context of non-farm enterprises.

The results are summarized in Table D.1. Indeed, implementing a simple version of the test I find that the separation hypothesis is violated. The number of adult males, adult females, and children in the household (as measures

[1]The term "separation" is used to capture the question of whether we can think of a household enterprise as being a separate, profit-maximizing entity apart from the household, or whether enterprise-specific decisions are affected by other household characteristics and choices, such as labor supply and consumption.

of household labor supply) are statistically significant predictors of household non-farm enterprise size, reflected in enterprise employment, after controlling for the age of the household head and location. We see that each additional male household member is correlated with an extra 0.09 labor units in the enterprise, with the corresponding figures being 0.05 for women and 0.07 for children. These figures are economically significant given that the average enterprise has less than 2 additional employees beyond the entrepreneur.[2]

These results suggest that if exposure to enterprise activity is important for acquiring entrepreneurship-specific human capital, labor market frictions might inhibit such transfers.

D.1.2 Proofs: The properties of the value function

I verify the claim in Proposition 5.2.3. The stated claim is that the value function $V : S \to \mathbb{R}$ exists and is unique, and furthermore, that V is bounded and continuous, and a stationary optimal policy $\zeta^* : S \to a$ exists. First I define a couple of objects. First, the state space, S, which contains vectors of the form (e_i, W_i, x_i, ξ), is defined as follows: $S = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{Z}_+ \times \mathbb{R}_+$. Second, the action set is defined as

$$A = \{0, 1\} \times \mathbb{R}_+ \times [0, \lambda(e_i, W_i)] \times [0, w(x_i) + W_i] \times [0, R(w, r; e_i, W_i) + W_i], \quad \text{(D.1)}$$

where the first component refers to the discrete choice of occupation, the second and third components the (potential) labor and capital allocations, respectively,

[2] In columns 3 & 4 of Table **??**, I show that if I just run total waged labor as the dependent variable (rather than all sources of labor), only male household labor is a statistically significant predictor, perhaps suggesting that male household labor is most important for monitoring. This point is also supported by the results in column 4 of the table, in which outside, hired labor is only a function of the household's own adult male labor supply.

the fourth component the savings decision in wage employment, and finally the fifth component the savings decision in self-employment. Individual action profiles are denoted by $a \in A$.

The basic assumptions necessary for the proposition to hold are provided in Bhattacharya and Majumdar (2007), pp. 388-390. Hence I must verify the following conditions:

1. S is a (nonempty) Borel subset of a complete, separable metric space. Clearly S is a subset of the complete, separable metric space $\mathbb{R}^4_+$.

2. A is a compact metric space. Since all of the objects at hand are in Euclidean spaces, I invoke the Bolzano-Weierstrass theorem, so that it is sufficient to show that A is closed and bounded. It is immediately apparent that for given values of (e_i, W_i, x_i, ξ), A is a cross product of closed and bounded sets, and hence is closed and bounded itself.

3. U is a bounded, continuous function on $S \times A$. U is bounded and continuous by assumption.

4. If $s_n \to s$, $a_n \to a$, then $q(.|s_n, a_n)$ converges weakly to $q(.|s, a)$, where we take q to denote the distribution of ξ. Again, this is trivial – since ξ is assumed to follow an i.i.d. distribution with finite mean and variance any sequence $q(.|s_n, a_n)$ on the state space will heap on $q(.|s, a)$; i.e., $q(.|s_n, a_n) = q(.|s, a)$ for any n, so the convergence result holds.

Next, I verify the claim in Proposition 5.2.3**.** Recall the statement, that for e_i, $\underline{e}$, $1/\sigma^2_\xi$, and δ sufficiently large, ℓ'' sufficiently large on $\left(0, \underline{e}\right)$, and $\lambda(e_i, W_i)$ sufficiently small, there exists an optimal policy ζ^* under which $s^* > 0$ in all states, with s^* increasing in e_i, and there exists a ratio of W_i/e_i such that $e'/e \gg 1$, whereas when such a ratio is not satisfied $e'/e \approx 1$.

Recall the Bellman equation provided in equation (5.6),

$$
\begin{aligned}
& V\left(e_i, x_i, W_i, \xi\right) \qquad\qquad (D.2)\\
= \; & \max\left\{ \max_{s\in[0,w(x_i)+W_i]} U\left[W_i + w\left(x_i\right) - s\right] + \delta EV\left(e_i, x_i+1, s, \xi'\right), \right.\\
& \left. \max_{\substack{s\in[0,R+W_i],\ l\geq 0,\\ k\in[0,\lambda(e_i,W_i)/r]}} U\left[W_i + R\left(k,l;w,r;e_i,W_i\right) - s\right] + \delta EV\left(\ell\left(c\left(k,l\right)/\left(1+e_i\right)\right), x_i, s, \xi'\right) \right\},
\end{aligned}
$$

First, notice that for e_i and $1/\sigma_\xi^2$ sufficiently large, the individual will be incentivized to select the discrete, self-employment occupation in each period. Given the i.i.d. shock distribution and the accumulation of EHC, the incentives supporting this action only increase over time. Hence let us assume that self-employment is chosen in almost all periods under an optimal policy ζ^*. Now the only question is the optimal investment policy.

Notice that as $\underline{e}$ increases and ℓ'' increases on $\left(0, \underline{e}\right)$, $\ell\left(c\left(k,l\right)/\left(1+e_i\right)\right)$ becomes significantly greater than 1 for $c\left(k,l\right) \approx \underline{e}$. Given this, and with δ sufficiently large and constraint $\lambda\left(e_i, W_i\right)$ sufficiently tight, it may be the case that it is optimal for the agent to follow a policy of undertaking saving ($s^* > 0$) for a number of periods, adding to W_i until a sufficient value of W_i is attained in order to undertake a large investment, which generates a significant amount of EHC accumulation. We can see the incentive tradeoff by differentiating just the self-employment payoff with respect to saving,

$$
U'\left[W_i + R\left(k,l;w,r;e_i,W_i\right) - s\right] = \delta EV_2\left(\ell\left(c\left(k,l\right)/\left(1+e_i\right)\right), W_i + s, x_i, \xi'\right). \qquad (D.3)
$$

It may be that the marginal future expected utility captured in V_2 may exceed the moderate marginal increase in subsequent returns from immediately investing any excess savings, if the learning function ℓ is sufficiently convex. Hence in the

significant saving periods $e'/e \gg 1$, while in other periods e'/e will be closer to 1.

Finally, in the savings periods, the value of s^* is growing in e_i due to the structure of ℓ – successively larger investments are needed to attain the higher-return segment of the learning function, necessitating increased investment (though of course this becomes increasingly feasible as e_i, and hence profits, increase).

D.2 Figures

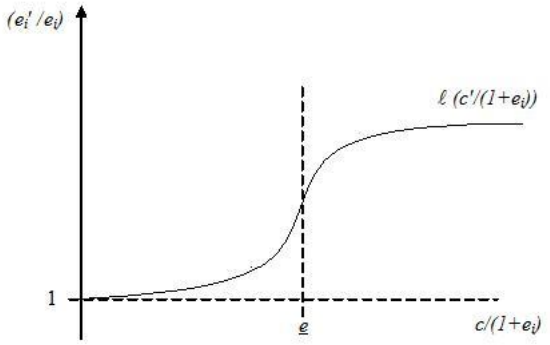


Figure D.1: EHC learning function

D.3 Tables

Table D.1: Evidence for labor market frictions: Dependent variable labor quantity, IFLS4 (2008), OLS

	(1)	(2)	(3)	(4)
	Total labor	Total labor	Wage labor	Wage labor
Number adult males	0.0925***	-0.0469	0.0509**	-0.147***
	(0.0234)	(0.0467)	(0.0205)	(0.0410)
Number adult males^2		0.0216***		0.0307***
		(0.00612)		(0.00538)
Number adult females	0.0508**	0.0742	0.00250	-0.0258
	(0.0238)	(0.0607)	(0.0209)	(0.0533)
Number adult females^2		-0.00218		0.00668
		(0.00903)		(0.00792)
Number children	0.0715***	0.106**	0.0234	0.0298
	(0.0211)	(0.0472)	(0.0186)	(0.0415)
Number children^2		-0.00773		-0.00191
		(0.00923)		(0.00810)
Age of head	0.0320**	0.0314**	0.0263**	0.0252**
	(0.0134)	(0.0134)	(0.0118)	(0.0117)
Age of head^2	-0.000356***	-0.000350***	-0.000293**	-0.000283**
	(0.000132)	(0.000132)	(0.000116)	(0.000116)
Urban dummy	YES	YES	YES	YES
Province dummies	YES	YES	YES	YES
Constant	1.848***	1.944***	0.437***	0.669***
	(0.102)	(0.120)	(0.0898)	(0.106)
Observations	4905	4905	4905	4905
R-squared adj.	0.0265	0.0285	0.0119	0.0185
F	8.409***	7.841***	4.279***	5.407***

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1

Table D.2: Log-linear fixed effect experience regressions; dependent variable monthly Net Profit

	(1) No controls	(2) With controls	(3) No controls	(4) With controls	(5) No controls	(6) With controls	(7) No exp. measures
Exp. 1 curr. ent.	0.2381***	0.0386***			0.2660***	0.1125***	
	(0.0131)	(0.0060)			(0.0181)	(0.0083)	
Exp. 1 curr. ent. ^{2}	0.0182***	-0.0020***			0.0198***	-0.0045***	
	(0.0009)	(0.0004)			(0.0010)	(0.0005)	
Exp. 2 curr. ent.	0.2597***	-0.0164**			0.3841***	0.0572***	
	(0.0146)	(0.0067)			(0.0211)	(0.0097)	
Exp. 2 curr. ent. ^{2}	0.0169***	-0.0016***			0.0155***	-0.0042***	
	(0.0010)	(0.0005)			(0.0011)	(0.0005)	
Exp. 3 curr. ent.	0.2939***	-0.0358			0.1727	0.1190**	
	(0.0709)	(0.0322)			(0.1120)	(0.0510)	
Exp. 3 curr. ent. ^{2}	0.0226***	0.0157***			0.0343***	0.0100***	
	(0.0069)	(0.0031)			(0.0075)	(0.0034)	
Total exp. 1			0.2902***	-0.0138***	-0.0085	-0.0667***	
			(0.0078)	(0.0035)	(0.0107)	(0.0050)	
Total exp. 1 ^{2}			-0.0056***	0.0002*	-0.0017***	0.0012***	
			(0.0002)	(0.0001)	(0.0003)	(0.0001)	
Total exp. 2			0.2827***	-0.0452***	-0.0913***	-0.0659***	
			(0.0090)	(0.0040)	(0.0128)	(0.0059)	
Total exp. 2 ^{2}			-0.0051***	0.0006***	-0.0001	0.0012***	
			(0.0003)	(0.0001)	(0.0003)	(0.0002)	
Total exp. 3			0.5130***	0.0072	0.1559*	-0.1548***	
			(0.0512)	(0.0217)	(0.0842)	(0.0383)	
Total exp. 3 ^{2}			-0.0181***	0.0006	-0.0107***	0.0040***	
			(0.0020)	(0.0009)	(0.0026)	(0.0012)	
Age		0.0059		0.0280***		0.0355***	0.0059
		(0.0070)		(0.0071)		(0.0072)	(0.0067)
Age ^{2}		0.0020***		0.0018***		0.0018***	0.0019***
		(0.0001)		(0.0001)		(0.0001)	(0.0001)
Education (years)		-0.0157***		-0.0121***		-0.0154***	-0.0142***
		(0.0039)		(0.0039)		(0.0039)	(0.0038)
Marriage (married=1)		8.3170***		8.2809***		8.2218***	8.2659***
		(0.0262)		(0.0264)		(0.0267)	(0.0257)
Constant	7.8529***	0.1465	7.9687***	-0.2930*	8.0381***	-0.5648***	0.4577***
	(0.0240)	(0.1667)	(0.0283)	(0.1668)	(0.0265)	(0.1712)	(0.1546)
Observations	60153	60138	60153	60138	60153	60138	62533
R-squared	0.191	0.833	0.061	0.833	0.200	0.835	0.827
F	1885.8	24026	522.04	24031	1003.1	15147	59854
Prob >F	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: The experience variables measure years of experience with 'Curr.' in current enterprise and 'Total' counting all years of experience. 1=enterprise with no employees, 2=enterprise with only h-hold/unpaid employees, and 3=enterprise with waged employees.

Table D.3: Cobb-Douglas fixed effect experience regressions; dependent variable monthly Net Profit

	(1) No controls	(2) With controls	(3) No controls	(4) With controls	(5) No controls	(6) With controls	(7) No exp. measure
Log exp. 1 curr. ent.	1.1863***	0.0549***			0.9683***	0.2335***	
	(0.0277)	(0.0129)			(0.0516)	(0.0228)	
Log exp. 2 curr. ent.	1.3263***	-0.0295**			1.3703***	0.1517***	
	(0.0306)	(0.0141)			(0.0611)	(0.0270)	
Log exp. 3 curr. ent.	1.1605***	0.2173***			0.5462*	0.4308***	
	(0.1371)	(0.0606)			(0.3088)	(0.1359)	
Log total exp. 1			0.9228***	-0.0290***	0.2194***	-0.1949***	
			(0.0238)	(0.0111)	(0.0439)	(0.0196)	
Log total exp. 2			1.0002***	-0.0829***	-0.0381	-0.1968***	
			(0.0267)	(0.0123)	(0.0527)	(0.0234)	
Log total exp. 3			1.0903***	0.1258**	0.6122**	-0.2216*	
			(0.1236)	(0.0541)	(0.2756)	(0.1213)	
Log of (age + 1)		6.3332***		6.6270***		6.6265***	6.3168***
		(0.0998)		(0.1022)		(0.1020)	(0.0907)
Log of (education + 1)		-0.1347***		-0.1261***		-0.1282***	-0.1353***
		(0.0180)		(0.0180)		(0.0180)	(0.0176)
Marriage (married=1)		8.5386***		8.5250***		8.4939***	8.4710***
		(0.0269)		(0.0269)		(0.0270)	(0.0264)
Constant	7.9312***	-19.2416***	7.8830***	-20.2419***	7.8836***	-20.2250***	-19.0804***
	(0.0262)	(0.3586)	(0.0295)	(0.3661)	(0.0292)	(0.3655)	(0.3272)
Observations	60153	60041	60153	60041	60153	60041	62436
R-squared	0.079	0.821	0.061	0.821	0.079	0.822	0.815
F	1364.9	36757	1041.1	36775	687.94	24614	73509
Prob >F	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: The experience variables measure years of experience with 'Curr.' in current enterprise and 'Total' counting all years of experience. 1=enterprise with no employees, 2=enterprise with only h-hold/unpaid employees, and 3=enterprise with waged employees.

Table D.4: Cobb-Douglas fixed effect experience regressions; dependent variable monthly Net Profit

	(1) No controls	(2) With controls	(3) No controls	(4) With controls	(5) No controls	(6) With controls	(7) No exp. measure
Log exp. 1 curr. ent.	1.1863***	0.0549***			0.9683***	0.2335***	
	(0.0277)	(0.0129)			(0.0516)	(0.0228)	
Log exp. 2 curr. ent.	1.3263***	-0.0295**			1.3703***	0.1517***	
	(0.0306)	(0.0141)			(0.0611)	(0.0270)	
Log exp. 3 curr. ent.	1.1605***	0.2173***			0.5462*	0.4308***	
	(0.1371)	(0.0606)			(0.3088)	(0.1359)	
Log total exp. 1			0.9228***	-0.0290***	0.2194***	-0.1949***	
			(0.0238)	(0.0111)	(0.0439)	(0.0196)	
Log total exp. 2			1.0002***	-0.0829***	-0.0381	-0.1968***	
			(0.0267)	(0.0123)	(0.0527)	(0.0234)	
Log total exp. 3			1.0903***	0.1258**	0.6122**	-0.2216*	
			(0.1236)	(0.0541)	(0.2756)	(0.1213)	
Log of (age + 1)		6.3332***		6.6270***		6.6265***	6.3168***
		(0.0998)		(0.1022)		(0.1020)	(0.0907)
Log of (education + 1)		-0.1347***		-0.1261***		-0.1282***	-0.1353***
		(0.0180)		(0.0180)		(0.0180)	(0.0176)
Marriage (married=1)		8.5386***		8.5250***		8.4939***	8.4710***
		(0.0269)		(0.0269)		(0.0270)	(0.0264)
Constant	7.9312***	-19.2416***	7.8830***	-20.2419***	7.8836***	-20.2250***	-19.0804***
	(0.0262)	(0.3586)	(0.0295)	(0.3661)	(0.0292)	(0.3655)	(0.3272)
Observations	60153	60041	60153	60041	60153	60041	62436
R-squared	0.079	0.821	0.061	0.821	0.079	0.822	0.815
F	1364.9	36757	1041.1	36775	687.94	24614	73509
Prob >F	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: The experience variables measure years of experience with 'Curr.' in current enterprise and 'Total' counting all years of experience. 1=enterprise with no employees, 2=enterprise with only h-hold/unpaid employees, and 3=enterprise with waged employees.

Table D.5: Log-linear fixed effect experience regressions, common experience; dependent

	(1) Common exp.	(2) Common exp.	(3) Common exp.	(4) Common exp.
Exp. 1 curr. ent.	0.4834***	0.2434***	-0.6884***	0.0319
	(0.0419)	(0.0200)	(0.0457)	(0.0252)
Exp. 1 curr. ent. ^{2}	-0.0101***	-0.0203***	0.1015***	0.0093***
	(0.0038)	(0.0019)	(0.0041)	(0.0024)
Exp. 2 curr. ent.	0.5855***	0.1617***	-0.7297***	-0.1075***
	(0.0450)	(0.0217)	(0.0543)	(0.0298)
Exp. 2 curr. ent. ^{2}	-0.0072*	-0.0155***	0.1156***	0.0160***
	(0.0038)	(0.0019)	(0.0044)	(0.0026)
Exp. 3 curr. ent.	1.4078***	0.2077***	-0.5327**	0.1075
	(0.1493)	(0.0689)	(0.2676)	(0.1419)
Exp. 3 curr. ent. ^{2}	-0.0582***	-0.0158**	0.1406***	0.0133
	(0.0171)	(0.0079)	(0.0222)	(0.0118)
Total exp. 1			1.3193***	0.0503
			(0.0613)	(0.0348)
Total exp. 1 ^{2}			-0.1122***	-0.0260***
			(0.0040)	(0.0022)
Total exp. 2			1.4995***	0.1538***
			(0.0659)	(0.0373)
Total exp. 2 ^{2}			-0.1313***	-0.0322***
			(0.0044)	(0.0025)
Total exp. 3			1.6856***	-0.0392
			(0.2665)	(0.1423)
Total exp. 3 ^{2}			-0.1814***	-0.0298***
			(0.0190)	(0.0102)
Age		0.2870***		0.3893***
		(0.0226)		(0.0216)
Age ^{2}		-0.0009***		-0.0015***
		(0.0002)		(0.0002)
Education (years)		-0.0027		-0.0053
		(0.0087)		(0.0082)
Marriage (married=1)		7.5763***		6.1517***
		(0.0859)		(0.0944)
Constant	7.2483***	-6.4288***	6.4443***	-7.2203***
	(0.1113)	(0.5367)	(0.2689)	(0.5026)
Observations	9125	9125	9125	9125
R-squared	0.154	0.823	0.449	0.846
F	174.3	2661	388.9	1960
Prof >F	0.000	0.000	0.000	0.000

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: The experience variables measure years of experience with 'Curr.' in current enterprise and 'Total' counting all years of experience. 1=enterprise with no employees, 2=enterprise with only h-hold/unpaid employees, and 3=enterprise with waged employees.
Note: 'Common exp.' refers to fixed effects panel regressions only on the subsample of individuals with 8-12 years of self-employment experience.

Table D.6: Log-Linear fixed effect experience regressions, common start age; dependent variable monthly net profit

	(1) Common age	(2) Common age	(3) Common age	(4) Common age
Exp. 1 curr. ent.	0.5127***	0.0483***	0.2669***	0.0562***
	(0.0206)	(0.0097)	(0.0247)	(0.0115)
Exp. 1 curr. ent. ^{2}	0.0055***	-0.0048***	-0.0152***	-0.0052***
	(0.0012)	(0.0006)	(0.0013)	(0.0006)
Exp. 2 curr. ent.	0.3150*	0.0695	0.1505	0.1705
	(0.1857)	(0.0789)	(0.2578)	(0.1182)
Exp. 2 curr. ent. ^{2}	-0.0279	0.0057	-0.0045	-0.0028
	(0.0415)	(0.0176)	(0.0420)	(0.0193)
Exp. 3 curr. ent.	-0.3169	-0.5604*	-3.5910*	-3.3290***
	(0.7930)	(0.3368)	(1.9517)	(0.8944)
Exp. 3 curr. ent. ^{2}	0.1442	0.1147	0.7639	0.5664**
	(0.1966)	(0.0835)	(0.5104)	(0.2339)
Total exp. 1			0.2524***	-0.0141
			(0.0162)	(0.0100)
Total exp. 1 ^{2}			0.0120***	0.0006*
			(0.0005)	(0.0004)
Total exp. 2			0.0968	-0.0822
			(0.1575)	(0.0723)
Total exp. 2 ^{2}			-0.0070	0.0034
			(0.0085)	(0.0039)
Total exp. 3			3.3251*	2.8642***
			(1.8463)	(0.8461)
Total exp. 3 ^{2}			-0.6133	-0.4616**
			(0.4708)	(0.2157)
Age		-0.2266***		-0.1981***
		(0.0145)		(0.0221)
Age ^{2}		0.0048***		0.0044***
		(0.0002)		(0.0003)
Education (years)		-0.0239***		-0.0227***
		(0.0064)		(0.0064)
Marriage (married=1)		9.2475***		9.2334***
		(0.0488)		(0.0496)
Constant	7.1609***	5.9775***	4.1908***	5.5535***
	(0.0473)	(0.2512)	(0.0821)	(0.3528)
Observations	16714	16714	16714	16714
R-squared	0.270	0.868	0.374	0.869
F	740.4	7928	599.0	4964
Prof >F	0.000	0.000	0.000	0.000

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
Note: The experience variables measure years of experience with 'Curr.' in current enterprise and 'Total' counting all years of experience. 1=enterprise with no employees, 2=enterprise with only h-hold/unpaid employees, and 3=enterprise with waged employees.
Note: 'Common age' refers to fixed effects panel regressions only on the subsample of individuals who first entered self-employment in their 20s.

Table D.7: The propensity of the household to engage in running an enterprise, 2001-2008, probit

	(1) Coef.		Std. Err.	P >|z|	(2) Coef.		Std. Err.	P >|z|
Age Head	-0.0002		0.008	0.979	-0.0042		0.008	0.596
Age Head^2	0.0000		0.000	0.662	0.0000		0.000	0.892
Head Ent Exp					-0.0278	***	0.003	0.000
Head Ent Exp^2					0.0002	***	0.000	0.000
Head Ent Exp No Empl					0.1855	***	0.016	0.000
Head Ent Exp Fam Empl					0.0880	***	0.018	0.000
Head Ent Exp Wage Lab					0.3891	***	0.069	0.000
Head First Occ No Empl					0.0923		0.138	0.504
Head First Occ Fam Empl					0.0113		0.227	0.960
Head First Occ Wage Empl					-0.1422		0.540	0.792
Number Adult Female	0.0393	**	0.016	0.014	0.0839	***	0.017	0.000
Number Adult Male	0.0222		0.015	0.147	0.0544	***	0.016	0.001
Number Child	0.0373	***	0.013	0.003	0.0363	***	0.013	0.005
Wealth 2000 USD	0.0000	***	0.000	0.005	0.0000	***	0.000	0.001
Wealth 2000 USD^2	0.0000		0.000	0.359	0.0000		0.000	0.143
Bank BRI	-0.0138		0.064	0.829	0.0171		0.066	0.795
People's Credit	-0.1120	**	0.048	0.020	-0.0823	*	0.049	0.095
Village Credit Union	0.0045		0.043	0.918	0.0302		0.045	0.497
Village Coop	-0.0088		0.042	0.833	0.0016		0.043	0.970
Other form Coop	-0.0840	**	0.041	0.040	-0.0828	**	0.042	0.048
State Private Bank	0.0051		0.050	0.920	-0.0010		0.052	0.984
BMT	0.0276		0.047	0.561	0.0345		0.049	0.479
Java Dummy	-0.0188		0.041	0.647	0.0338		0.042	0.423
Urban Dummy	0.3003	***	0.041	0.000	0.2783	***	0.043	0.000
Constant	-0.8777	***	0.057	0.000	-0.8864	***	0.060	0.000
N	7968				7968			
LR Chi-square(2)	162.1100				635.7500			
Log-likelihood	-4335.9184				-4099.0964			

*** 1 percent, ** 5 percent, * 10 percent significance.
Note: Wealth values are the lagged, year 2000 values. Monetary values converted to USD.

BIBLIOGRAPHY

[Angrist, 1990] Angrist, Joshua. 1990. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *The American Economic Review*, **80**(3), 313–336.

[Astebro & Bernhardt, 2005] Astebro, Thomas, & Bernhardt, Irwin. 2005. The Winner's Curse of Human Capital. *Small Business Economics*, **24**, 63–78.

[Azariadis & Drazen, 1990] Azariadis, Costas, & Drazen, Allen. 1990. Threshold Externalities in Economic Development. *Quarterly Journal of Economics*, **105**, 521–526.

[Banerjee & Duflo, 2010] Banerjee, Abhijit, & Duflo, Esther. 2010. *Giving Credit Where It is Due*. MIT working paper.

[Banerjee & Newman, 1993] Banerjee, Abhijit, & Newman, Andrew. 1993. Occupational Choice and the Process of Development. *The Journal of Political Economy*, **101**(2), 274–298.

[Banerjee *et al.*, 2009] Banerjee, Abhijit, Duflo, Esther, Glennerster, Rachel, & Kinnan, Cynthia. 2009. *The Miracle of Microfinance? Evidence from a Randomized Evaluation*. MIT working paper.

[Barrett *et al.*, 2005] Barrett, Christopher, Bezuneh, Mesfin, Clay, Daniel C., & Reardon, Thomas. 2005. Heterogeneous Constraints, Incentives and Income Diversification Strategies in Rural Africa. *Quarterly Journal of International Agriculture*, **44**(1), 37–60.

[Becker, 1964] Becker, Gary. 1964. *Human Capital*. New York, NY: Columbia University Press.

[Behrman & Rosenzweig, 1999] Behrman, Jere, & Rosenzweig, Mark. 1999. 'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates. *Economics of Education Review*, **18**(2), 159–167.

[Benjamin, 1992] Benjamin, Dwayne. 1992. Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica*, **60**(2), 287–322.

[Bianchi & Bobba, 2010] Bianchi, Milo, & Bobba, Matteo. 2010 (December). *Liquidity, Risk and Occupational Choices*. Working paper.

[Bloom *et al.*, 2010] Bloom, Nicholas, Mahajan, Aprajit, McKenzie, David, & Roberts, John. 2010. Why Do Firms in Developing Countries Have Low Productivity. *American Economic Review Papers and Proceedings*, **2**, 619–623.

[Bruhn *et al.*, 2010] Bruhn, Miriam, Karlan, Dean, & Schoar, Antoinette. 2010. What Capital is Missing in Developing Countries? *American Economic Review Papers and Proceedings*, **2**, 629–633.

[Buera, 2009] Buera, Francisco. 2009. A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence. *Annals of Finance*, **5**(3), 443–464.

[Cagetti & Nardi, 2006] Cagetti, Marco, & Nardi, Mariacristina De. 2006. Entrepreneurship, Frictions and Wealth. *Journal of Political Economy*, **114**(5), 835–870.

[Card, 1999] Card, David. 1999. *The Causal Effect of Education on Earnings*. Vol. 3A. Elsevier.

[Carter & Olinto, 2003] Carter, Michael, & Olinto, Pedro. 2003. Getting Institutions Right for Whom? Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment. *American Journal of Agricultural Economics*, **85**(1), 173–186.

[CIA, 2011] CIA. 2011. *CIA Factbook*. Online. Source: https://www.cia.gov/library/publications/the-world-factbook/geos/id.html.

[Cope, 2005] Cope, Jason. 2005. Toward a Dynamic Learning Perspective of Entrepreneurship. *Entrepreneurship Theory and Practice*, **29**(4), 373–397.

[Cunha *et al.*, 2006] Cunha, Flavio, Heckman, James J., Lochner, Lance, & Masterov, Dimitriy V. 2006. *Interpreting the Evidence on Life Cycle Skill Formation*. Vol. I. Elsevier.

[Dasgupta & Ray, 1986] Dasgupta, Partha, & Ray, Debraj. 1986. Inequality as a Determinant of Malnutrition and Unemployment. *The Economic Journal*, **96**(384), 1011–1034.

[de Mel *et al.*, 2008] de Mel, Suresh, McKenzie, David, & Woodruff, Christopher. 2008. Returns to Capital: Results from a Randomized Experiment. *Quarterly Journal of Economics*, **123**(4), 1329–1372.

[de Mel *et al.*, 2010] de Mel, Suresh, McKenzie, David, & Woodruff, Christopher. 2010. *Who are the Microenterprise Owners? Evidence from Sri Lanka on Tokman V. De Soto*. The University of Chicago Press. Pages 63–88.

[Dunn & Holtz-Eakin, 2000] Dunn, Thomas, & Holtz-Eakin, Douglas. 2000. Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links. *Journal of Labor Economics*, **18**(2), 282–305.

[Evans & Jovanovic, 1989] Evans, David, & Jovanovic, Boyan. 1989. An Estimated Model of Entrepreneurial Choice under Liquidity Constraints. *Journal of Political Economy*, **97**(4), 808–827.

[Fairlie & Robb, 2007a] Fairlie, Robert, & Robb, Alicia. 2007a. Families, Human Capital, and Small Business: Evidence from the Characteristics of Business Owners Survey. *Industrial and Labor Relations Review*, **60**(2), 225–245.

[Fairlie & Robb, 2007b] Fairlie, Robert, & Robb, Alicia. 2007b. Why Are Black-Owned Businesses Less Successful Than White-Owned Businesses? The Role of Families, Inheritances, and Business Human Capital. *Journal of Labor Economics*, **25**(2), 289–323.

[Feder, 1985] Feder, Gershon. 1985. The Relation Between Farm Size and Farm Productivity. *Journal of Development Economics*, **18**, 297–313.

[Field & Torero, 2006] Field, Erica, & Torero, Maximo. 2006. *Do Property Titles Increase Credit Access Among the Urban Poor? Evidence from a Nationwide Titling Program*. Harvard University mimeo.

[Fields & Pfefferman, 2003] Fields, Gary, & Pfefferman, Guy (eds). 2003. *Pathways Out of Poverty: Private Firms and Economic Mobility in Developing Countries*. Norwell, MA: Kluwer Academic Publishers.

[Foster & Rosenzweig, 1995] Foster, Andrew D., & Rosenzweig, Mark R. 1995. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, **103**(6), 1176–1209.

[Frankenberg & Karoly, 1995] Frankenberg, Elizabeth, & Karoly, Lynn. 1995 (November). *The 1993 Indonesia Family Life Survey: Overview and Field Report*. Tech. rept. RAND Corporation, Santa Monica, CA.

[Frankenberg & Thomas, 2000] Frankenberg, Elizabeth, & Thomas, Duncan. 2000 (March). *The Indonesia Family Life Survey (IFLS): Study Design and Re-*

sults from Waves 1 and. Tech. rept. RAND Corporation, Santa Monica, CA. DRU-2238/1-NIA/NICHD.

[Galor & Zeira, 1993] Galor, Oded, & Zeira, Joseph. 1993. Income Distribution and Macroeconomics. *Review of Economic Studies*, **60**(1), 35–52.

[Ghatak *et al.*, 2007] Ghatak, Maitreesh, Morelli, Massimo, & Sjostrom, Tomas. 2007. Entrepreneurial Talent, Occupational Choice, and Trickle Up Policies. *Journal of Economic Theory*, **137**, 27–48.

[Hurst & Lusardi, 2004] Hurst, Erik, & Lusardi, Annamaria. 2004. Liquidity Constraints, Household Wealth, and Entrepreneurship. *Journal of Political Economy*, **112**(2), 319–347.

[Ikhwan & Johnson, 2009] Ikhwan, Andi, & Johnson, Don Edwin. 2009. *The Constraints in Accessing Credit Faced by Rural Non-Farm Enterprises*. Singapore: Institute of South Asian Studies.

[Iyigun & Owen, 1998] Iyigun, Murat, & Owen, Ann. 1998. Risk, Entrepreneurship, and Human-Capital Accumulation. *American Economic Review Papers and Proceedings*, **88**(2), 454–457.

[Jovanovic, 1982] Jovanovic, Boyan. 1982. Selection and the Evolution of Industry. *Econometrica*, **50**(3), 649–670.

[Jovanovic & Nyarko, 1996] Jovanovic, Boyan, & Nyarko, Yaw. 1996. Learning by Doing and the Choice of Technology. *Econometrica*, **64**(6), 1299–1310.

[Kanbur, 1979] Kanbur, Ravi. 1979. Of Risk Taking and the Personal Distribution of Income. *Journal of Political Economy*, **87**(4), 769–797.

[Karlan & Zinman, 2009] Karlan, Dean, & Zinman, Jonathan. 2009. Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment. *Econometrica*, **77**(6), 1993–2008.

[Karlan & Zinman, 2010] Karlan, Dean, & Zinman, Jonathan. 2010 (January). *Expanding Microenterprise Credit Access: Using Randomized Supply Decisions to Estimate Impacts in Manila*. Working paper.

[Keane & Wolpin, 1997] Keane, Michael, & Wolpin, Kenneth. 1997. The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3), 473–522.

[Kihlstrom & Laffont, 1979] Kihlstrom, Richard, & Laffont, Jean-Jacques. 1979. A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion. *Journal of Political Economy*, **87**, 719–749.

[Klein & Cook, 2006] Klein, Peter, & Cook, Michael. 2006. T.W. Schultz and the Human-Capital Approach to Entrepreneurship. *Review of Agricultural Economics*, **28**(3), 344–350.

[Lerner, 2009] Lerner, Joshua. 2009. *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed–and What to Do About It*. Princeton, NJ: Princeton University Press.

[Liedholm & Mead, 1999] Liedholm, Carl, & Mead, Donald. 1999. *Small Enterprises and Economic Development: The Dynamics of Micro and Small Enterprises*. New York, NY: Routledge.

[Lucas, 1978] Lucas, Robert. 1978. On the Size Distribution of Business Firms. *Bell Journal of Economics*, **9**, 508–523.

[McLeod, 1980] McLeod, Ross. 1980 (December). *Finance and Entrepreneurship in the Small-Business Sector in Indonesia*. Ph.D. thesis, Australian National University, Canberra, Australia.

[Nafziger & Terrell, 1996] Nafziger, Wayne, & Terrell, Dek. 1996. Entrepreneurial Human Capital and the Long-Run Survival of Firms in India. *World Development*, **24**(4), 689–696.

[Naude, 2008] Naude, Wim. 2008. *Entrepreneurship in Economic Development*. UNU-WIDER Research Paper No. 2008/20.

[Nicolau *et al.*, 2008] Nicolau, Nicos, Shane, Scott, Hunkin, Janice, Cherkas, Lynn, & Spector, Tim. 2008. Is the Tendency to Engage in Entrepreneurship Genetic? *Management Science*, **54**(1), 167–179.

[Otani, 1996] Otani, Kiyoshi. 1996. A Human Capital Approach to Entrepreneurial Capacity. *Economica*, **63**(250), 273–289.

[Paulson & Townsend, 2004] Paulson, Anna, & Townsend, Robert. 2004. Entrepreneurship and Financial Constraints in Thailand. *Journal of Corporate Finance*, **10**, 229–262.

[Politis, 2008] Politis, Diamanto. 2008. *The Process of Entrepreneurial Learning*. New York, NY: Routledge.

[Poppele *et al.*, 1999] Poppele, J., Sumarto, Sudarno, & Pritchett, Lant. 1999. *Social Impacts of the Indonesian Crisis: New Data and Policy Implications*. Tech. rept. World Bank, Jakarta, Indonesia. SMERU report.

[Porta & Shleifer, 2008] Porta, Rafael La, & Shleifer, Andrei. 2008. The Unofficial Economy and Economic Development. *Brookings Papers on Economic Activity*.

[Roy, 1951] Roy, A.D. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers (New Series)*, **3**, 135–146.

[Schoar, 2010] Schoar, Antoinette. 2010. *The Divide Between Subsistence and Transformational Entrepreneurship*. Vol. 10. Chicago, IL: The University of Chicago Press.

[Schultz, 1980] Schultz, T.W. 1980. Investment in Entrepreneurial Ability. *The Scandanavian Journal of Economics*, **82**(4), 437–448.

[Shane, 2003] Shane, Scott. 2003. *A General Theory of Entrepreneurship: The Individual-Opportunity Nexus*. Cheltenham: Edward Elgar.

[Shane, 2008] Shane, Scott. 2008. *Illusions of Entrepreneurship: The Costly Myths That Enbtrepreneurs, Investors, and Policy Makers Live By*. New Haven, CT: Yale University Press.

[Smith *et al.*, 2002] Smith, James, Thomas, Duncan, Frankenberg, Elizabeth, Beegle, Kathleen, , & Teruel, Graciela. 2002. Wages, Employment and Economic Shocks: Evidence from Indonesia. *Journal of Population Economics*, **15**, 161–193.

[Strauss *et al.*, 2004] Strauss, John, Beegle, Kathleen, Sikoki, Bondan, Dwiyanto, Agus, Herawati, Yulia, & Witoelar, Firman. 2004 (March). *The Third Wave of the Indonesia Family Life Survey (IFLS)*. Tech. rept. RAND Corporation, Santa Monica, CA. DRU-2238/1-NIA/NICHD.

[Strauss *et al.*, 2009] Strauss, John, Witoelar, Firman, Sikoki, Bondan, & Wattie, Ana Marie. 2009 (April). *The Fourth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report*. Tech. rept. WR-675/1-NIA/NICHD.

[Taveras, 2010] Taveras, Carmen. 2010. *Entrepreneurship, Learning and Wealth*. MIT mimeo.

[Thomas *et al.*, 2000] Thomas, Duncan, Beegle, Kathleen, & Frankenberg, Elizabeth. 2000. *Labor Market Transitions of Men and Women During an Economic Crisis: Evidence from Indonesia*. RAND Labor and Population Program working paper series 00-11.

[Udry & Anagol, 2006] Udry, Christopher, & Anagol, Santosh. 2006. The Return to Capital in Ghana. *American Economic Review Papers and Proceedings*, **2**, 388–393.