Essays on

Entrepreneurship and Economic Development

Dissertation
zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Universität Göttingen

vorgelegt am 7. 9. 2007
von Jagannadha Pawan Tamvada
aus New Delhi

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die eingereichte Dissertation

Essays on Entrepreneurship and Economic Development
selbständig verfasst habe. Anderer als der von mir angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht.

Göttingen, den 7. September 2007, Jagannadha Pawan Tamvada

Dedicated to Amma, Nannagaru and Sai Maa

Abstract

Two compelling reasons motivate the work in this dissertation. While enormous literature on developed countries has emerged identifying the determinants of entrepreneurship and its impact, very little is known about the characteristics and the role of entrepreneurship in less developed countries. Who are entrepreneurs in such contexts and who amongst them create jobs for others? What is the impact of geographic location on the initial size of new firms entering markets? What are the welfare implications of entrepreneurship in a developing country and what are the dynamics of entrepreneurial choice? Furthermore, recent methodological advances in econometrics allow rigorous analysis of occupational choice problems and the determinants of new firm formation. In this dissertation, I employ new tools of spatial analysis, Bayesian semi-parametric and non-parametric methods and some recent advances in econometrics to examine these questions.

Publications

The paper *Religion and Entrepreneurship*, co-authored with David Audretsch and Werner Boente, is published as a Center for Economic Policy Research (CEPR) Discussion Paper. The other papers in the dissertation are authored by me and have been presented at international conferences, doctoral colloquiums and faculty seminars.

The research work in this dissertation has been accepted for presentation at the First World Congress of Spatial Econometrics (Cambridge, 2007), the 44th European Regional Science Association's Annual Congress (Paris, 2007), the International Council for Small Business Research (Finland, 2007), the IZA-World Bank Conference on Employment and Development (Bonn, 2007) and the Second Annual Max Planck Indian Institute of Science (IISc) International Conference on Entrepreneurship, Innovation and Economic Growth (Bangalore, 2007).

The research in this dissertation has been presented at the Schumpeter Conference (Nice, 2006), the 20th Research in Entrepreneurship Conference (Bruessels, 2006), the First Annual Max Planck India Workshop on Entrepreneurship, Innovation, and Economic Growth (Bangalore, 2006), Hellenic Workshop on Entrepreneurship and Productivity (Patras, 2006), the European Summer School in Industrial Dynamics (Corsica, 2006), the Babson Doctoral Consortium (Bloomington, 2006), Augustin Cournot Doctoral Days (Strasbourg, 2006), the Technology Transfer Society's Annual Conference (Kansas City, 2005) and the G-Forum's Annual Conference (Jena, 2006).

The work has also been presented at internal seminars at the Max Planck Institute of Economics, Jena and at the Faculty of Economics, University of Göttingen.

Acknowledgements

I am greatly indebted to Prof. Stephan Klasen for giving me an opportunity to pursue doctoral studies in economics, a dream I had cherished since my high school days. Without his intellectual guidance, this dissertation would not have seen its completion. I am also grateful to Prof. David Audretsch, whose inspiration, support, and kindness have helped me to complete the task at hand. I owe enormous gratitude to him not just for giving me the prestigious Max Plank PhD scholarship, but also for introducing me to entrepreneurship and constantly guiding me through the scholarship. I also show my gratitude to Prof. Walter Zucchini for his guidance and encouragement. My discussions with my supervisors form the foundations of this dissertation, and I would like to express my gratitude for giving me their precious time and intellectual support.

I am grateful to Prof. Amartya Sen for his inspiration not just for me, but for many young Indians to study economics. I thank him for sparing some of his valuable time for me when I visited him at Harvard. I also thank Professors TVS Ram Mohan Rao and Vishwanath Pandit for their constant guidance and support. I also thank Swami Supernanada, a monk of the Rama Krishna Mission order, who introduced me to this fascinating subject.

I thank all my colleagues at the Max Planck Institute, especially Werner Boente and Max Keilbach. I would also like to thank Taylor Aldridge, Melanie Aldridge, Iris Beckman, Saradindu Bhaduri, Andreas Chai, Andrea Conte, Sameeksha Desai, Stephan Heblich, Anja Klaukien, Stefan Krabel, Adam Lederer, Prashanth Mahagaonkar, Erik Monsen, Pamela Mueller, Holger Patzelt, Stephan Schütze, Jörg Zimmerman, and all my colleagues for making my stay in Jena both intellectually and personally rewarding. My special thanks go to Kerstin Schük, Madeleine Schmidt, and Lydia Nobis for their continuous efforts at making my stay at the Institute memorable. I also show my gratitude towards

Thomas Bauman and his IT team and Katja Müller and her library team at the Max Plank Institute for attending to all my academic requirements. I also thank Marten Koppenhagen, Thilo Klein, and particularly Alex Audretsch for providing valuable research assistance. Thus, I thank the Max Planck Institute of Economics and all its wonderful staff for providing an environment that is so conducive to research.

I also would like to thank my colleagues at the Chair for Development Economics at Göttingen, particularly Micheal Grimm, Melanie Grosse, Isabal Günther, Andrey Launov, Felicitas Nowak-Lehman Danzinger, Ken Harttgen, Mark Misselhorn, Jan Priebe, Dana Schüler, Sebastian Vollmer and Julian Wiesbrod. My special thanks also go to Michaela Beckmann. Roswitha Brinkmann and the international office of the Goettingen University have my great appreciation for giving me a scholarship during my stay there in 2004, as does Prof. Manfred Denker for accepting me as a member of the Center for Statistics in Göttingen.

I give my sincerest thanks to the Lindau Nobel Council for selecting me to participate at the Second Lindau meeting of Nobel laureates, giving me an unparalleled opportunity to meet and listen to some of the greatest living legends in the field of economics. I also extend my thanks to the Max Planck Society for giving me a grant to organize the First Annual Max Planck India Workshop on Entrepreneurship, Innovation, and Economic Growth in partnership with the Indian Institute of Science, Bangalore. This conference reaffirmed to me that further entrepreneurship research focusing on developing countries such as India is an absolute necessity to promoting their development. Finally, I would like to thank the Kauffman Foundation for sponsoring my participation at the Babson Doctoral Consortium.

The Ministry of Small Scale Industries provided much assistance to me by providing me firm-level data, as did the Reserve Bank of India by inviting me for a research stay at the Rural Credit and Policy Department.

I am greatly indebted to my parents for their love, affection, prayers, and blessings. In storm and in calm, they have stood by me. I am equally grateful to Sai Baba, who showed me that man is born not for pursuing self-interest, but to serve humanity. He stayed with me through-and-through. Like my parents, he always showered unconditional love and affection on me and inspired me to follow my heart and pursue this path. As his student in Prashanti Nilayam, I learnt that God is always with the man who has faith and who never stops dreaming about doing some good to the world he lives in. My siblings, Sirish and Mallika, have always inspired me to finish my PhD. Their love and affection has been overwhelming. Last but not the least, I am thankful to Mili Shrivastava for believing in me, inspiring me to walk this path, and treating me with great care and concern over the last five years. I am also thankful for her intellectual insights that had prompted me to constantly evaluate my progress.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Almost four decades ago, Baumol (1968, p. 71) proclaimed that "in a growth conscious world I remain convinced that encouragement of the entrepreneur is the key to the stimulation of growth." Entrepreneurship, however, remained hidden and elusive from the grasp of economists. Fortunately, in recent years, the economics of entrepreneurship emerged as a compelling subject, providing insights into the entrepreneurial processes. Bringing together this literature on entrepreneurship, Parker (2004, p. 1) notes that "entrepreneurship has only recently come to be regarded as a subject." While the debate in scholarly community has still not conclusively accepted even the definition of entrepreneurship, a vast literature has emerged over the last two decades providing insights into the many facets of entrepreneurship. Though each such facet is incomplete by itself, together they offer a comprehensive understanding of the entrepreneurial choice, new firm formation and the role of entrepreneurship in economic growth. Reflecting a broad consensus that has emerged in recent times, Lazear (2002, p. 1) claims that "the entrepreneur is the single most important player in the economy." This dynamically expanding subject, *the economics of entrepreneurship*, however, neglected entrepreneurship in less developed countries. This dissertation exploits recent advances in Bayesian semiparametric methods and geoadditive models (Fahrmeir and Lang, 2001a) and large databases of individual and firm-level micro-data from India to provide fresh perspectives of the entrepreneurial processes and their relationship to economic development. This dissertation underlines the nexus between the entrepreneur, the firm, and the region by emphasizing the role of the spatial location in simultaneously determining the entrepreneurship choice and the size of new firms. The returns to occupational choice and the spatio-temporal dynamics of self-employment choice form another major part of this dissertation. The role of the caste system and religion in determining the entrepreneurship choice is studied, as such factors play a crucial and important role in determining the occupational choice in India.

The theme of the second and third chapters is the determinants of self employment and the role of exogenous constraints in occupational choice. While a vast literature has emerged examining the determinants of entrepreneurship, the role of spatial location and the neighborhood of an individual have rarely been considered as determinants of entrepreneurship choice. There are compelling reasons, however, to assume that such factors play an important role in shaping the occupational choice of people. Thus, in chapter 2, I analyze the role of geographic location as a micro-determinant of self-employment choice. I also study the impact of human capital accumulation on occupational choice in agricultural and nonagricultural sectors in India. In chapter 3, I analyze the role of religion as an exogenous constraint on the occupational choice of individuals. Recent studies (Iannaccone, 1998; McCleary and Barro, 2006a; Guisa et al., 2006) link religion with economy but the channels through which religion influences the economy are not examined by the existing literature. One such channel through which religion might influence the economy is through entrepreneurship. Religions impose behavioral constraints and influence economic outcomes. For instance, the institution of the caste system in Hindusim is likely to act as an exogenous constraint on the occupational choice of Hindus. In this paper, I examine the role of religion and class structures in promoting or inhibiting entrepreneurial behavior.

The theme of the fourth chapter is the impact of ownership structure and geographic location on the size of new entrants. In this chapter, I revisit the question of firm size at entry. A number of studies show that, for new entrants at least, the initial size influences growth and survival. The determinants of the size of firms at entry, however, remained under-researched and neglected in this discussion, for a long time. The few studies on start-up size show that the industry characteristics such as turbulence, minimum efficient scale, and industry growth (Mata and Machado, 1996; Mata, 1996) and human capital of entrepreneurs (Astebro and Bernhardt, 2005; Colombo et al., 2004; Colombo and Grilli, 2005), determine the start-up size of new firms. However, the role of spatial location on the start-up size has never been studied. Chapter 4 incorporates ownership structure and geographic location as micro-determinants of start-up size, using micro data from India.

The theme of the fifth chapter is entrepreneurship and welfare. A growing body of literature identifies returns to self-employment in developed countries (Hamilton, 2000). Historically, the development economics literature has considered self-employment in less developed countries, to be a part of the so-called informal sector (Harris and Todaro, 1970). More recently, a growing body of literature argues that the informal sector is a blend of a low-productive disadvantaged sector and a voluntary competitive sector (Cunningham and Maloney, 2001; Fields, 2005; Günther and Launov, 2006). In chapter 5, I link occupational decisions of the household with a direct measure of welfare, per-capita consumption. Using quantile regressions, I estimate occupational choice as a determinant of welfare. Furthermore, using selection methods that allow for corrections after multinomial logit estimation (Bourguignon et al., 2007), I test if a process of endogenous non-random selection determines the selection of individuals into different occupations. Thus, the underlying process of selection into occupations and subsequent returns in terms of welfare are examined to see whether people are compelled to opt for low-productivity self-employment or whether they voluntarily self-select based on their unobserved abilities, in a developing country.

The theme of the sixth chapter is the evolution of the entrepreneurial choice over time and space. The post-liberalisation era of Indian economy has witnessed a surge in entrepreneurial activity. The dynamics of occupational choice in this context are not explored in the literature. Using two cross-sectional databases of the National Sample Survey Organization of India (NSSO) data, I examine the spatial dynamics of self-employment choice and in particular, the role of education as a determinant of entrepreneurship. In addition, using three surveys of the NSSO (1994-1995, 1999-2000 and 2004), I also construct a psuedo-panel (Deaton, 1997; Moffitt, 1993; Verbeek, 2006) to examine the dynamics of entrepreneurial activity in India. The final chapter constructs the dual theory of entrepreneurship, linking results of the chapters of this dissertation. This chapter showcases a coherent theory of self-employment, firm formation, and geographic location and concludes this dissertation.

Chapter 2

The Geography and Determinants of Entrepreneurship

In this chapter, we examine the determinants of self-employment choice in India. In addition to standard determinants such as age, sex and education, we examine geographic location as a micro-determinant of self-employment choice using Bayesian semiparametric methodologies and geoadditive techniques. The analysis suggests the presence of spatial patterns in self-employment and a clear north-south divide when individuals of agricultural and nonagricultural sectors are considered together; however, such spatial patterns are less pronounced when individuals in the nonagricultural sector alone are considered in the analysis. The results further suggest nonlinear relationships between age, wealth and the probability of self-employment and demonstrate a contrasting link between education and self-employment choice in the two sectors.

2.1 Introduction

Referred to as self-employment in some studies and new firm formation, entry rate or start-up activity in others, entrepreneurship has captured the attention of not just labor economists or scholars of industrial dynamics, but also growth theorists.[1] Parallel to this body of literature linking entrepreneurship to the eco-

[1]For instance, Aghion et al. (2004) show that entry induces productivity growth of incumbent firms. Iyigun and Owen (1999) argue that, in an economy where both entrepreneurial and professional human capital affect the future level of technology, the initial stocks of both types of human capital are important for the process of development and countries that have too little entrepreneurial or professional human capital end up in a development trap in which production is carried out in the unskilled sector only and there is no human capital investment of any type.

nomic progress,[2] a vast literature has emerged examining the determinants of entrepreneurship. A proliferation of studies aimed at explaining the characteristics of entrepreneurs, the determinants of occupational choice of individuals and the contexts that promote entrepreneurship has resulted (see Parker, 2004, for a survey of this literature).

However, the spatial location and the neighborhood of an individual have rarely been considered as determinants of the entrepreneurship choice, while there are compelling reasons, to assume that such factors play an important role in shaping the occupational choice of people.[3] Beginning with the seminal paper of Krugman (1991), the role of economic geography in determining economic outcomes is widely recognized. This study uses a new approach to analyze the determinants of entrepreneurship using recent advances in Bayesian semiparametric geoadditive models that allow incorporation of spatial location as a micro determinant of self-employment choice.

Until recently, the entrepreneurship literature has also largely ignored the labor markets of Less Developed Countries (LDCs). An impression of non-competitive labor markets in LDCs rendered the entrepreneurial sector of LDCs uninteresting to scholars researching the personality of entrepreneurs. Harris and Todaro (1970), for instance, predicted that absence of economic opportunities, combined with high unemployment, forces individuals into low productivity self employment in LDCs. As Blau (1986, p. 839) notes, "In most studies of LDC labor markets the self-employed are either ignored or treated as part of the so-called informal sector." There is growing evidence, however, that the labor markets of LDCs are actually competitive and that self-employment is not merely a subsistence level activity in LDCs (Maloney, 2004; Mohapatra et al., 2007). In such a context, it is surprising to note that neither the determinants of entrepreneurship nor the role of entrepreneurship in some emerging economies is analyzed in empirical literature. This paper also bridges this gap, by examining the deter-

[2]See Murphy et al. (1991), Banerjee and Neuman (1993), Iyigun and Owen (1999), Baumol (2002) for theoretical and Berkowitz and DeJong (2005) and Audretsch et al. (2006) for empirical studies linking occupational choice and entrepreneurship to economic development.

[3]Some studies do recognize the importance of regional factors in determining the self-employment choice. However, most of these studies are based on aggregated data and assess the quantum of entrepreneurial activity as a function of regional variables such as unemployment, tax rates and small business employment (Blanchflower, 2000; Blanchflower and Oswald, 1998; Reynolds et al., 1994). Another strand of literature examines the effect of new business formation on the region (Fritsch and Müller, 2004).

minants of self-employment choice in one such growing economy, India, that has in recent years, experienced substantial leaps in both its entrepreneurial activity and growth rates.

Household level data collected by the National Sample Survey Organization in 2004 are used for the empirical analysis. The effects of individual personal characteristics, educational background, household characteristics and non-linear effects of continuous covariates such as age and geographic location on the probability of being self-employed are jointly estimated using geoadditive models. The results suggest that outside of agriculture, educated individuals are more likely to be salaried employees while in the agricultural sector, educated individuals are more likely to be self-employed. Strong spatial patterns are observed and these are primarily attributable to spatial self-employment patterns in the agricultural sector. Consistent with earlier empirical studies on the determinants of entrepreneurship, the results suggest that Indian males, married and older citizens are more likely to be self-employed as well.

The next section discusses the literature and states the hypotheses on the determinants of self-employment in a developing economy. The third and fourth sections describe the semiparametric geoadditive modeling techniques and the dataset. The fifth section presents the empirical analysis. The final section provides conclusions and discusses possible avenues for future research.

2.2 Theoretical Background

2.2.1 Determinants of Self-employment

Empirical research on occupational choice in developed economies suggests that individuals' personal characteristics (Kihlstrom and Laffont, 1979; Evans and Leighton, 1989b) and regional factors (Georgellis and Wall, 2000) play an important role in influencing the entrepreneurial decisions. The decision of individuals to become entrepreneurs is generally modeled in terms of utility maximization, where the economic returns from entrepreneurship are compared to returns of wage employment (Lucas, 1978; Holmes and Schmitz Jr., 1990; Jovanovic, 1994).

Individual-specific characteristics such as risk aversion (Kihlstrom and Laffont, 1979), prior self-employment experience (Evans and Leighton, 1989b), education, human capital, and age (Zucker et al., 1998; Bates, 1990; Rees and Shaw, 1986; Blanchflower and Meyer, 1994) and personality characteristics (McCelland, 1964), are found to have an impact on an individual's entrepreneurship choice. As Parker (2004, p. 106) succinctly summarizes the broadly agreed determinants of entrepreneurship,

> The clearest influences on measures of entrepreneurship (usually the likelihood or extent of self-employment) are age, labor market experience, marital status, having a self-employed parent and average rates of income tax (all with positive effects). Greater levels of risk and higher interest rates generally have negative effects, although to date only a handful of studies have satisfactorily investigated the former.

Region specific characteristics such as industry structure (Acs and Audretsch, 1989; White, 1982), unemployment rates (Blanchflower, 2000; Blanchflower and Oswald, 1998), local job layoffs (Storey and Jones, 1987), small business employment (Reynolds et al., 1994) and public policy variables such as state retirement benefits (Blau, 1987), unemployment benefits (Carrasco, 1999), and adherence to welfare state principles (Fölster, 2002) are also found to influence occupational choice.[4]

2.2.2 Labor Markets in Developing Countries

The disadvantage theory and the comparative advantage theory are two competing theories of labor markets in developing countries. The disadvantage theory hypothesizes that people who are rationed out of the formal labor markets are compelled to take up self-employment or work as workers in household enterprises. Such people are considered to constitute the *informal sector*. Thus, beginning with the labor surplus model of Lewis (1954), the labor markets of developing countries are viewed as segmented dualistic markets along the formal-informal lines (also see Sen, 1966; Ranis and Fei, 1961; Harris and Todaro, 1970).[5]

[4] Other examples of studies analyzing the determinants of entrepreneurship include Evans and Jovanovic (1990) and Parker et al. (2005).

[5] Lewis (1954) argued that if wage rate is determined competitively in the rural areas of a LDC then it will be below the subsistence levels. Harris and Todaro (1970) predicts that workers who migrate from rural to urban areas face unemployment and are forced to work in household enterprises at subsistence levels. Models of rural-urban migration following this line of thought hypothesize that the urban informal sector acts as a refuge for migrants and excess labor in urban areas are forced to take up low productivity self employment.

Many studies find evidence against these theories of low level subsisting self-employment in LDCs (Chiswick, 1976; Majumdar, 1981; Blau, 1986; Rosenzweig, 1980; Mohapatra et al., 2007).[6] The comparative advantage theory, thus hypothesizes that individuals voluntarily choose employment in the so called *informal sector*, when they perceive competitive opportunities there (Gindling, 1991; Magnac, 1991; Maloney, 2004).[7]

In this paper, we do not distinguish between the formal and the informal sectors for two reasons. First, Maloney (2004, p.1159) notes that, "as a first approximation we should think of the informal sector as the unregulated, developing country analogue of the voluntary entrepreneurial small firm sector found in advanced countries, rather than a residual comprised of disadvantaged workers rationed out of good jobs." As most empirical research on the determinants of self-employment is based on data from the developed economies, the results will stand comparable to the results of earlier studies if we consider both the sectors together and treat the informal sector akin to the entrepreneurial small firm sector of the developed countries. Second, the other main purpose of the paper is to examine the determinants of self-employment choice in agriculture and nonagriculture in India through the lens of economic geography. Though the characteristics of the informal sector in a developing country are well debated in the literature, examining the determinants of self-employment in this light is an interesting avenue for future research.

2.2.3 Hypotheses: Determinants of Self-employment

Though there are compelling reasons to posit that there are sectoral differences in self-employment choice, male, married and older individuals are more likely

[6] Blau (1985) positively tests for competitive labor markets in the nonagriculture sector in LDCs but finds negative selection into self-employment based on managerial ability in the farm sector. His results suggest that self-employed earn more than wage employees in urban areas whereas in rural areas the self-employed earn much less than the wage employees.

[7] More recently, a growing body of literature attempts to capture the heterogeneity within the informal sector. This strand of literature argues that the informal sector is a blend of both disadvantaged and competitive sectors (Cunningham and Maloney, 2001; Fields, 2005; Günther and Launov, 2006) and claims simultaneous presence of disadvantaged "lower" and voluntary "upper" tiers within the informal sector. Pratap and Quintin (2006) do not find any evidence for segmented labor markets in Argentina. Yamada (1996) finds evidence of voluntary self-selection and higher earnings in self-employment in informal sector in Peru.

to be self-employed in general.[8] The probability for individuals in both agricultural as well as nonagricultural sectors to be self-employed increases with age as individuals accumulate more human capital and resources needed for starting a new venture with time.[9] Some theoretical studies claim that younger workers choose entrepreneurship as they are more likely to enter riskier projects (Johnson, 1978; Jovanovic, 1979; Miller, 1984). However, younger workers may not be able to accumulate capital needed to start a new business. Calvo and Wellisz (1980) argue that individuals acquire managerial skills through learning over time. Older individuals are also more likely to be successful in entrepreneurship. Most empirical evidence suggests a positive relationship between age and entrepreneurship (Evans and Leighton, 1989a; Blanchflower and Meyer, 1994; Blanchflower, 2000). Thus we hypothesize a positive relationship between age and the probability of self-employment.

The empirical literature on the determinants of entrepreneurship suggests that married individuals are also more likely to be self-employed. Borjas (1986) suggests that risk in self-employment reduces if the partner alone works in the business. Moreover, married couple can together raise a greater amount of capital for the start-up and self-employment may appear to be less risky if there is financial support from spouse (Blanchflower and Oswald, 1998). For these reasons, we hypothesize a positive relationship between marriage and the probability to be self-employed.

Empirical evidence on the role of education is mixed. Education increases managerial ability and this leads to a higher probability of entrepreneurship (Lucas, 1978; Calvo and Wellisz, 1980; van Praag and Cramer, 2001). In van Praag and Cramer (2001), education increases entrepreneurial ability and expected entrepreneurial performance. This increases the expected utility of entrepreneurship. However, in a meta analysis of studies linking education and entrepreneurship, Sluis et al. (2005) find that more educated workers become salaried employees. They further find that relative to farming, more educated workers choose nonfarm entrepreneurship. Bates (1990) finds that start-ups by highly educated people are more likely to survive and owner educational background is a signif-

[8] By sector, we refer to the broad sectors of agriculture and nonagriculture here. See Le (1999) for a survey of empirical studies on self-employment.

[9] Goedhuys and Sleuwaegen (2000), for instance, find this to be true for individuals in Côte d'Ivoire.

icant determinant of the financial capital structure of small business start-ups. Thus, there is no consensus in the literature on the effect of education. While education expands the knowledge base of an individual and makes him alert to new opportunities, education also increases the opportunity cost of being self-employed. While Rees and Shaw (1986), Taylor (1996), Blanchflower and Meyer (1994) and Blanchflower (2000) find positive effects of education on self employment, Evans and Leighton (1989b) and Evans and Jovanovic (1989) find no significant effects and Blanchflower et al. (2001) find negative effects of education on the probability of selecting self-employment. Thus, educated individuals may not be willing to take the risks associated with entrepreneurship.

Iyigun and Owen (1999, pp. 213-215) argue that "entrepreneurial human capital plays an important role in intermediate income countries, whereas professional human capital is relatively more important in richer economies." Under the assumption that entrepreneurship is riskier than providing professional services they show that as an economy develops, individuals invest time in accumulating professional skills through education than accumulating entrepreneurial human capital. In their words,

> As per capita income grows and the payoff to being a professional increases, individuals are less willing to gamble on entrepreneurial ventures. This phenomena occurs even though the expected value of entrepreneurship rises with per capita income. While entrepreneurs in a more developed economy face a clearly better lottery than entrepreneurs in a less developed economy, the price of the lottery ticket-foregone professional earnings-is higher in the developed economy, making individuals less willing to take the bet. . . . when individuals are compensated for their manual labor as well as their aggregate human capital input, skill-biased technological change induces more variability in the entrepreneurial payoff. Thus, as the return to the safe activity increases and the payoffs to the risky activity becomes more variable, human capital accumulators devote more time to schooling and less time to gaining entrepreneurial experience. In essence, individuals in high-income economies with higher wages to professionals have more to lose by gambling on an entrepreneurial venture. In contrast, individuals in low income countries face less variable payoffs to entrepreneurship and a lower return to their investment in professional skills and are therefore more willing to invest in entrepreneurial skills.

This suggests that returns to salaried employment increase faster than returns to entrepreneurship as the per-capita income grows, and this makes individuals more risk averse and decreases their willingness to become entrepreneurs (also see Lucas, 1978). Thus, there are compelling reasons to posit that individuals who are more educated opt for salaried employment relative to self-employment in an LDC context (see Sluis et al., 2005, for a survey). Hence, we hypothesize that individuals with greater human capital might prefer salaried employment as opposed to self-employment.

Another determinant of self-employment that is discussed in the literature is wealth. Wealth possessed by the individuals provides a degree of security for entering self-employment and helps them to ease their credit constraints.[10] As Boháček (2006, p.2196) notes,

> In order not to default on loan contracts, entrepreneurs can borrow only limited amounts secured by collateral. This collateral (accumulated assets) guarantees not only the repayment of the loan but also positive consumption of the entrepreneur in the case of a project's failure. As the financial constraint is endogenously related to a borrower's wealth, entrepreneurship becomes positively correlated with wealth.

Households with very high levels of wealth have a higher propensity to take risk (Carroll, 2000). Hurst and Lusardi (2004) argue that as households with higher levels of wealth have a higher tolerance for risk, they are most likely to be business owners.[11] Blanchflower and Oswald (1998) find that inheritance increases the probability of self-employment. Banerjee and Neuman (1993) argue that wealth distribution determines the occupational structure. For these reasons, we hypothesize a positive relationship between household wealth and the entrepreneurship choice.

Borjas and Bronars (1989) present differences in self-employment rates amongst racial minorities in US. They show that consumer discrimination af-

[10] Lindh and Ohlsson (1996) test if the presence of credit constraints inhibit people from becoming self-employed. Many other studies also find that credit constraints act as barriers to entry of individuals into self-employment (Evans and Jovanovic, 1989; Evans and Leighton, 1989b; Blanchflower and Oswald, 1998).

[11] However, Hurst and Lusardi (2004) find that the relationship between wealth and entrepreneurship is flat over the majority of the wealth distribution. They discover a positive relationship only after the ninety-fifth percentile. They argue that the reason could be that capital needed for a start-up in the United States is relatively low (also see Bhidé, 2000).

fects the earnings of self-employed blacks and other minority communities, making them less likely to select into self-employment relative to whites. Some other studies find that self-employment is higher in minority communities (Clark and Drinkwater, 1998). In an Indian context, the presence of caste system leads us to hypothesize that individuals of the backward classes may have a lesser propensity to be self-employed.

Based on insights from the theory of new economic geography (Krugman, 1991; Fujita and Krugman, 2003), we hypothesize that individuals in neighboring regions exhibit similar occupational preferences and in some neighborhoods individuals are more likely to be self-employed than in others and that this effect is non-linear in shaping economic outcomes over space. The presence of many self-employed people in a wealthy neighborhood may induce others to choose self-employment. Thus, it may have an inducement effect on the local population. People in such regions are likely to be more entrepreneurial and risk loving. However, presence of many self-employed people in poor neighborhoods indicates that dearth of viable employment opportunities compells people to select into self-employment in such neighborhoods.

2.3 Bayesian Semiparametric Methodology

Semiparametric regression technique based on Bayesian P-Splines and geoadditive models is used for the empirical analysis. The methodology allows for the estimation of non-linear effects of the continuous variables and the neighborhood effects of spatial units on the probability of individuals selecting self-employment. A brief outline of the method is presented here.[12]

2.3.1 Geoadditive Models

Let (y_i, x_i, v_i) for i in {1,2,...N} describe a dataset of N observations. Let y_i be the response variable and x_i be a m-dimensional vector of continuous covariates and

[12]This section draws on Lang and Brezger (2004) and Brezger and Lang (2005). This methodology has been applied earlier by Kandala et al. (2001) and Kandala et al. (2002) to examine the determinants of under-nutrition in African countries.

v_i be a vector of categorical variables.[13] Assume y_i are independent and Gaussian with mean $\eta_i = f_1(x_{i1}) + ... + f_p(x_{ip}) + v_i\gamma$, and a common variance σ^2. If f_i are unknown smooth functions of the continuous variables and $v_i\gamma$ corresponds to the parametric part of the regression, the regression model is called the Additive Model or a Semiparametric regressor. Eilers and Marx (1996) use polynomial regression splines that are parameterized in terms of B-Spline basis functions, the P-Splines, in the context of an Additive Model, to estimate the smooth functions within the semiparametric framework. Fahrmeir and Lang (2001a,b) use simple random walk priors in a bayesian version of the Additive Model. Kammann and Wand (2003) introduce Geoadditive models within the Additive Mixed Model framework to deal with unobserved heterogeneity across different spatial units.[14] Furthermore, Lang and Brezger (2004) and Brezger and Lang (2005) generalize the work of Fahrmeir and Lang (2001a,b) and develop the Bayesian version of the P-Spline approach of Eilers and Marx (1996), Bayesian P-Splines.[15]We use these methods in the empirical analysis.

Assume that the unknown functions f_j can be approximated by a l degree spline with equally positioned knots in the domain of x_j (Eilers and Marx, 1996). By writing such a spline in the form of a linear combination of k B-Spline basis functions, B_{jk}, where k is equal to the number of knots plus the degree of the spline, $f_j(x_j) = \Sigma\beta_{jk}B_{jk}$ and, in matrix notation, $\eta = \Sigma X_j\beta_j + V\gamma$. By defining a roughness penalty based on the differences of adjacent B-Spline coefficients, for ensuring smoothness of the estimated functions, the penalized likelihood assumes the form:

$$L = l(y, \beta_1, ..., \beta_p, \gamma) - \lambda_1\Sigma(\triangle^k\beta_1)^2 - ....\lambda_p\Sigma(\triangle^k\beta_p)^2 \tag{2.1}$$

[13]We first present the case of the gaussian response distribution and then show how the family of binomial probit models can be generalized to the family of gaussian response, using a link function.

[14]Generalized Additive Mixed Models (Lin and Zhang, 1999) for cases with unobserved heterogeneity are extensions of Generalized Additive Models (Hastie and Tibshirani, 1990). For an overview of semiparametric regressions, see Fahrmeir and Tutz (2001). Additive Mixed Models in the Bayesian framework have also been considered by Hastie and Tibshirani (2000) and Fahrmeir and Lang (2001a,b) but these approaches do not consider the unobserved heterogeneity, the spatially correlated random effects.

[15]The difference penalties are replaced by Gaussian (intrinsic) random walk priors that serve as smoothness priors for the unknown regression coefficients. A related approach is the Bayesian smoothing splines methodology of Hastie and Tibshirani (2000).

In the Bayesian framework, β_j for $j = 1....p$ and γ are considered as random variables and assigned prior distributions. Independent diffuse priors are assumed for the fixed effects parameters, $\gamma_j \propto const$ for $j = 1....q$. The priors for the coefficients of the non-linear functions, β_j, are obtained by substituting the stochastic analogues of the difference penalties. In case of first differences, a first order random walk and for second differences, a second order random walk are considered. Hence, $\beta_{jk} = \beta_{j,k-1} + u_{jk}$ or $\beta_{jk} = 2\beta_{j,k-1} - \beta j, k - 2 + u_{jk}$ with Gaussian errors $u_{jk} \sim N(0, \tau_j^2)$ and constant diffuse priors for the initial values of β_{j1} and β_{j2}. τ_j^2 controls the smoothness of the fitted function. For Bayesian inference, τ_j^2 are also treated as random variables and simultaneously estimated with the β_j. Highly dispersed inverse gamma priors $IG(a_j, b_j)$ are assigned to the variances τ_j^2.

The geoadditive model is obtained if a spatial effect, $f_{spatial}$, is added to the above predictor. The spatial effect may be split into spatially correlated and uncorrelated effects, $f_{spatial} = f_{str} + f_{unstr} = X_{str}\beta_{str} + X_{unstr}\beta_{untr}$, as the spatial effect may comprise of a component that has strong spatial structure and a component that is only locally present. Following Besag et al. (1991) Markov Random Field (MRF) priors are assumed for the regression coefficients β_{str}. If $s \in 1, ...S$ are pixels of a lattice or regions of a geographical map, the MRF prior is given as,

$$\beta_{str,s} \backslash \beta_{str,u} \sim N(\sum_{u \in \partial_s} \frac{1}{N_s} \beta_{str,u}, \frac{\tau_{str}^2}{N_s}) \tag{2.2}$$

for, $u \neq s$, where, N_s is the number of adjacent regions (pixels) and ∂_s is the neighborhood of s. This prior may be seen as an extension of a first order random walk into two dimensional space. For the second component, β_{unstr}, independent and identically distributed (i.i.d.) Gaussian random priors, $\beta_{unstr}(s) \sim N(0, \tau_{unstr}^2)$, are assumed for s=1.....,S. For τ_{str}^2 and τ_{unstr}^2 inverse gamma priors, $IG(a_{str}, b_{str})$ and $IG(a_{unstr}, b_{unstr})$ are assumed.

Inference is based on the posterior and uses recent Monte Carlo Markov Chain (MCMC) techniques. If α is a vector of the unknown parameters, assuming conditional independence of the parameters, the posterior is given by:

$$\begin{aligned} p(\alpha \backslash y) \propto L(y, \beta_1, ..., \beta_p, \beta_{str}, \beta_{unstr}, \gamma, \sigma^2) \times \prod_{j=1}^{p} (p(\beta_j \backslash \tau_j^2) p(\tau_j^2)) \\ \times p(\beta_{str} \backslash \tau_{str}^2) p(\tau_{str}^2) p(\beta_{unstr} \backslash \tau_{unstr}^2) p(\tau_{unstr}^2) p(\gamma) p(\sigma^2) \end{aligned} \tag{2.3}$$

The probit model in this setting, where y_i assumes only binary values 0 or 1, requires slight modifications of the posterior. Here y_i follows Bernoulli distribution $y_i \sim B(1, \mu_i)$, conditional on the covariates and parameters. The mean $\mu_i = \Phi(\eta_i)$ where Φ is the cumulative normal distribution function. Considering the latent variables, we have, $U_i = \eta_i + \epsilon_i$, with $\epsilon_i \sim N(0, 1)$. By defining $y_i = 1$ if $U_i \geq 0$ and $y_i = 0$ otherwise, the model corresponds to a binary probit model. The new posterior also depends on the extra parameters of the latent variable U_i.

2.3.2 Model Diagnostics

Following Spiegelhalter et al. (2002), the Deviance Information Criteria (DIC) is used as a measure of complexity and fit for model selection. The DIC is defined as the (p. 603) "classical estimate of fit, plus twice the effective number of parameters." The unstandardized deviance is given by $-2log\{p(y\backslash\mu)\}$. Assuming that $f(y)$ as a standardizing term that is a function of the data alone, the classical estimate of fit, $D(\overline{\theta})$ is obtained from $D(\theta) = -2log\{p(y\backslash\theta)\} + 2logf(y)$, by evaluating $D(\theta)$ at the mean of the parameters $\overline{\theta}$. $D(\theta)$ is also referred to as the Bayesian deviance or the saturated deviance. For members of the exponential family with $E(Y) = \mu(\theta)$, $D(\theta)$ is obtained by setting $f(y) = p\{y\backslash\mu(\theta) = y\}$. That is, $D(\theta) = -2log\{p(y\backslash\theta)\} + 2log\{p(y\backslash\mu(\theta) = y)\}$. The measure of the effective number of parameters, p_D, is the difference between the posterior mean of deviance $\overline{D(\theta)}$ and deviance at the posterior means of the parameters $D(\overline{\theta})$. That is, $p_D = \overline{D(\theta)} - D(\overline{\theta})$. Then, $DIC = D(\overline{\theta}) + 2p_D = \overline{D(\theta)} + p_D$. Of the competing models, the specification with the least DIC is selected and reported.

2.3.3 Explaining the Residual Spatial Patterns

Consider estimating the geoadditive model with only the spatial component, in a binary probit setting. In our analysis, this would show the propensity of people to be self-employed in a region. However, when individual characteristics (also called fixed effects) are also introduced into the geoadditive model, the resulting spatial patterns show the residual spatial patterns after these characteristics are controlled for. Thus, the spatial patterns estimated in this paper are the residual spatial patterns, as we simultaneously introduce individual characteristics and the spatial components in the geoadditive framework. These estimated residual spatial patterns can be explained using one of the following econometric approaches. A simple strategy is to regress the mean residual spatial effects on the regional variables. Thus, after estimating the geoadditive model, the total spatial effect of each region is explained by regressing the posterior mean of the estimated spatial residual effect on the regional variables. However, this empirical strategy does not consider the estimated posterior variance of spatial effects. In order to overcome this problem, a discrete choice model of the 95% or 80% spatial effects can be estimated. In this case, a variable is constructed that takes a value of (-1) when the region has a significant negative effect, takes a value of (0) if the effect is insignificant and takes a value of (1) if the effect is significant and positive. This leads to a straightforward multinomial specification. This variable is then regressed on the regional variables. We employ both strategies to examine the determinants of the residual spatial patterns.

2.4 Data

The data used for the analysis is the 60th round employment-unemployment survey of the National Sample Survey Organization (NSSO) of India conducted in 2004. As the focus of the paper is on economically active individuals, we restrict the sample to those who are older than 15 years but younger than 70 years. This reduces the sample size from 303,811 to 204,298.[16] While the principal economic activity of this sample ranges from domestic duties to full time employment (in the form of salaried employment, self-employment, casual labor or unemployment), 17% of the individuals in this sample are engaged in subsidiary activities. For the rest of the analysis, we consider the principal economic activity alone for two reasons. First, all individuals are not engaged in subsidiary activities. Second, as less than one sixth of the entire sample are engaged in subsidiary activities, considering such activities would further complicate the analysis when individuals report as both self-employed and paid employees. Furthermore, the principal economic activity is the activity to which the individuals devote most of their time. For these reasons, we consider only the primary occupation for classifying workers into self-employment and paid employment. Table 2.1 lists the number of

[16]We drop 17 individuals who adhere Zoharastrianism for reasons of consistency with the next chapter.

individuals in different occupational categories. We also drop individuals who are unpaid family workers, students, workers involved in domestic duties, pensioners, those who are unable to work due to disabilities and people who reported to belong to the occupational class 'other'. This reduces the final sample to 88,623 economically active individuals.[17] We thus only consider those who have reported their primary occupation as self-employed (includes own account workers and employers), salaried employees, casual laborers, or unemployed.[18]

The descriptive statistics in Table 2.2 show that 65% percent of the individuals have attended at least primary school, 65% live in rural areas and 40% are in the agricultural sector. Table 2.3 presents the descriptive statistics of self-employed and others in agricultural as well as nonagricultural sectors. Self-employed are older in both sectors. 13% of the self-employed in nonagriculture have university education compared to 3.7% of those who are self-employed in agriculture. A higher proportion of educated individuals are self-employed in agriculture and a higher proportion of educated individuals are salaried employees in nonagriculture.

In the absence of an appropriate measure for wealth, we proxy it using the land-possed by the household. We thus posit that individuals who own large areas of land are more likely to be self employed. While in agriculture, land enables self-employed farming, and this makes people to choose self-employment over other modes of occupation, in the nonagricultural sector, land serves as potential collateral to obtain credit for starting an enterprise.[19]

These descriptive tables also show that more than 50% of individuals in agri-

[17] 21.91% of these individuals are engaged in some subsidiary economic activity but for reasons listed earlier, we only consider the primary occupation in classifying individuals as self-employed workers or paid employees.

[18] We merge the occupations into self-employment and paid-employment for the rest of the analysis in this chapter. In the next chapter, we consider the four occupational categories as distinct classes.

[19] On the one hand, self-employed individuals in agriculture may possess more land as they need it for agricultural purposes. On the other hand, only those who possess land may be able to choose self-employment. Thus, the land possessed is also likely to determine the self-employment status. Hence the problem of endogeneity with respect to land even in the agricultural sector may not be so severe. The dataset has some information on the purchases made on the some durable commodities for some households. However, the information is missing for a number of households and for a number of items in the representative consumption bundle. Hence, we are not in a position to use this data. Furthermore, as income data is not available for the majority of individuals in the sample, we are not able to instrument the land possessed using income data.

culture are self-employed in comparison to a relatively lower proportion in nonagriculture. The presence of agricultural sector in the data poses several problems in analyzing the determinants of self-employment. The farm sector is usually found in rural areas with mainly farmers as self employed individuals. There are compelling reasons to posit that they are different from self-employed individuals in nonagriculture. As some scholars have noted before, the process of economic development reduces participation in farm sector and this induces a bias when analyzing the changes in self-employment rates with time if the agricultural sector is included in the analysis (Parker, 2004).[20] Researchers have usually analyzed the determinants of self-employment only in the non-farm sector in order to get around these problems. As the farm sector is very important in a developing country like India, we also study self-employment in this sector.

2.5 Empirical Analysis

In order to use the entire data set on hand and to make robust inferences on the determinants of self-employment, three different models are estimated.

2.5.1 Aggregate Model

In the first model, participation in the agricultural sector is controlled using a dummy variable. The following semiparametric geoadditive probit model is estimated:

$$\eta = \gamma_{const} + \gamma_{female} + \gamma_{marital\_status} + \gamma_{education\_general} + \gamma_{education\_technical} + \gamma_{wealth} + \gamma_{urban} + \gamma_{agri} + \gamma_{hindu} + \gamma_{backward} + f_{age} + f_{spatial}(district) + f_{random}(district)$$

The non-linear effect of age is modeled as third degree P-Spline with second order random walk penalty.[21] Figure 2.1(a) shows that the probability of being

[20]However, as our study is cross-sectional and does not analyze self-employment rates over time, this limitation does not apply here. Furthermore, we analyze the determinants of self-employment in agriculture and nonagriculture separately.

[21]The number of equidistant knots is assumed to be 20. The structured spatial effects are estimated based on Markov random field priors and random spatial effects are estimated with gaussian priors. The variance component in all the cases are estimated based on inverse gamma priors with hyperparameters a=0.001 and b=0.001. The number of iterations is set to 110000 with burnin parameter set to 10000 and the thinning parameter set to 100. The autocorrelation files and the sampling paths show that the MCMC algorithm has converged. These plots are available from the author.

self-employed increases with age, confirming the age-effect. The derivative of the 'age' function in Figure 2.1(b) indicates that the marginal effect of age on the self-employment choice first increases, drops and then increases very rapidly for individuals older than 55 years. The rise in the 50s is consistent with the findings of empirical literature on developed countries (Blanchflower and Meyer, 1994; Blanchflower, 2000) that older individuals are more likely to be self-employed.[22] As Fuchs (1982, p.356) claims: "Men who change to self-employment late in life are primarily those who have had previous experience in self-employment or who are in wage-and-salary occupations such as managers or salesmen that have many characteristics similar to self-employment." The self-employed continue to work even after the retirement age when the salaried employees stop. This leads to over-sampling of older self-employed, and could be a reason for the jump at 55. It is also possible that switches to self-employment reflects a partial-retirement effect, as salaried workers switch to self-employment instead of dropping from the labor force towards the end of the life cycle (Quinn, 1980).

The results of the parametric part of this regression model, also referred to as fixed effects, in Table 2.4, suggest that both married and divorced people are more likely to be self-employed compared to unmarried individuals.[23] Marriage reduces entrepreneurial risk if the spouse is economically active. It also provides an additional unpaid family worker for the household enterprises. It is also possible that marriage gives additional money in the form of dowry, which can enable start-up activity.[24] The positive coefficients of the education variables of informal and school education suggest that lower levels of education are positively related to self-employment. The negative coefficient of the variable 'University', however, suggests that higher education decreases the probability of self-employment. The Indian education system allows students to choose between technical education at professional colleges or general education at universities after high school. Students who are successful in competitive exams are selected to join the technical institutions primarily consisting of the engineering, medical and agricultural colleges. They also have an option to do diploma courses that are usually shorter

[22]Retirements effects are also associated with this phenomena. However some studies (Blau, 1987; Evans and Leighton, 1989b; Evans and Jovanovic, 1989) do not find significant effects of age on self-employment.

[23]This is consistent with Taylor (1996), Fairlie and Meyer (1996) and other studies that find positive effects of marital status on self-employment.

[24] Though dowry is legally prohibited in India, it is prevalent in numerous forms.

in duration than technical degree courses. People with technical education may choose to be self-employed as their professional training enables this possibility. For this reason, we introduce technical education dummies in the estimation, with "having no technical education" as the base variable. The results suggest that the effect of having technical degree is insignificant and having a technical diploma is negative and significant at the 5% level. This is possibly because the foregone professional earnings for individuals with a technical degree is much higher than for those with a diploma.[25] The results also suggest that Hindus and members of backward castes are less likely to be self-employed. This remarkable observation is analyzed in greater detail in the next chapter. The probability to be self-employed also increases with the wealth of the individual's household, proxied here by the land possessed. However, this result should be interpreted with a degree of care, as land is potentially endogenous with respect to occupation.[26] We keep the land variables as there are compelling reasons to assume that wealth determines the entrepreneurial choice, in the Indian context.[27]

The map of structured spatial effects in Figure 2.2(a) shows the presence of strong spatial effects and a clear north-south divide in the probability of self-employment choice. This is confirmed by Figures 2.2(c) and 2.2(d) that plot the 95% and 80% confidence intervals for the estimated structured spatial effect that show presence of neighborhood effects that spill over district as well as state boundaries. The local unstructured random effects in Figure 2.2(b) are very small compared to the structured effects.[28] While people in the northern states of Uttar Pradesh and Bihar have a higher likelihood to be self-employed, people in southern regions are less likely to be so. In order to shed more light on these spatial patterns, sector specific models are estimated.

[25]When self-employed are separated into those who are only self-employed and those who employ others in a multinomial setting, it is found that education is positively related to employers while it is still negative for the self-employed. There are only very few employers in the database and the results are available from the author.

[26]In the absence of a good instruments for wealth, we do simple probit estimations with and without the land variables to check if the land variable adversely affects the coefficients of the other variables, but we do not find such evidence. We also do a hausman test to test for changes in coefficients of other variables.

[27]One of the primary reasons for keeping these indicators of household wealth is that there is evidence of the financial institutions rationing credit to individuals who are able to provide collateral. This indicates that wealth should strongly predict the self-employment choice as lack of finance is one of the biggest obstacles to being self-employed.

[28]The structured spatial effects range from -0.8 to 0.8, the random unstructured local spatial effects range from -0.2 to 0.18.

2.5.2 Sector Specific Models

Agricultural and Nonagricultural Self-employment

The first model assumes that the determinants of self-employment are same for all self-employed individuals in agricultural as well as nonagriculture. In order to examine the differences in the two sectors, the following semiparametric model is estimated for individuals in agricultural and nonagricultural sectors separately:

$\eta = \gamma_{const} + \gamma_{female} + \gamma_{marital\_status} + \gamma_{education\_general} + \gamma_{education\_technical} + \gamma_{wealth} + \gamma_{urban} + \gamma_{hindu} + \gamma_{backward} + f_{age} + f_{spatial}(district) + f_{random}(district)$

The parameters for a, b, the number of iterations, burnin, and the thinning parameter are set equal to the first model's parameters.[29] The relationship of age with self-employment is very close to being linear in the agricultural sector, as seen in Figure 2.1(e), while in the nonagricultural sector, as Figure 2.1(c) shows, the age function increases at a decreasing rate until the age of 55 years and then increases at an increasing rate. Table 2.5 and Table 2.6 show considerable differences in relative human capital endowments of self-employed individuals in the two sectors. While in the agricultural sector, those who are endowed with higher levels of human capital (proxied by age and education) are more likely to be self employed, in the nonagricultural sector such individuals are more likely to be salaried employees. Belonging to a backward class is significantly negatively related to being self-employed in both the sectors, and being a Hindu has a significant negative relationship only in nonagriculture.

For people in nonagriculture, as maps in Figure 2.3 suggest, the north-south divide seen in the spatial effect on the self-employment choice for individuals in the aggregate model is less pronounced. People of Kerala and some districts of Tamil Nadu in the south, Maharastra and Madhya Pradesh in western and central parts of India, and the majority of districts in the north-eastern states are less likely to be self-employed. People living in Uttar Pradesh, Bihar, Rajasthan, some districts of Andhra Pradesh, and West Bengal are more likely to be self-employed.

The maps of spatial effects in agriculture in Figure 2.4 show that the result of north-south spatial divide observed in the first model can be attributed mainly to such a phenomenon in the agricultural sector. In sharp contrast to some districts in the western and the northern parts of India, people are very less likely to be

[29] The autocorrelation files and plots of the sampling paths show that sufficient convergence is achieved in these models also.

self-employed in agriculture in southern and central states. As Figures 2.3(b) and 2.4(b) demonstrate, the unstructured random effects are negligible compared to the structured spatial effects. The confidence interval plots for the random spatial effects also show that the local effects are small and insignificant compared to the effects of structured spatial effects in all the three estimated models.[30]

2.5.3 Determinants of Residual Spatial Patterns

The presence of spatial patterns, as shown by the empirical analysis, suggests that it is not just personal characteristics of individuals that totally explain their occupational choice. As discussed below, regional characteristics also play an important role in determining self-employment choice. In particular, financial constraints, level of economic development, unemployment and small business employment are found to influence the self-employment rates in a region by earlier studies. Hence, we hypothesize that these variables can explain the residual spatial patterns. We follow the empirical approach described in subsection 2.3.3.

Holtz-Eakin et al. (1994) test the role of liquidity constraints in the formation of new enterprises. Their analysis suggests that the size of inheritance has an effect on entrepreneurial choice and also on investment in the capital of a new enterprise. Many studies find that credit constraints are barriers to entry for individuals into self-employment (Evans and Jovanovic, 1989; Evans and Leighton, 1989b; Blanchflower and Oswald, 1998). Lindh and Ohlsson (1996) test for the presence of credit constraints as inhibitors to self-employment, by seeing if those who win a lottery are more likely to enter self-employment. They also find that such individuals start firms with higher capital. Cabral and Mata (2003) find that the presence of binding financial constraints inhibit firms from growing to their optimal size. Hence, we hypothesize that the level of financial development in the region, measured by the per-capita credit or the credit-deposit ratio in a district can explain the residual spatial pattern.

Lucas (1978) predicts that entrepreneurship decreases with economic development. Calvo and Wellisz (1980) show that the growth rate of total stock of knowledge requires greater ability of the marginal entrepreneur in a steady state equilibrium. This suggests that, given a fixed ability distribution in a population, the number of entrepreneurs decreases and average firm size increases with tech-

[30]These plots are available from the author.

nological progress. Empirical studies of Acs et al. (1994) and Fölster (2002) find that per-captia gross net product (GNP) is negatively related to self-employment. Acs et al. (1994) argue that self-employment decreases in the early stages of development as technological change shifts output from agriculture and small scale industry to large scale manufacturing. We thus hypothesize that level of economic development determines the propensity to be self-employed in a region.

Cross-sectional evidence gives a mixed impression about the effect of unemployment on the propensity to be self-employed. The recession-push hypothesis claims that high unemployment decreases the probability of getting paid employment and thus pushes individuals into self-employment. However, the prosperity-pull hypothesis suggests that high unemployment reduces demand for goods and services of the self-employed, leading to a reduction in self-employment. Many cross-sectional studies find a negative relationship between unemployment and the probability of self-employment (Taylor, 1996; Blanchflower and Oswald, 1998). However, many studies also indicate that the self-employed experience a spell of unemployment (Evans and Leighton, 1989b; Blanchflower and Meyer, 1994). As Storey (1991) notes, time series studies show a positive relationship but cross-sectional studies suggest a negative relationship. Hence we hypothesize that unemployment could explain the residual self-employment pattern.

We also introduce a number of demographic controls. In particular, we control for size of the district and the population density. Armington and Acs (2002) suggest that these factors play an important role in explaining the spatial patterns of new firm formation. We also control for agglomeration, measured by the density of firms in the region, as presence of a large number of firms in the neighborhood is likely to result in spillovers that induce new firm formation. As Krugman (1991, p. 484) notes, "the concentration of several firms in a single location offers a pooled market for workers with industry-specific skills, ensuring both a lower probability of unemployment and a lower probability of labor shortage." Furthermore, as Armington and Acs (2002, p.38) argue, "informational spillovers give clustered firms a better production function than isolated producers have. The high level of human capital embodied in their general and specific skills is another mechanism by which new firm start-ups are supported." Thus regions with high agglomeration are more likely to be associated with higher probability of people entering self-employment.

In Table 2.7 the determinants of spatial variation are estimated using the above set of regional indicators. The dependent variable is the estimated mean residual spatial effect in the district, after controlling for individual characteristics. In Table 2.8 and Table 2.9, we estimate multinomial logit models with the dependent variable as the estimated 95% spatial effects in the maps in Figure 2.2, Figure 2.3 and Figure 2.4. Thus the dependent variable takes value (-1) if the effect is significantly negative (black areas in the maps), (0) if the value is insignificant (grey areas) and (1) if the value is significantly positive (white areas). In Table 2.8, we use per-capita credit as a proxy for financial development and in Table 2.9, we use the credit-deposit rate as a proxy for financial development of the region.

The coefficient of the first proxy for financial development in Table 2.7, per-capita credit, is insignificant in agriculture as well as nonagriculture. The coefficient of the second proxy, the credit-deposit ratio, is significant and positive in nonagriculture and negative in agriculture. It is also seen that level of economic development, measured by the per-capita net state domestic product, is negatively related to the probability of self-employment in both sectors. These observations support the claim of Acs et al. (1994) that technological change shifts output from agriculture and small scale industry to large scale manufacturing, resulting in a decrease in self-employment. However, unemployment appears to increase self-employment in nonagriculture, but is negatively related to self-employment in agricultural sector. Thus, we find evidence of a "push" effect in nonagriculture and a "pull" effect in the agricultural sector.[31] Size of district and population density also have a similar relationship with the residual spatial pattern of self-employment. While they increase the probability of self-employment in the nonagricultural sector, they lower it in the agricultural sector. This is plausible as a highly dense region induces people into nonagricultural self-employment for reasons listed above. The negative sign in the agricultural sector may be referring to the lesser availability of per-capita land that is an important determinant of self-employment in this sector. The agglomeration index is insignificant in the agriculture and the nonagriculture equations.

[31] It is also possible that the measure of unemployment rate we use leads to this result. The unemployment rate in a district is constructed as the proportionate number of people in the district who have registered with the unemployment office. People registered with the unemployment office are mostly educated individuals looking for employment. In the absence of data on unemployment, we proxy it using this measure.

The R-squared in the model explaining determinants of self-employment in agricultural sector is 0.16 when the per-capita credit is included as a measure of financial development and 0.22 when the credit-deposit ratio is included as a measure of financial development. However, the R-squared in the models explaining the determinants of self-employment in the nonagricultural sector is 0.40 in both models. This suggests a better fit for the nonagricultural sector. This may be because the independent variables mostly measure trends that are more relevant to the nonagricultural sector.[32] However, these results should be interpreted carefully as they are based on the estimated mean residual spatial effect, and do not consider the variance.

The multinomial logit estimation of the 95% significant spatial effects in Table 2.8 and Table 2.9 suggest that neither per-capita credit nor credit-deposit ratio have a significant positive effect on self-employment. However, they confirm most of the above results. The interpretation of the results is straightforward. For example, in Table 2.8 it can be seen that an increase in the per-capita net state domestic product decreases the probability of a region to be significant positive effect region (white) and increases the probability to be a significant negative effect region (black) in Figure 2.3(c). Similarly, the positive effect of unemployment vanishes in the nonagricultural sector in the multinomial estimations. This shows that the results of Table 2.7 should be interpreted carefully as they are based only on the posterior mean of the estimated residual spatial effect.

In summary, the analysis suggests that while economic development has a significant negative effect on self-employment, financial development has no effect, when other factors are controlled for.

2.5.4 Self-employment in Rural and Urban Areas

The data used in the earlier analysis consists of individuals in rural and urban areas. This is essential as we estimate the spatial effects and the individual effects jointly in the geoadditive framework. Considering individuals of only urban or only rural regions would be incorrect because the spatial component is modeled as a continuous variate. Hence, we estimate a binary probit model, for examining

[32] The measure of agglomeration index, for instance, is more likely to explain the spatial pattern of self-employment in the nonagricultural sector than the spatial pattern in the agricultural sector.

the determinants in urban and rural areas. We control for regional effects using a set of state level regional dummies. We estimate this for the sub-sample of individuals in the nonagricultural sector alone.[33] We also check the robustness of the estimates, with respect to the presence of land variables, by running separate regressions with and without land variables. We estimate the regressions with the land variables excluded in the first specification and land variables included in the second specification (Table 2.10). However, the regression estimates for the two specifications are not very different. It can be argued that in the Indian context, wealth plays a definite role in self-employment choice. As argued earlier, this is possible if credit is rationed in favor of individuals possessing assets such as land. We interpret the results of the specification with the land variables, as Table 2.10 suggests that the estimates of models with and without them are similar.

The results are broadly consistent with results of the semi-parametric estimation. The estimated signs of higher education variables are negative in rural as well as urban areas. The absolute value of the coefficients are, however, slightly higher in the rural areas suggesting that educated people in the rural areas have a still lower propensity for self-employment. The returns to self-employment in rural areas may be lower in comparison to the returns to self-employment in urban areas and this could explain this result. This issue is analyzed more extensively in chapter 5. While technical education is insignificant in rural estimations, it is significant and negative in urban regressions. The land variables are positive and increase the propensity to be self-employed in rural and urban areas. However, the coefficients are larger in urban areas, indicating that people in urban areas with more land have a higher propensity to choose self-employment. This may be because land in urban areas is more expensive relative to land in rural areas. This has a direct implication for obtaining credit from financial institutions. The estimates of the religion and caste variables are consistent with the semi-parametric model for the nonagricultural sector estimated earlier and the coefficients are significant and negative. The absolute value of the coefficient of the 'Hindu' variable is larger in the urban regression than in the rural regression equation. This is counter intuitive to some degree, because cultural institutions responsible for lower likelihood of Hindus and individuals of backward classes to be self-employed are expected to be stronger in rural areas. A plausible explanation is that individ-

[33] As the agricultural sector is mostly found in the rural areas only, we restrict the urban-rural analysis to the nonagricultural sector.

uals of other religions face greater discrimination in urban areas when it comes to wage-employment. Thus the probability of Hindus entering wage-employment may be higher in urban areas.

2.6 Conclusion

The field of entrepreneurship in economics provides insights into the individual determinants of the self-employment choice in developed countries. We contribute to one aspect of this literature that remained neglected for a long time. We use recent advances in Bayesian semiparametric methodologies to examine the spatial as well as individual determinants of self-employment choice in a developing country, India. Consistent with studies based on datasets from developed countries, we find age to have a non-linear relationship with the probability to be self-employed, particularly in nonagriculture. A clear jump after the age of 55 is noticed, which could be a direct result of the retirement effect. The effect is linear and monotonically increasing in agriculture. Married individuals are more likely to be self-employed in both sectors. In nonagriculture, educated people are less likely to be self-employed while in agriculture, they are more likely. The results are consistent with empirical studies of developed economies and also shed light on the unexplored agricultural self-employment in a developing country context. The analysis further suggests that in the nonagriculture, self-employed people are more or less uniformly distributed across different spatial units but in agriculture self-employed individuals are concentrated in certain geographic pockets. In both sectors, the regions with the highest propensity of self-employment are the states of Uttar Pradesh and Bihar. While it can be argued that these regions are more entrepreneurial, these regions are also the poorest regions in India, in terms of per-capita income and human development. This leads to an important conclusion that self-employment in Indian context may actually support the view that self-employment in a fast growing economy like India continues to be the main occupational option in the poorest neighborhoods and not for individuals with high human capital. Furthermore, an analysis of the determinants of nonagricultural self-employment in rural and urban areas suggests that in rural areas educated individuals have still lower propensity to become self-employed.

Table 2.1: Distribution of Occupation

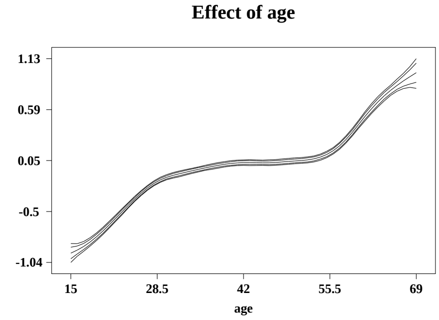
	Total Number	Percentage	Cumulative
Self-employed(Own Account Workers)	37,197	18.21	18.21
Self-employed(Employers)	922	0.45	18.66
Household Helpers (Unpaid Family Worker)	23,505	11.51	30.16
Salaried Employees	21,223	10.39	40.55
Casual Labor(Public)	310	0.15	40.70
Casual Labor(Other)	23,823	11.66	52.36
Unemployed	5,148	2.52	54.88
Students	25,853	12.65	67.54
Only Domestic Duties	40,894	20.02	87.56
Domestic Duties and Collection of Wood etc.	18,045	8.83	96.39
Pensioners	2,645	1.29	97.68
Not working due to disability	1,381	0.68	98.36
Beggars and Prostitutes	3352	1.65	100
Total	204,298	100	

Table 2.2: Agricultural and Nonagricultural Sectors (Descriptives)

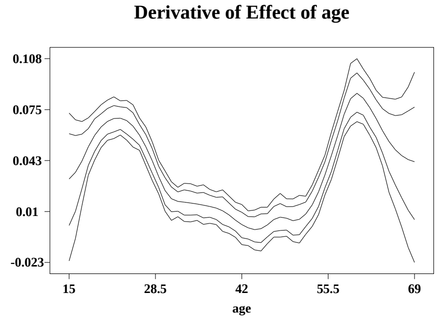
Variable	All	Non-Agri	Agri
Selfemployed	0.430	0.341	0.564
Age	37.130	34.910	40.464
Age (Std. Dev)	12.88	12.08	13.31
Male	0.809	0.837	0.767
Female	0.191	0.163	0.233
Unmarried	0.210	0.278	0.108
Married	0.745	0.690	0.828
Divorced	0.045	0.032	0.064
No Education	0.256	0.135	0.437
Informal Education	0.085	0.066	0.115
Primary School	0.310	0.319	0.298
High School	0.227	0.294	0.126
University	0.122	0.186	0.024
No Technical Education	0.948	0.919	0.991
Technical Degree	0.009	0.014	0.001
Technical Diploma	0.043	0.067	0.007
Rural	0.649	0.453	0.943
Urban	0.351	0.547	0.057
Agriculture	0.400	0.000	1.000
Land $<$ 0.2 Hectares	0.214	0.295	0.093
0.2$<$ Land $<$0.4 Hectares	0.461	0.526	0.362
0.4$<$ Land $<$2 Hectares	0.241	0.137	0.396
Land $>$ 2 Hectares	0.084	0.041	0.149
Hindu	0.792	0.777	0.815
Backward	0.676	0.612	0.772
N	88623	53202	35421

Table 2.3: Sector Specific Self-employment (Descriptives)

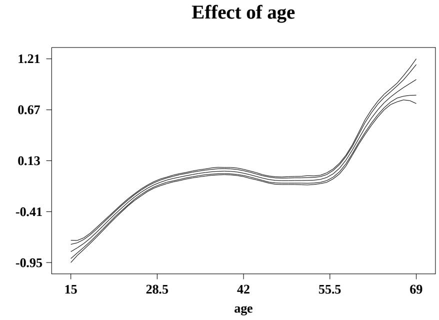
	Nonagri Employee	Nonagri Self-employed	Agri Employee	Agri Self-employed
Age	33.304	38.015	35.542	44.266
Age (Std. Dev)	11.712	12.189	12.639	12.552
Male	0.812	0.885	0.634	0.870
Female	0.188	0.115	0.366	0.130
Unmarried	0.335	0.168	0.168	0.062
Married	0.636	0.795	0.772	0.871
Divorced	0.029	0.037	0.060	0.068
No Education	0.125	0.154	0.553	0.348
Informal Education	0.057	0.083	0.107	0.121
Primary School	0.301	0.353	0.250	0.335
High School	0.304	0.276	0.083	0.159
University	0.213	0.134	0.008	0.037
No Technical Education	0.906	0.946	0.995	0.988
Technical Degree	0.016	0.009	0.001	0.002
Technical Diploma	0.078	0.044	0.005	0.010
Rural	0.450	0.460	0.954	0.935
Urban	0.550	0.540	0.046	0.065
Land $<$ 0.2 Hectares	0.311	0.265	0.188	0.020
0.2$<$ Land $<$0.4 Hectares	0.508	0.560	0.576	0.197
0.4$<$ Land $<$2 Hectares	0.139	0.134	0.208	0.541
Land $>$2 Hectares	0.041	0.041	0.028	0.242
Hindu	0.790	0.751	0.875	0.769
Backward	0.626	0.584	0.856	0.707
N	35064	18138	15440	19981



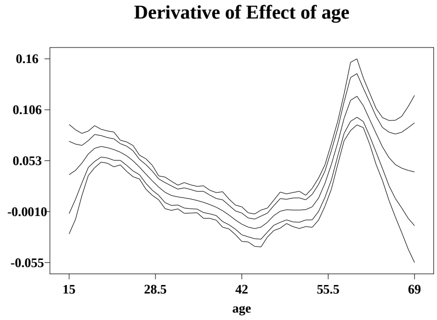
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in the Aggregate Model.



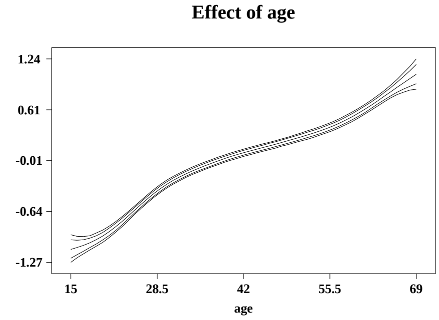
(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in the Aggregate Model.



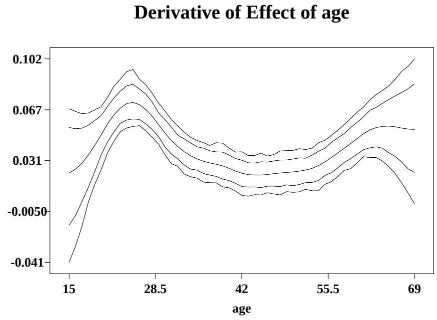
(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture.



(d) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Nonagriculture.



(e) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Agriculture.



(f) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Agriculture.

Figure 2.1: Non-linear Effects of Age on Self-employment

Table 2.4: Determinants of Self-employment

Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Personal Characteristics				
Female	-0.398	0.014	-0.426	-0.372
Married	0.175	0.018	0.141	0.211
Divorced	0.317	0.029	0.259	0.376
General Education				
Informal	0.265	0.019	0.227	0.304
Primary School	0.332	0.014	0.304	0.360
High School	0.193	0.016	0.163	0.224
University	-0.181	0.020	-0.218	-0.141
Technical Education				
Technical Degree	-0.127	0.057	-0.232	0.016
Technical Diploma	-0.117	0.026	-0.168	-0.068
Land Possessed				
0.2$<$ Land $<$0.4 Hectares	0.149	0.014	0.120	0.176
0.4$<$ Land $<$ 2 Hectares	0.791	0.017	0.758	0.824
Land $>$ 2 Hectares	1.180	0.024	1.132	1.226
Location				
Urban	0.253	0.013	0.227	0.279
Agriculture	0.336	0.013	0.312	0.361
Religion & Social Group				
Hindu	-0.205	0.014	-0.233	-0.179
Backward	-0.183	0.012	-0.206	-0.160
Constant	-0.545	0.027	-0.599	-0.492
N	86140			
Deviance(Mean)	93422.587			
Std. Dev.	36.196992			
deviance($\bar{\mu}$)	92973.92			
pD	448.66642			
DIC	93871.253			

Notes: Dependent variable is binary self-employment status of the individual. Base categories for marital status, general education, technical education, land dummies are unmarried, no general education, no technical education and less than 0.2 hectares of land respectively.

Table 2.5: Determinants of Self-employment in Nonagriculture

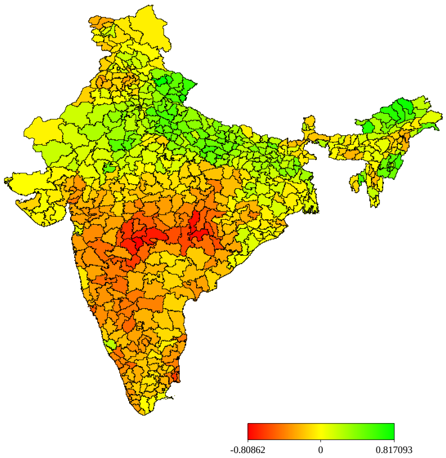
Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Personal Characteristics				
Female	-0.256	0.018	-0.290	-0.221
Married	0.203	0.019	0.165	0.240
Divorced	0.218	0.042	0.137	0.298
General Education				
Informal	0.141	0.028	0.085	0.195
Primary School	0.130	0.021	0.086	0.169
High School	-0.039	0.022	-0.078	0.004
University	-0.349	0.024	-0.395	-0.301
Technical Education				
Technical Degree	-0.109	0.057	-0.217	0.004
Technical Diploma	-0.134	0.025	-0.183	-0.084
Land Possessed				
0.2< Land <0.4 Hectares	0.151	0.015	0.122	0.181
0.4< Land < 2 Hectares	0.112	0.022	0.070	0.153
Land > 2 Hectares	0.160	0.033	0.097	0.222
Location				
Urban	0.029	0.015	0.001	0.059
Religion & Social Group				
Hindu	-0.180	0.016	-0.213	-0.149
Backward	-0.150	0.014	-0.179	-0.121
Constant	-0.222	0.031	-0.282	-0.163
N	51674			
Deviance(Mean)	60166.724			
Std. Dev:	34.978124			
deviance($\bar{\mu}$)	59807.524			
pD	359.20045			
DIC	60525.925			

Notes: Dependent variable is binary self-employment status of the individual. Base categories for marital status, general education, technical education, land dummies are unmarried, no general education, no technical education and less than 0.2 hectares of land respectively.

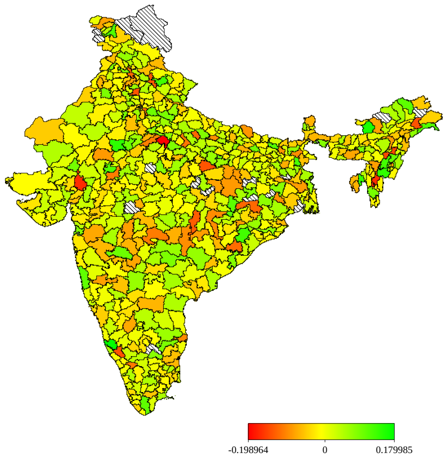
Table 2.6: Determinants of Self-employment in Agriculture

Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Personal Characteristics				
Female	-0.540	0.027	-0.594	-0.487
Married	0.206	0.042	0.122	0.288
Divorced	0.447	0.058	0.336	0.558
General Education				
Informal	0.233	0.032	0.164	0.296
Primary School	0.435	0.025	0.387	0.484
High School	0.758	0.035	0.689	0.827
University	0.862	0.076	0.722	1.018
Technical Education				
Technical Degree	0.157	0.274	-0.377	0.702
Technical Diploma	0.193	0.114	-0.034	0.413
Land Possessed				
0.2$<$ Land $<$0.4 Hectares	0.533	0.042	0.443	0.614
0.4$<$ Land $<$ 2 Hectares	1.986	0.042	1.903	2.074
Land $>$ 2 Hectares	2.787	0.050	2.686	2.892
Location				
Urban	0.459	0.044	0.378	0.543
Religion & Social Group				
Hindu	-0.015	0.035	-0.083	0.054
Backward	-0.286	0.027	-0.339	-0.235
Constant	-1.031	0.064	-1.155	-0.908
N	34466			
Deviance(Mean)	22493.237			
Std. Dev:	35.860231			
deviance($\bar{\mu}$)	22042.36			
pD	450.87693			
DIC	22944.114			

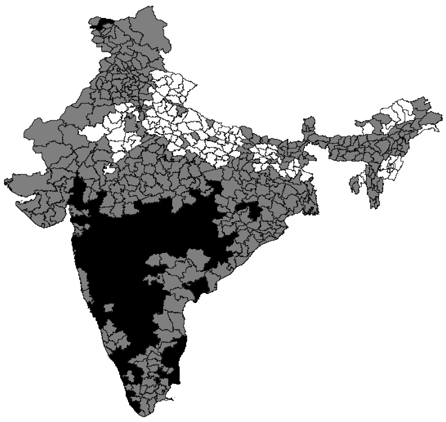
Notes: Dependent variable is binary self-employment status of the individual. Base categories for marital status, general education, technical education, land dummies are unmarried, no general education, no technical education and less than 0.2 hectares of land respectively.



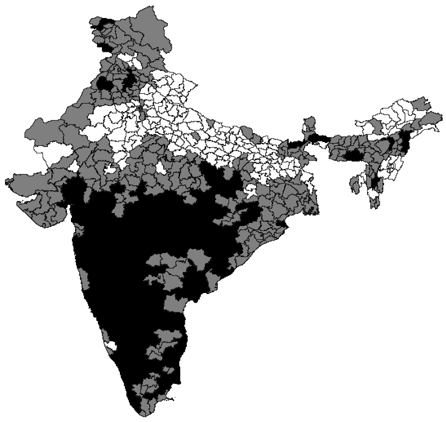
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(b) Unstructured Random Effect of 'District'. Shown are the posterior means.

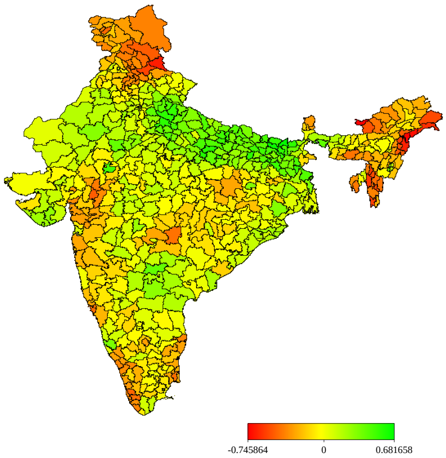


(c) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

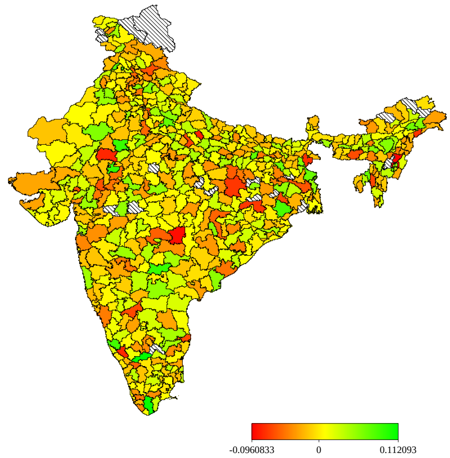


(d) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

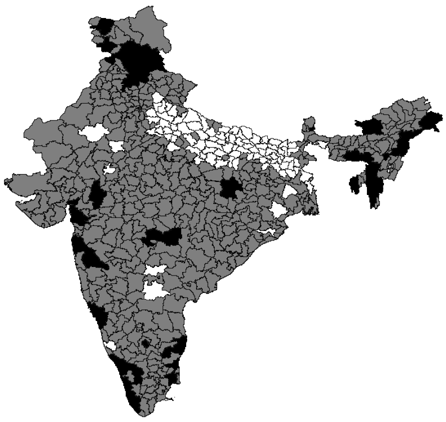
Figure 2.2: Spatial Effects on Self-employment Choice



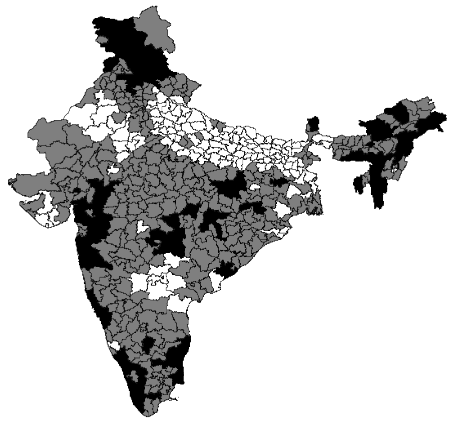
(a) Structured Non-linear Effect of 'District'. Shown are the posterior means.



(b) Unstructured Random Effect of 'District'. Shown are the posterior means.

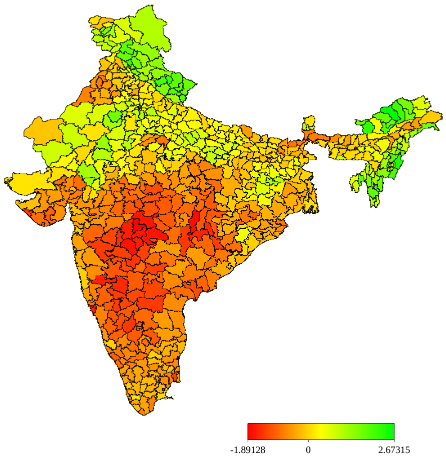


(c) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

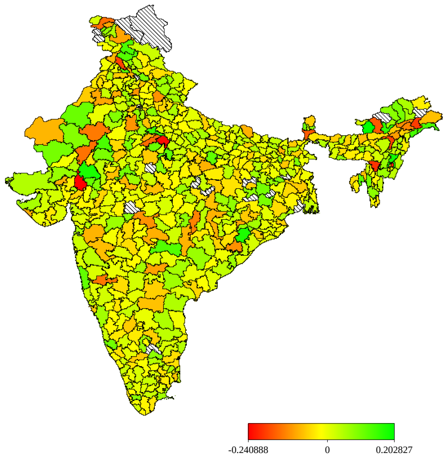


(d) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

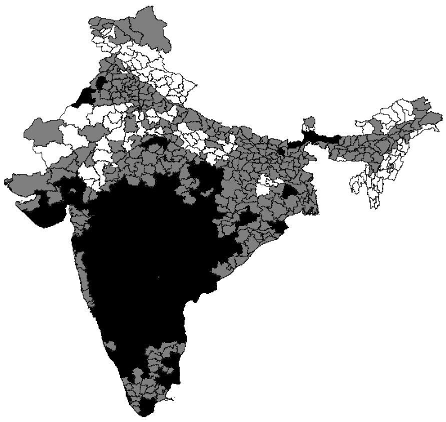
Figure 2.3: Spatial Effects in 'Nonagriculture'



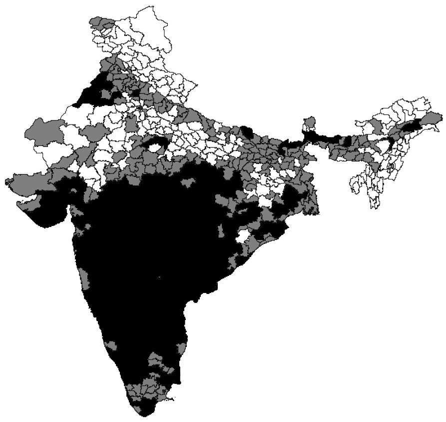
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(b) Unstructured Random Effect of 'District'. Shown are the posterior means.



(c) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(d) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

Figure 2.4: Spatial Effects in 'Agriculture'

Table 2.7: Determinants of Spatial Patterns in Figure 2.2, Figure 2.3 and Figure 2.4

	All		Nonagriculture		Agriculture	
Financial Development						
Per-capita Credit	0.00622		-0.0183		0.0275	
	(0.016)		(0.012)		(0.044)	
Credit-Deposit Ratio		-0.102***		0.0436**		-0.402***
		(0.023)		(0.018)		(0.061)
Economic Development						
Per-Capita NSDP	-0.310***	-0.268***	-0.291***	-0.325***	-0.418***	-0.253***
	(0.034)	(0.030)	(0.026)	(0.023)	(0.091)	(0.078)
Unemployment	-0.0603***	-0.0471***	0.0406***	0.0369***	-0.291***	-0.239***
	(0.016)	(0.015)	(0.012)	(0.012)	(0.041)	(0.040)
Demographics						
Mid Size District	0.00325	0.0141	0.0869***	0.0819***	-0.191**	-0.147*
	(0.030)	(0.029)	(0.022)	(0.022)	(0.079)	(0.076)
Large District	0.0280	0.0305	0.0750	0.0719	-0.176	-0.161
	(0.091)	(0.090)	(0.068)	(0.068)	(0.24)	(0.23)
Population Density	-0.0189	-0.0183	0.0594***	0.0554***	-0.129***	-0.129***
	(0.015)	(0.014)	(0.011)	(0.011)	(0.040)	(0.037)
Agglomeration Index						
Firm Density	-0.00767	-0.00213	-0.00290	-0.00874	-0.0179	0.00453
	(0.013)	(0.012)	(0.0094)	(0.0090)	(0.033)	(0.031)
Constant	2.926***	2.534***	2.618***	2.749***	4.280***	2.778***
	(0.35)	(0.36)	(0.27)	(0.27)	(0.95)	(0.93)
Observations	534	534	531	531	532	532
R^2	0.20	0.23	0.40	0.40	0.16	0.22
F	19.08	22.46	49.43	50.30	14.11	21.42
R^2 Adjusted	0.192	0.220	0.390	0.394	0.147	0.212

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is the mean spatial effect per district after estimation of the geoadditive models.

Table 2.8: Determinants of 95% Spatial Patterns in Figure 2.2, Figure 2.3 and Figure 2.4

	All			Nonagriculture			Agriculture		
	-ve black	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*
Financial Development									
Per-Capita Credit	-0.0149	0.0651**	-0.0501**	0.0262	0.0269	-0.0531***	0.0113	-0.00676	-0.00453
	(0.026)	(0.029)	(0.025)	(0.017)	(0.023)	(0.019)	(0.030)	(0.031)	(0.025)
Economic Development									
Per-Capita NSDP	0.425***	-0.140**	-0.285***	0.272***	-0.0459	-0.226***	0.282***	-0.122*	-0.161***
	(0.056)	(0.062)	(0.052)	(0.041)	(0.054)	(0.042)	(0.062)	(0.062)	(0.053)
Unemployment	0.0401*	0.0249	-0.0650***	-0.0570***	0.0380	0.0190	0.147***	-0.0196	-0.128***
	(0.024)	(0.028)	(0.023)	(0.017)	(0.025)	(0.022)	(0.030)	(0.030)	(0.023)
Demographics									
Mid Size District	0.0970**	-0.189***	0.0924*	-0.0487	-0.118**	0.167***	0.175***	-0.165***	-0.0103
	(0.049)	(0.052)	(0.048)	(0.030)	(0.051)	(0.048)	(0.055)	(0.051)	(0.044)
Large District	0.00916	-0.231	0.222	-0.0208	-0.135	0.155	0.445***	-0.310***	-0.135
	(0.14)	(0.15)	(0.19)	(0.079)	(0.18)	(0.18)	(0.12)	(0.086)	(0.085)
Population Density	-0.0115	0.0221	-0.0106	-0.00672	-0.0686***	0.0753***	-0.0855***	0.166***	-0.0807***
	(0.023)	(0.027)	(0.023)	(0.017)	(0.024)	(0.020)	(0.029)	(0.032)	(0.023)
Agglomeration Index									
Firm Density	0.0300	-0.0519**	0.0220	0.0131	-0.0282	0.0151	0.0158	-0.0368	0.0210
	(0.020)	(0.022)	(0.019)	(0.014)	(0.018)	(0.014)	(0.022)	(0.023)	(0.019)
Observations	534			531			532		
Log Likelihood	-490.2			-357.4			-504.2		
χ^2 (14)	140.0			265.7			129.2		
Pseudo R^2	0.125			0.271			0.114		

Notes: Marginal effects after multinomial logit estimation. *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the 95% spatial effect of the district, after estimation of the geoadditive models.

Table 2.9: Determinants of 95% Spatial Patterns in Figure 2.2, Figure 2.3 and Figure 2.4

	All			Nonagriculture			Agriculture		
	-ve black	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*
Financial Development									
Credit-Deposit Ratio	0.143***	0.00945	-0.152***	-0.0790***	0.100***	-0.0212	0.233***	-0.0656	-0.168***
	(0.037)	(0.043)	(0.036)	(0.030)	(0.039)	(0.030)	(0.046)	(0.047)	(0.038)
Economic Development									
Per-Capita NSDP	0.368***	-0.0897	-0.278***	0.316***	-0.0333	-0.283***	0.225***	-0.115**	-0.110**
	(0.050)	(0.055)	(0.047)	(0.039)	(0.053)	(0.041)	(0.057)	(0.057)	(0.047)
Unemployment	0.0232	0.0159	-0.0390*	-0.0500***	0.0154	0.0345	0.124***	-0.0183	-0.106***
	(0.024)	(0.028)	(0.023)	(0.017)	(0.025)	(0.022)	(0.030)	(0.031)	(0.023)
Demographics									
Mid Size District	0.0707	-0.177***	0.106**	-0.0316	-0.153***	0.185***	0.143**	-0.157***	0.0138
	(0.049)	(0.052)	(0.049)	(0.033)	(0.052)	(0.049)	(0.056)	(0.053)	(0.046)
Large District	0.00133	-0.211	0.210	-0.0152	-0.217	0.232	0.441***	-0.321***	-0.119
	(0.14)	(0.15)	(0.19)	(0.079)	(0.20)	(0.21)	(0.13)	(0.089)	(0.092)
Population Density	-0.0160	0.0364	-0.0204	0.00542	-0.0591**	0.0537***	-0.0731***	0.157***	-0.0841***
	(0.022)	(0.025)	(0.022)	(0.017)	(0.023)	(0.018)	(0.028)	(0.031)	(0.023)
Agglomeration Index									
Firm Density	0.0198	-0.0350*	0.0152	0.0223*	-0.0250	0.00264	0.0113	-0.0313	0.0200
	(0.018)	(0.021)	(0.018)	(0.014)	(0.018)	(0.013)	(0.021)	(0.022)	(0.019)
Observations	534			531			532		
Log Likelihood	-481.0			-357.8			-486.7		
$\chi^2(14)$	158.4			265.0			164.2		
Pseudo R^2	0.141			0.270			0.144		

Notes: Marginal effects after multinomial logit estimation. *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the 95% spatial effect of the district, after estimation of the geoadditive models.

Table 2.10: Self-employment in Nonagriculture

Rural and Urban Regressions

	Model I		Model II	
Independent Var.	Rural	Urban	Rural	Urban
Personal Characteristics				
Age	0.0298***	0.0332***	0.0294***	0.0335***
	(0.0052)	(0.0049)	(0.0052)	(0.0049)
Age Square	-0.0224***	-0.0229***	-0.0221***	-0.0239***
	(0.0064)	(0.0059)	(0.0065)	(0.0060)
Female	-0.232***	-0.275***	-0.231***	-0.276***
	(0.027)	(0.024)	(0.027)	(0.024)
Married	0.252***	0.298***	0.255***	0.302***
	(0.028)	(0.027)	(0.028)	(0.027)
Divorce/Widow	0.376***	0.250***	0.380***	0.268***
	(0.061)	(0.053)	(0.061)	(0.053)
General Education				
Informal Education	0.175***	0.0874**	0.170***	0.0799**
	(0.038)	(0.040)	(0.038)	(0.040)
Primary School	0.159***	0.0759***	0.155***	0.0614**
	(0.027)	(0.028)	(0.027)	(0.029)
High School	-0.0540*	-0.0248	-0.0567**	-0.0510*
	(0.028)	(0.029)	(0.029)	(0.030)
Diploma/University Education	-0.410***	-0.278***	-0.412***	-0.317***
	(0.036)	(0.032)	(0.036)	(0.032)
Technical Education				
Technical Degree	0.168	-0.211***	0.164	-0.220***
	(0.12)	(0.063)	(0.12)	(0.063)
Technical Diploma	0.0251	-0.205***	0.0262	-0.208***
	(0.042)	(0.033)	(0.042)	(0.033)
Household Controls				
0.2< Land <0.4 Hectares			0.117***	0.166***
			(0.027)	(0.018)
0.4< Land <2 Hectares			0.0603**	0.226***
			(0.030)	(0.043)
Land >2 Hectares			0.113***	0.344***
			(0.041)	(0.066)
Hindu	-0.128***	-0.237***	-0.128***	-0.238***
	(0.024)	(0.020)	(0.024)	(0.020)
Backward	-0.117***	-0.157***	-0.119***	-0.157***
	(0.021)	(0.018)	(0.021)	(0.018)
Total Observations	23916	28611	23895	28589
Log Likelihood	-14191	-16930	-14169	-16865
LR (χ^2)	2472	2685	2492	2789
Degrees of freedom	47	47	50	50
Pseudo R^2	0.0801	0.0735	0.0808	0.0764

Notes: Probit estimation. *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is 'selfemployed'. State dummies are included in all the regressions and are not reported here. The coefficients of the constant are not reported.

Chapter 3

Religion and Entrepreneurship

While considerable concern has emerged about the impact of religion on economic development, little is actually known about how religion impacts the decision making of individuals. This chapter examines the influence of religion on the decision for people to become an entrepreneur. Based on a large-scale data set of nearly ninety thousand workers in India, this chapter finds that religion shapes the entrepreneurial decision. In particular, some religions, such as Islam and Christianity, are found to be more conducive to entrepreneurship than Hinduism. In addition, the caste system is found to influence the propensity to become an entrepreneur. Individuals belonging to a backward caste exhibit a lower propensity to become an entrepreneur. Thus, the empirical evidence suggests that both religion and the tradition of the caste system influence entrepreneurship, suggesting a link between religion and economic behavior.

3.1 Introduction

Religion and economics have had a tenuous relationship. On the one hand, scholars dating back at least to Adam Smith and Max Weber have argued that religion plays a fundamental role in shaping economics.[1] On the other hand, only scant attention has recently been given as to how and why religion might influence economics. The omission of religion as a determinant of economic activity is startling, given the recent suggestion by Iannaccone (1998, pp. 1492) that "the economics of religion will eventually bury two myths - that of homo economicus as a cold

[1] Anderson (1988, p. 1068) notes, "In Wealth, Smith was not interested in theological issues or even in the nature of religious belief. Rather, he was concerned with two basic problems: (1) the economic incentives involved in the individual's decision to practice religion and (2) the economic effects of different systems of religious belief as reflected in individual behavior. He did not attempt to develop an economic theory of the emergence of religious beliefs... Smith attempted the more limited task of defining the logical economic consequences of certain kinds of religious beliefs."

creature with neither need nor capacity for piety, and that of homo religiosus as a benighted throwback to pre-rational times." Moreover, as Edmund Phelps argues, "values and attitudes are as much a part of the economy as institutions and policies are. Some impede, others enable."[2]

In India, for instance, Hinduism is strongly associated with the emergence of the caste system. Although some aspects of the caste system such as untouchability, were abolished by the government, it remains formidable and imposing in practice. There remains a heated public debate in India on the impact of the caste system on the economic status of what is widely referred to "*backward classes*". For example, in an article announcing, "Indian College Quota Law Suspended", *The New York Times* reports that, "Caste discrimination is outlawed but continues to persist in obvious and subtle ways, and the contest over the latest university admissions quotas revolve around how to best redress an entrenched and often ugly social bias."[3]

Recent studies suggest the existence of a relationship between religion and economic performance (Barro and McCleary, 2003; McCleary and Barro, 2006b; Guisa et al., 2006). For example, Barro and McCleary (2003) estimate the impact of adherence to religious beliefs on economic performance using international survey data on religiosity. They find that increases in church attendance tend to reduce economic growth while increases in the belief in hell and an afterlife increase economic growth. These empirical findings raise several important but unanswered questions: (1) What are the channels by which religion influences economic activity? and (2) Is the impact of religion on economic activity homogeneous across all religions?

The purpose of this paper is to shed light on these questions by examining whether religion has any impact on one particular channel of economic decision-making influencing economic growth – the decision to become an entrepreneur. Recent studies suggest that entrepreneurship may be a key factor generating growth and development (Baumol, 2002). As Lazear (2002, p. 1) concludes, "The Entrepreneur is the single most important player in a modern economy." Lazear's conclusion is supported by considerable theoretical and empirical literature linking entrepreneurship to economic growth.[4]

---

[2]"It's All About Attitude," *Newsweek International Edition*, 30 April, 2007.

[3]"India College Law Suspended," *The New York Times*, 29 March, 2007.

[4]See for example the studies by Holtz-Eakin and Kao (2003) and Audretsch et al. (2006).

In particular, this paper links the decision of people in India to start a business to their religion as well as their caste status. What this paper does not at all address is whether India, or any other country for that matter, needs more or less entrepreneurship. Rather, the focus of this paper is on the impact of religion on the economic decision making process of individuals.

This paper consists of five sections. The following section discusses the link between religion, culture and entrepreneurship in the Indian context and posits that both religion and culture will influence the decision to become an entrepreneur. The third section describes our data set, which consists of a large sample of individuals. The fourth section presents our empirical analysis testing the hypotheses that both religion and culture influence economic behavior. The final section provides a summary and conclusion. In particular, the empirical evidence suggests that both religion and the cultural tradition of the caste system influence economic behavior, and in particular the decision to become an entrepreneur.

3.2 Religion, Entrepreneurship and the Indian Context

Scholars have generally framed the decision of an individual (*homo oeconomicus*) to become an entrepreneur in terms of the model of occupational choice, where the income generated from entrepreneurship is compared to the wage earned as an employee (Lucas, 1978; Kihlstrom and Laffont, 1979; Holmes and Schmitz Jr., 1990; Parker, 2004; Jovanovic, 1994).

A broad spectrum of individual-specific characteristics, ranging from risk aversion (Kihlstrom and Laffont, 1979), to personality attributes (McCelland, 1964), to education and human capital (Zucker et al., 1998; Bates, 1990; Blanchflower and Meyer, 1994) and unemployment (Evans and Leighton, 1989a) are found to influence individuals' entrepreneurship choice. Thus, an important strand of research has emerged trying to identify why some individuals choose to start a new business while others abstain from entrepreneurship.[5]

Why should religion influence the decision of an individual to become an entrepreneur? Iannaccone (1998, p. 1475) concludes that "At the level of individuals and households, economic behavior and outcomes do correlate with religion." However, to our best knowledge there are no studies, with the exception of Minns and Rizov (2005), that have considered the role that religion plays in shaping the entrepreneurial decision.[6] Yet, there are compelling reasons to posit that religion does influence an individual's decision to become an entrepreneur.

Eisenstadt (1968, p. 10), for instance, emphasizes the importance of the "transformative potential" of a religion for economic motives and activities. By "transformative potential", he means the "capacity to legitimize, in religious or ideological terms, the development of new motivations, activities, and institutions which were not encompassed in their original impulses and views." Moreover, he postulates that "the transformative potential of a given religion is greater the stronger is the emphasis in it on transcendentalism, on individual responsibility and activism, on an open unmediated relationship between the individual and the sacred tradition with the concomitant possibility of its continuous redefinition and reformulation, and a high degree of social openness among the religiously active groups" (Eisenstadt, 1968, p. 20). Hence, it can be argued that religions with great transformative potential may facilitate entrepreneurial behavior. Conversely, those religions with a low transformative potential may inhibit entrepreneurship.

There are also compelling reasons to posit that religion will influence economic behavior in the Indian context. The main religions of South Asia are Hinduism, Islam, Christianity, Buddhism, Jainism and Sikhism. Given that Buddhism and Sikhism have historical links with Hinduism and majority of South Asians are Hindus, the Hindu religion may influence the choice to become an entrepreneur

[5]As Parker (2004, p. 106) notes "The clearest influences on measures of entrepreneurship (usually the likelihood or extent of self employment) are age, labor market experience, marital status, having a self-employed parent and average rates of income tax (all with positive effects). Greater levels of risk and higher interest rates generally have negative effects, although to date only a handful of studies have satisfactorily investigated the former." Lazear (2005) argues that entrepreneurs do not excel in one skill but are competent in many.

[6]Minns and Rizov (2005) use 1901 census of Canada to historically link religion and self employment at the beginning of the 20th century. They find that Canadian Jews were more entrepreneurial than Catholics. They also find that "Catholics were only somewhat less likely to be self-employed than Church Protestants, and no meaningful difference is apparent between mainstreem Protestants and members of other sects."[p. 275]

in India. Compared to the other main religions of India, Hinduism provides little encouragement or value to change one's situation in terms of material well being (Singer, 1966). According to Uppal (2001, p. 20), "The people of South Asia are deeply religious and all facets of their lives including their endeavors to achieve material advancement are affected greatly by religious beliefs and values."[7]

According to Hinduism every human being is *Amrutasya Putraha*, a child of immortality and a spark of divinity. The purpose of life is to attain liberation which essentially is freedom from re-birth and the chain of cause and effect. One should live to understand reality and not for transitory material pursuits.

Dharma Righteousness, *Artha* Earnings, *Kama* Desire, *Moksha* Liberation are supposed to guide the lives of Hindus. The scriptures ordain individuals to follow righteousness, perform duties and earn their livelihood, satisfy their desires and finally seek liberation. *Dharma, Artha, Kama, Moksha* can also be interpreted differently: one should righteously earn his livelihood and desire only for liberation (also referred to as self-realization). An individual has to do his duty as dictated by the scriptures and should not loose himself in material pursuits.

Varna refers to classification of individuals into different classes, categories or castes. Historically Hindus were classified into four major castes. Initially their occupation determined their caste and caste affiliation akin to the religious identity was passed on to their progeny. Brahmins were scholars, priests, advisors to kings, intelligentsia of the community. *Kshatriyas* were kings and noblemen. Their duties involved protection of the community from enemies and administration. Traders, businessmen and entrepreneurs were *Vyshyas* and people of all other occupations were classified as *Shudras*. Thus the *Varna* System that initially categorized individuals into different classes persisted across generations and later determined the occupations of Hindus to a great extent.

In his third major work on the sociology of religion, Weber (1958, pp. 103-104) states that "If the stability of the caste order could not hinder property differentiation it could at least block technological change and occupational mobility, which from the point of view of caste were objectionable and ritually dangerous." In summary, he claims that the impact of caste system on the economy is essentially negative (Medhora, 1965).

In one of the few studies analyzing the effects of the caste system, Munshi and

[7] Uppal (2001) also provides an excellent overview of the philosophy of Hinduism.

Rosenzweig (2006) examine the influence of the caste within the context of an educational choice model in Bombay. They find that lower caste boys are more likely to study in schools where the medium of instruction is the local language and not English. This is very likely to lead them into traditional occupations as defined by the caste structure. Munshi and Rosenzweig (2006, p. 1230) note, "caste networks might place tacit restrictions on the occupational mobility of theirs members to preserve the integrity of the network" and "although these restrictions might have been welfare enhancing and indeed equalizing when they were first put in place, such restrictions could result in dynamic inefficiencies when the structure of the economy changes."

The clear demarcation of occupations based on castes, the persistence of occupation decisions across generations and the other tenets that entail Hindus not to live a life of material pursuits, lead us to hypothesize that these factors might continue to influence the occupational choices of Hindus, and in particular inhibit the propensity to become an entrepreneur. We have no strong predictions how other religions in India, like Islam or Christianity, might influence an individual's entrepreneurial decision. It is likely, however, that the impact of the caste system on economic behaviors is stronger for Hindus as compared to non-Hindus.

In the following sections we will analyze whether Hinduism, as well as belonging to a lower caste, will influence the propensity to become an entrepreneur.

3.3 Data

The main source of data to link religion and caste affiliation to entrepreneurship is the National Sample Survey Organization (NSSO) of India. We use the NSSO's 60th round Employment-Unemployment Survey. This household level survey was conducted in 2004. Almost three hundred thousand individuals in sixty thousand households were questioned about their economic status, religious affiliation and personal background. The households were selected based on a stratified sampling methodology. Since the focus of this paper is on economically active individuals, we only consider those who have reported to be: self employed (includes own account workers and employers), salaried employees, casual laborers and unemployed. For similar reasons, we restrict our sample to those who are older than 15 years but younger than 70 years. We thus exclude from our analysis family members who assist household enterprises, such as children and the elderly, as well as people classified into other miscellaneous occupational categories. These individuals can also be located according to their region. The final sample consists of 87,181 individuals.

Table 3.1 provides the means and standard deviations of the independent variables. 79% of the final sample are Hindus, 11.2% are Muslims, 5.6% are Christians, 1.4% are Sikhs, 0.3% are Jains, 1% are Buddhists and 1.1% are individuals of other religions or without religion. This roughly corresponds to the distribution of religion within the overall population of India.[8] 66.5% of Jains in the sample are self-employed, 50.4% of Christians and 48.6% of Muslims, 41% of Hindus and Sikhs and 38% of Buddhists. (Figure 3.1 and Table 3.2).

Individuals included in the database are also classified according to class affiliation. They belong to either one of the three backward classes (Schedule Castes, Schedule Tribes, Other Backward Classes) or to the forward castes. 12.5% of the sample belong to schedule castes, 18% to schedule tribes, 36.8% to other backward classes. These three classes combine to account for 67.5% of the entire sample. It should be emphasized that although the caste system is a distinct feature of Hinduism and the Constitution of India (Schedule Castes) Order, 1950 notes that, "no person who professes a religion different from the Hindu, the Sikh or the Buddhist religion shall be deemed to be a member of a Scheduled Caste", almost 66% of Christians are classified in the Schedule Caste. As Table 3.3 suggests, the other religions also have a share of their population that claims to be backward. While in Christianity this may be the result of conversion of individuals of the lower castes of Hinduism, in other religions this possibly reflects the economic backwardness rather than social backwardness. The presence of caste system, a characteristic of Hinduism, is also reflected in other religions in India. Within Islam certain sects are considered to be nobler than others. In Christianity, converts from lower castes of Hindu society are treated as lower caste members of Christianity. We cannot rule out conversions into Christianity giving rise to this phenomena. Also, we cannot rule out the possibility of the caste system diffusing into other religions in India.

When we examine class based occupational behavior specifically in Hinduism,

[8]According to the 2001 Census, the religious composition of population in India is as follows: 80.9% are Hindus, 12.9% are Muslims, 2.4% are Christians, 1.9% are Sikhs, 0.4% are Jains, 0.8% are Buddhists, and 0.7% are others. See Premi (2004, p. 4294).

we find that there is a lower representation of schedule caste and schedule tribe individuals in the self employed category and a far higher representation in the casual laborer category (Figure 3.2).

3.4 Empirical Analysis: Discrete Choice Models

In order to test the hypotheses that both religion, and in particular Hinduism, as well as membership in a lower caste, have a negative impact on entrepreneurship, we estimate multinomial probit models of occupational choice.[9] Individuals are either self employed, or salaried, or casual laborer or unemployed.

In the first model (Table 3.6), the effect of religion on self employment is isolated by controlling for a number of variables that are likely to influence the probability of self employment such as age, gender, marital status, educational background, land possessed, rural or urban location. The results show that Hindus are less likely to be self employed compared to individuals of other religions. In particular, the probability of becoming self employed is 8.6% less for Hindus.

The control variables are generally consistent with results already well established in the literature. As has been commonly found, the evidence suggests a quadratic relationship between age and the probability to become an entrepreneur.[10] In addition, both married and divorced people are more likely to be entrepreneurs compared to unmarried individuals.[11] There is not much consensus on the effect of eduction in the literature.[12] These results for India suggest

[9]We do not make use of the multinomial logit model as the Hausman-test suggets that its basic assumption, independence from irrelevant alternatives (IIA), is violated.

[10]This is consistent with the findings of empirical literature on developed countries that older individuals are more likely to be self employed. Evans and Leighton (1989a), Blanchflower and Meyer (1994), Blanchflower (2000) and many other studies find a positive and quadratic effect of age on the probability of becoming self-employed; however Blau (1987), Evans and Leighton (1989b), Evans and Jovanovic (1989) do not find significant effects of age on self-employment.

[11]Consistent with Taylor (1996) and Fairlie and Meyer (1996) and others who find positive effects of marital status on self-employment.

[12]Education expands the knowledge base of an individual and makes him alert to new opportunities. Rees and Shaw (1986), Taylor (1996), Evans and Leighton (1989a), Blanchflower and Meyer (1994), Blanchflower (2000) and others find positive effects of education on self employment. However, education also increases the opportunity cost of being self employed. Educated individuals may not be willing to take the risks associated with entrepreneurship. For instance, while Evans and Leighton (1989b) and Evans and Jovanovic (1989) find no significant effects Blanchflower et al. (2001) and Georgellis and Wall (2000) find negative effects of education on the probability of selecting self employment.

that increases in education reduce the probability of self employment in the Indian context. The effect is non-linear with individuals having lowest levels of education showing a higher-propensity to be self-employed. The effect is positive for low levels and becomes negative for those with university education.

The negative coefficients on the variable Hinduism suggest that religion does, in fact, influence the decision to become an entrepreneur; however these results do not shed much light on the channels through which such inhibition might take place. Thus in Table 3.7, we include a dummy variable reflecting membership in backward class along with personal characteristics, educational background and regional factors. As explained earlier, the class structures of Hinduism have had considerable influence on the formation of class structures in other religions in India. The results presented in Table 3.7 suggest that individuals in the backward classes of all the religions are less likely to be entrepreneurs. Further, as the negative coefficient on the variable reflecting the Hindu religion, Hindus are still less likely to be entrepreneurs compared to individuals of other religions even after controlling for the class structure.

The strong presence of class structures within Hinduism leads us to posit that Hindus of all classes, forward as well as backward, might have a lower propensity to become an entrepreneur than do individuals of other religions. As mentioned earlier, amongst Hindus, only the *Vyshyas* are expected to do business. Thus, the impact of being both a Hindu and a member of different classes on the decision to become an entrepreneur was estimated and the results are presented in Table 3.8. The evidence suggests that an individual who is both a Hindu as well as a member of the backward class *scheduled caste* is almost 14% and backward class *scheduled tribe* is 19% less likely to be self employed. Hindus belonging to the other backward classes are 5.7% and forward castes as well are 2.2% less likely to be entrepreneurs relative to the individuals of the other religions. This confirms our hypothesis that the class structures of Hindusim are binding and continue to influence their occupational choice, particularly with respect to becoming an entrepreneur.

In contrast, the probability of being a salaried employee is higher for Hindus, irrespective of the class as compared to non Hindus. The positive effect of being a member of a backward class might be explained by the reservation system established in India by the government that supports Hindus belonging to backward class but not members of other religions. One might therefore argue that the reservation system enables Hindu backward class to favor salaried employment instead of self employment whereas members of other religions choose self employment. However, the values of estimated marginal effects suggest that the positive coefficients for salaried employment category are negligible compared to the negative coefficients in the self-employment category. This suggests that the effect of caste system in inhibiting Hindus from selecting self-employment is significant. In fact, the backward class Hindus have a higher propensity to be casual laborers.

In order to focus on the impact of caste system we estimate the model based on the sample of Hindus only (Table 3.9) . The strong presence of class structures within Hinduism leads us to posit that Hindu individuals belonging to the backward class might have a lower propensity to become an entrepreneur than Hindus belonging to the forward class. Thus, the impact of both religion and caste system, by being both a Hindu and a member of the backward class on the decision to become an entrepreneur was estimated and the results are presented in Table 3.9. The evidence suggests that a Hindu who is a member of the backward class *scheduled caste* is almost 14.6% and backward class *scheduled tribe* is 18% less likely to be self employed than a forward class Hindu.

The four estimated models confirm our hypotheses that Hindus are less likely to be entrepreneurs than are individuals of other religions. This leads us to the last question. How does the propensity to enter into entrepreneurship compare between the non-Hindu and the Hindu religions? Thus, the results included in Table 3.10 take Hinduism as the base class and show the marginal effect on the probability to be self employed for individuals of other religions. The results suggest that Muslims are 7.9%, Christians 2.9%, and Jains 27% more likely to be self employed compared to Hindus. By contrast, individuals of other minor religions and those without religion are almost 13.4% more likely to be entrepreneurs compared to Hindus. Buddhists and followers of Sikhism are pretty much in the same boat as Hindus.

As a further check of the robustness of the results, we estimate a model by considering the self-employed separated as employers and only self-employed people (Table 3.11). It is startling to observe that the coefficients of the Hindu variable and the backward class variable are significant and negative even for the employer group. This suggests that the Hindus have a lesser propensity to be entrepreneurs.[13] An important qualification of the results is that the self-employed includes both agricultural and non-agricultural self-employed people. However, when the sample is restricted to non-agriculture (Table 3.12), the results confirm that there is virtually no difference.

It is important to note that minority communities are associated with higher self-employment rates even in the developed countries (Clark and Drinkwater, 1998). However, the insight from our analysis is that even when we consider the Hindus alone, the caste system has an effect on the propensity to be self-employed. This supports our theory that the caste-system continues to exert an influence on the occupational choice of Hindus.

3.5 Conclusion

Religion is rarely attributed to shaping economic phenomena. So it is with the decision to become an entrepreneur. While a rich and robust literature has emerged identifying a number of important characteristics and factors alternatively conducive to or impeding entrepreneurship, religion has been noticeably absent.

The results of this paper suggest that religion matters. While India is rich with diverse religions, some of them, such as Islam and Christianity, are conducive to entrepreneurship. By contrast, others, and in particular Hinduism, inhibit entrepreneurship. We control for regional specific effects by introducing state level dummies and the results are robust to these controls as well.

Similarly, the caste system is found to influence the propensity to become an entrepreneur. In particular, belonging to a backward caste inhibits entrepreneurship. The least entrepreneurial people tend to be Hindus in the lower class. One reason for this might be the long shadow of caste system that persists and limits the freedom of occupational choice to some extent not only to all individuals of backward classes but to Hindus in particular.

Hence, the results of this paper suggest that elements of religion and the caste system need to be explicitly considered in understanding what influences important economic phenomena, such as entrepreneurship. Just as religion plays

[13] The marginal effects are very small but this could partly be attributed to the very small number of employers in the sample.

a major role in influencing entrepreneurial activity, so too does the caste system. At least in the case of India, Max Weber's insight is found to hold - religion is an important influence on economic behavior.

It may be fruitful for future research to consider not just the impact of religion on economic activity, such as entrepreneurship, but also the conditioning effect of the particular locational context. One clue about the importance of location is provided by the results of studies showing that Indian and other Asian immigrants in the United Kingdom and North America actually exhibit a greater propensity for entrepreneurship (Clark and Drinkwater, 1998). While the specific religion of the immigrants is not explicitly identified, the inhibiting impact of a specific religion and particular caste may, in fact, disappear along with the change in location and institutional context. Without the painstaking future research, however, such a conjecture will remain simply that, a conjecture.

Figure 3.1: Entrepreneurship and Religion

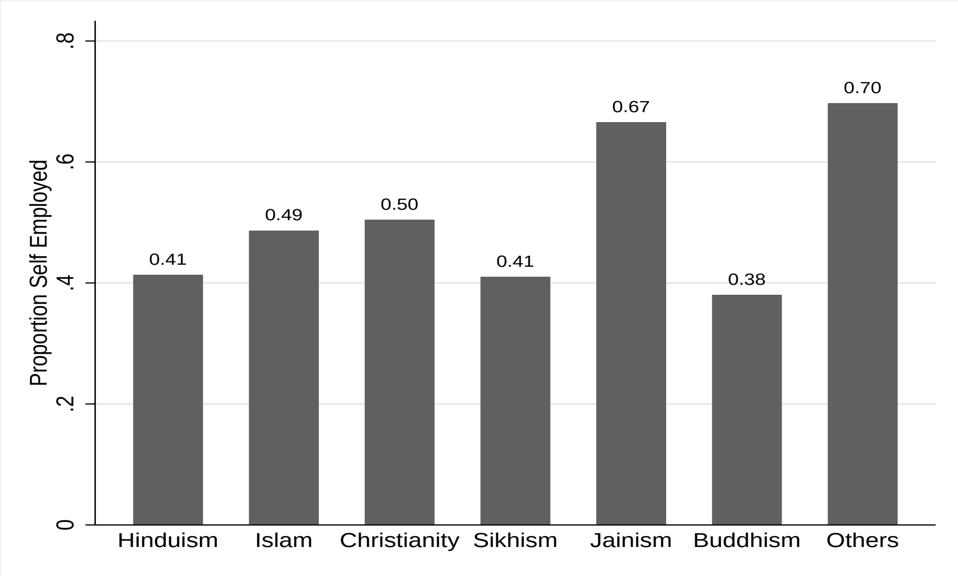
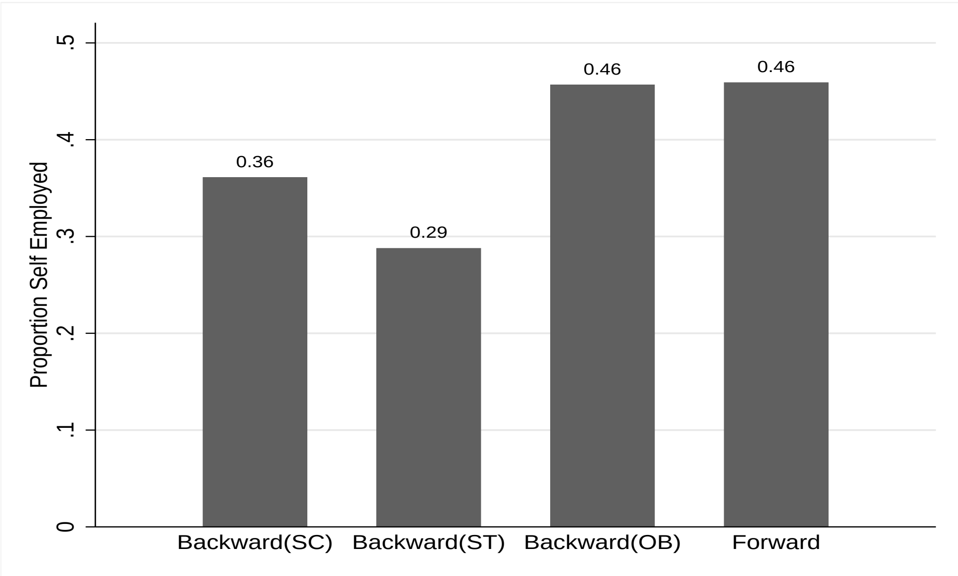


Figure 3.2: Entrepreneurship and Caste System in Hinduism



For explanation on SC, ST, OB see notes of Table 3.1

Table 3.1: Means and Standard Deviations

Variable	Mean	Standard Deviation
Self Employed	0.43	0.50
Salaried	0.24	0.43
Casual Labor	0.27	0.45
Unemployed	0.06	0.23
Hinduism	0.79	0.41
Islam	0.11	0.32
Christianity	0.06	0.23
Sikhism	0.01	0.12
Jainism	0.003	0.05
Buddhism	0.01	0.10
Other Religions	0.01	0.11
Backward Caste (SC)	0.13	0.33
Backward Tribe (ST)	0.18	0.39
Backward Others (OB)	0.37	0.48
Backward Class	0.68	0.47
Forward Caste	0.32	0.47
Age	37.13	12.88
Male	0.81	0.39
Female	0.19	0.39
Unmarried	0.21	0.41
Married	0.74	0.44
Divorced	0.04	0.21
No Education	0.26	0.44
Informal Education	0.09	0.28
Primary	0.31	0.46
High School	0.23	0.42
University Diploma/Degree	0.12	0.33
No Technical Education	0.95	0.22
Technical Degree	0.01	0.09
Technical Diploma	0.04	0.20
Rural	0.65	0.48
Urban	0.35	0.48
Land ($>$0.4 & $<$ 2 Hectares)	0.24	0.42
Land ($>$ 2 Hectares)	0.08	0.27

Notes: Individuals of backward classes belong to one of the three categories: Scheduled Castes(SC), Scheduled Tribes(ST) and Other Backward Classes(OB). The variable 'Backward' is all the three categories together.

Table 3.2: Religion and Occupational Choice (Descriptives)

Religion	Self Employed	Salaried Employee	Casual Labor	Unemployed	Total
Hinduism	41.30	23.90	28.99	5.81	100
Islam	48.62	20.92	24.28	6.17	100
Christianity	50.43	30.01	13.58	5.98	100
Sikhism	41.00	30.53	22.2	6.26	100
Jainism	66.54	28.08	4.23	1.15	100
Buddhism	37.97	26.00	32.15	3.88	100
Others	69.69	16.45	9.70	4.16	100
Total	43.01	23.95	27.23	5.81	100

Table 3.3: Religion and Caste System (Descriptives)

Religion	Backward Caste(SC)	Backward Tribe(ST)	Backward Other(OB)	Forward Caste	Total
Hinduism	8.84	21.28	40.06	29.82	100
Islam	2.98	0.99	35.67	60.37	100
Christianity	66.24	3.69	11.60	18.47	100
Sikhism	0.56	31.56	19.90	47.98	100
Jainism	7.31	0.00	2.69	90.00	100
Buddhism	39.27	50.81	5.83	4.10	100
Others	85.36	1.30	11.68	1.67	100
Total	12.52	18.17	36.88	32.43	100

For explanation on SC, ST, OB see notes of Table 3.1.

Table 3.4: Caste System and Occupation (Descriptives)

Social Group	Self Employed	Salaried Employee	Casual Labor	Unemployed	Total
Backward Caste(SC)	46.91	18.69	29.77	4.62	100
Backward Tribe(ST)	28.32	18.72	47.39	5.57	100
Backward Other(OB)	45.75	21.59	27.50	5.17	100
Forward Caste	46.62	31.59	14.66	7.13	100
Total	43.01	23.95	27.23	5.81	100

For explanation on SC, ST, OB see notes of Table 3.1.

Table 3.5: Caste System and Occupation in Hinduism (Descriptives)

Social Group	Self Employed	Salaried Employee	Casual Labor	Unemployed	Total
Backward Caste(SC)	36.10	13.72	45.70	4.48	100
Backward Tribe(ST)	28.78	18.29	47.45	5.47	100
Backward Other(OB)	45.67	21.44	27.84	5.05	100
Forward Caste	45.90	34.23	12.43	7.44	100
Total	41.3	23.9	29	5.8	100

For explanation on SC, ST, OB see notes of Table 3.1.

Table 3.6: Hinduism and Entrepreneurship

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion:				
Hinduism	-0.0861***	0.0293***	0.0534***	0.00346***
	(0.0052)	(0.0044)	(0.0042)	(0.00088)
Personal Characteristics:				
Age	0.0123***	0.00758***	-0.0160***	-0.00397***
	(0.0011)	(0.0010)	(0.00093)	(0.00031)
Agesq/100	-0.00424***	-0.00834***	0.00939***	0.00318***
	(0.0013)	(0.0012)	(0.0011)	(0.00040)
Female	-0.133***	0.0630***	0.0425***	0.0272***
	(0.0055)	(0.0052)	(0.0048)	(0.0019)
Married	0.0883***	-0.0445***	0.000897	-0.0447***
	(0.0066)	(0.0058)	(0.0056)	(0.0026)
Divorced	0.106***	-0.0540***	-0.0375***	-0.0149***
	(0.012)	(0.0096)	(0.0089)	(0.0011)
General Education:				
Informal Education	0.0308***	0.0721***	-0.102***	-0.000700
	(0.0084)	(0.0087)	(0.0045)	(0.0026)
Primary School	0.0148**	0.170***	-0.202***	0.0171***
	(0.0060)	(0.0060)	(0.0035)	(0.0022)
High School	-0.0763***	0.312***	-0.286***	0.0499***
	(0.0065)	(0.0066)	(0.0029)	(0.0037)
University	-0.226***	0.426***	-0.297***	0.0958***
	(0.0070)	(0.0081)	(0.0022)	(0.0066)
Technical Education:				
Technical Degree	0.0139	0.0930***	-0.107***	0.000122
	(0.025)	(0.021)	(0.027)	(0.0033)
Technical Diploma	-0.00744	0.105***	-0.111***	0.0134***
	(0.010)	(0.0090)	(0.0084)	(0.0021)
Household Characteristics:				
Urban	0.0439***	0.171***	-0.218***	0.00384***
	(0.0047)	(0.0042)	(0.0033)	(0.00088)
0.2<Land<0.4 Hectares	0.0730***	-0.0762***	0.00339	-0.000272
	(0.0054)	(0.0044)	(0.0045)	(0.0010)
0.4< Land < 2 Hectares	0.325***	-0.146***	-0.176***	-0.00309***
	(0.0055)	(0.0045)	(0.0039)	(0.0011)
Land > 2 Hectares	0.397***	-0.154***	-0.237***	-0.00606***
	(0.0053)	(0.0047)	(0.0026)	(0.0012)
Observations	87181			

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base categories for marital status, general education, technical education, land dummies are unmarried, no general education, no technical education and less than 0.2 hectares of land respectively. Full set of state level regional dummies are also included in the regression.

Table 3.7: Hinduism, Backwardness and Entrepreneurship

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion and Class:				
Hinduism	-0.0669***	0.0323***	0.0309***	0.00373***
	(0.0053)	(0.0045)	(0.0044)	(0.00089)
Backward Class	-0.0817***	-0.0114***	0.0942***	-0.00106
	(0.0046)	(0.0040)	(0.0038)	(0.00086)
Controls:				
Personal Characteristics	YES			
General Education	YES			
Technical Education	YES			
Household Characteristics	YES			
Regional Dummies	YES			
Observations	87175			

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for religion is non-Hindu and for caste is non-backward class.

Table 3.8: Hinduism, Caste System and Entrepreneurship

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion and Class:				
Hindu SC	-0.141***	0.0332***	0.108***	0.000583
	(0.0090)	(0.0093)	(0.0088)	(0.0020)
Hindu ST	-0.191***	0.0219***	0.162***	0.00727***
	(0.0065)	(0.0063)	(0.0066)	(0.0016)
Hindu OB	-0.0571***	0.0203***	0.0356***	0.00122
	(0.0060)	(0.0054)	(0.0052)	(0.0012)
Hindu Forward	-0.0223***	0.0491***	-0.0326***	0.00574***
	(0.0063)	(0.0057)	(0.0055)	(0.0013)
Controls:				
Personal Characteristics	YES			
General Education	YES			
Technical Education	YES			
Household Characteristics	YES			
Regional Dummies	YES			
Observations	87181			

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for “Religion and Class” variables is Nonhindu. Individuals of backward classes belong to one of the three categories: Scheduled Castes(SC), Scheduled Tribes(ST) and Other Backward Classes(OB).

Table 3.9: Backward Classes and Entrepreneurship (Only Hindus)

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion and Class:				
Hindu SC	-0.146***	-0.0331***	0.183***	-0.00331**
	(0.0084)	(0.0078)	(0.0090)	(0.0016)
Hindu ST	-0.181***	-0.0415***	0.222***	0.000495
	(0.0063)	(0.0054)	(0.0067)	(0.0012)
Hindu OBC	-0.0446***	-0.0425***	0.0926***	-0.00547***
	(0.0057)	(0.0048)	(0.0055)	(0.0010)
Controls:				
Personal Characteristics	YES			
General Education	YES			
Technical Education	YES			
Household Characteristics	YES			
Regional Dummies	YES			
Observations	69705			

Notes: *Signifies $p < 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for the Hindu caste is Hindu Forward. Set of state level regional dummies that have nonzero observations in all the four categories are included in the regression.

Table 3.10: Religion and Entrepreneurship

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion and Class:				
Muslim	0.0792***	-0.0475***	-0.0271***	-0.00462***
	(0.0063)	(0.0052)	(0.0052)	(0.00098)
Christian	0.0290**	0.0200**	-0.0490***	-0.0000146
	(0.012)	(0.010)	(0.0090)	(0.0020)
Sikh	0.00315	-0.0224	0.0145	0.00476
	(0.021)	(0.016)	(0.020)	(0.0048)
Jain	0.271***	-0.132***	-0.124***	-0.0155***
	(0.029)	(0.018)	(0.027)	(0.00094)
Buddhist	-0.0194	0.0350*	-0.0111	-0.00444
	(0.021)	(0.018)	(0.016)	(0.0031)
Others	0.134***	-0.0493**	-0.0827***	-0.00196
	(0.022)	(0.019)	(0.017)	(0.0044)
Backward Class	-0.0778***	-0.0150***	0.0941***	-0.00126
	(0.0047)	(0.0041)	(0.0039)	(0.00087)
Controls:				
Personal Characteristics	YES			
General Education	YES			
Technical Education	YES			
Household Characteristics	YES			
Regional Variables	YES			
Observations	87175			

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for religion is Hindu.

Table 3.11: Self-employed and Employers

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Employer	Salaried Employee	Casual Labor	Unemployed
Religion and Class:					
Hinduism	-0.0720***	-0.00161**	0.0109***	0.0588***	0.00386***
	(0.0048)	(0.00072)	(0.0042)	(0.0039)	(0.00087)
Backward Class	-0.0727***	-0.00731***	-0.0182***	0.1000***	-0.00168*
	(0.0044)	(0.00083)	(0.0038)	(0.0037)	(0.00088)
Controls:					
Personal Characteristics	YES				
General Education	YES				
Technical Education	YES				
Household Characteristics	YES				
Regional Dummies	YES				
Observations	87175				

Notes: Employers are treated as a separate class here. *Signifies $p< 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for religion is non-Hindu and for caste is non-backward class. Set of state level regional dummies that have nonzero observations in all the five categories are included in the regression.

Table 3.12: Entrepreneurship in Nonagriculture

(Marginal Effects after Multinomial Probit Estimation)

Independent	Self Employed	Salaried Employee	Casual Labor	Unemployed
Religion and Class:				
Hinduism	-0.0721***	0.0548***	0.00949**	0.00776***
	(0.0061)	(0.0061)	(0.0040)	(0.0020)
Backward Class	-0.0552***	-0.00502	0.0581***	0.00207
	(0.0054)	(0.0055)	(0.0035)	(0.0019)
Controls:				
Personal Characteristics	YES			
General Education	YES			
Technical Education	YES			
Household Characteristics	YES			
Regional Dummies	YES			
Observations	52484			

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is primary occupation of the individual. Base category for religion is non-Hindu and for caste is non-backward class. Full set of state level regional dummies are also included in the regression.

Chapter 4

The Geography of Start-up Size

In this chapter, spatial location is analyzed as a micro-determinant of the start-up size of new firms using a dataset of 150 thousand start-ups in India. Geoadditive models are used to estimate the effect of ownership structure, knowledge endowments and spatial location on start-up size. The results suggest that firm size distribution exhibits distinct regional patterns. Ownership structure influences the initial size with proprietary owners starting small micro firms. The spatial patterns are found to be explainable, to some extent, by the economic and financial development of the regions.

4.1 Introduction

It is widely recognized that entry of new firms has profound influence on the economy. Entrants generate disproportionately higher number of jobs than incumbents and drive innovation and competition. Stochastic as well as deterministic theories predict who amongst them grow to survive and who decline or exit.[1] A number of such studies show that, for new entrants at least, the initial size influences growth and survival (see Geroski, 1995, for a survey). The determinants of the size of the firms at entry, however, remained under-researched and neglected in this discussion, for a long time.[2]

[1]The Stochastic approach suggests that the likelihood of survival is random across firms and entrepreneurs learn from their post entry experience (Jovanovic, 1982, 1994). The deterministic approach suggests that firm and industry specific characteristics determine the post-entry performance of the firms (Dixit, 1989; Audretsch, 1995).

[2]As Mata and Machado (1996, pp.1306) note, "in spite of the increased attention recently devoted to study of entry and to the birth of new firms, and of the unequivocal role that start-up size has also been found to play in the post entry performance of firms, the fact remains that the analysis of the choice of firms' start-up size has been relatively neglected. In a more recent study, Colombo et al. (2004, pp. 1184) write, "Unfortunately, the analysis of the determinants of the size of new firms has so - far remained rather undeveloped."

The few studies on start-up size show that the industry characteristics (Mata and Machado, 1996; Mata, 1996) and human capital of entrepreneurs (Astebro and Bernhardt, 2005; Colombo et al., 2004; Colombo and Grilli, 2005), determine the start-up size of new firms. However, the role of spatial location on the start-up size has never been studied although the economic geography literature emphasizes the geographic location as an important determinant in shaping economic activity (Krugman, 1991; Fujita and Krugman, 2003). This paper contributes to the growing literature on the start-up size by highlighting that the firm size distribution of start-ups (FSDS) is not independent of the spatial context. Using recent methodological advances in spatial econometrics and a dataset of 150,000 firms that registered as small firms in India from 1998-2000, we find that the FSDS is remarkably spatially skewed and displays distinct spatial patterns.

The paper consists of five sections. In the next section, we discuss the theoretical framework and present the hypotheses on the FSDS in an Indian context. In the third section, we present the geoadditive modeling techniques with Bayesian inference based on Monte Carlo Markov Chain(MCMC) methods. In the fourth section, we give the empirical results linking the region with the FSDS. In the final section, we provide the conclusions and summary and present possible avenues for future research.

4.2 The Start-Up Size

One of the stylized facts in the industrial dynamics literature is that the magnitude of firm entry, across industries, time periods, and regions is quite startling. Firm size distribution is skewed and the majority of entrants are small (Cable and Schwalbach, 1991). The likelihood of survival for new entrants is low and those that do survive grow at a higher rate than the incumbents. Firms that have a higher start-up size have a higher likelihood of survival (Dunne et al., 1989; Guimaraes et al., 1995).[3] Many empirical studies categorically reject the Gibrat's Law which, in essence, claims that the firm growth is independent of size. Three

[3]However, there are some exceptions. Agarwal and Audretsch (2001) show that the entry size is more important in the early stages of the industry life cycle but not in the mature stages. Audretsch et al. (1999), however, find that there is no relationship between start-up size and firm survival in a sample of Italian firms. They also find that growth rates are even neagtively correlated with initial size.

important surveys (Geroski, 1995; Sutton, 1997; Caves, 1998) summarize these and other major findings of the literature on entry, growth, survival and exit of firms.

While the effects of entry are extensively discussed, the determinants of the start-up size have received little attention. As Colombo et al. (2004, p. 1184) note, "if a larger start-up size positively affects the likelihood of survival of new firms and if surviving new firms that started operations at smaller scale struggle to catch up, the question arises why there are firms with small initial size." The few empirical studies on the determinants of the start-up size of firms include Mata and Machado (1996), Mata (1996), Görg et al. (2000), Görg and Strobl (2002), Astebro and Bernhardt (2005), Colombo et al. (2004), Colombo and Grilli (2005). These studies examine the role of industry characteristics such as the minimum efficient scale (MES) of the industry, industry growth, effects of operation at suboptimal scale (defined as the proportion of those employed in firms that are operating at sub-optimal scale), impact of market size, role of human capital characteristics of founders, such as previous work experience and education, and credit constraints, on the initial size of firms.

As Mata and Machado (1996, p. 1321)[4] note, "entry on a relatively large scale in each industry is much more sensitive to the minimum efficient scale and to the extent of firm turnover in the industry than entry in small scale. Put differently, it seems that small new firms appear everywhere, while relatively large ones only appear where economies of scale make it crucial, or where sunk costs are low, therefore leading to low losses in case of failure." A similar study on Irish firms shows comparable results, but finds a negative effect of industry size and positive effect of industry growth on start-up size (Görg et al., 2000). The start-up size increases with age and education of the founder, and is higher in industries with higher minimum efficient scale (MES), greater turbulence, and in industries where few suboptimal firms operate (Mata, 1996). Industry-specific professional knowledge and managerial and entrepreneurial experience have been found to have a greater positive impact than education and working experience on the start-up size (Colombo et al., 2004).[5]

[4] Mata and Machado (1996) analyze a sample of 1079 new firms from Portugal. In their sample, not more than 25% have greater than the average size of 17 employees, and 50% of the firms employ less than 10 people.

[5] Colombo et al. (2004) investigate start-up size of 391 technology based young Italian firms in both manufacturing and services.

Görg and Strobl (2002) find that the presence of multinationals negatively effects the size of domestic Irish entrants. Astebro and Bernhardt (2005) show that entrepreneurial human capital of founders co-determines their household wealth and the firms start-up capital. According to (Colombo and Grilli, 2005), firms receiving external private equity financing have greater start-up size. Advertising costs and R&D expenditures are important in determining the start-up size of large firms than small firms (Arauzo-Carod and Segarra-Blasco, 2005). Nurmi (2006) studies sectoral differences in start-up size in Finland and finds that results for manufacturing and service sectors are very similar. In addition, some studies show that start-up size is higher when entrepreneurs receive inheritances (Holtz-Eakin et al., 1994). Evans and Jovanovic (1989) discover the presence of binding liquidity constraints that limit start-up capital of entrepreneurs. They find that "entrepreneurs are limited to a capital stock that is no more than about one and one-half times of their wealth." Thus, almost all entrepreneurs in their sample "devote less capital to their business than they would like to." (p. 825)

As mentioned earlier, we hypothesize that start-up size is not independent of the geographic region. The growing literature of economic geography (Krugman, 1991; Fujita and Krugman, 2003) gives us compelling reasons to hypothesize that the spatial location should play an important role in determining the size of new start-ups. In particular, there are compelling reasons to posit that some regions give birth to firms with a greater start-up size while others lead to creation of very small firms. We also hypothesize that initial knowledge endowments of the firm and the ownership structure influence the start-up size. Entrepreneurs who possess technical knowhow are more likely to start with larger firms. Firms that have single proprietary ownership are more likely to be small compared to those that have partnership or co-operative ownership structures.

4.3 Geoadditive Models

We use semiparametric regression techniques based on Bayesian P-Splines and geoadditive models for the empirical analysis. The method allows estimating the non-linearities of continuous variables and the neighborhood effects on the start-up size of new firms.[6] A brief outline of the methodology is presented here.

[6]This section draws from Lang and Brezger (2004); Brezger and Lang (2005).

Let (y_i, x_i, v_i) for i in {1,2,...N} describe a dataset of N observations. Let y_i be the response variable and x_i be a m-dimensional vector of continuous covariates and v_i be a vector of categorical variables. If y_i are independent and Gaussian with mean, $\eta_i = f_1(x_{i1}) + ... + f_p(x_{ip}) + v_i\gamma$ and a common variance σ^2, f_i are unknown smooth functions of the continuous variables and $v_i\gamma$ corresponds to the parametric part of the regression then the regression model is called the Additive Model or a Semiparametric regressor.

Eilers and Marx (1996) assumed that the unknown functions f_j can be approximated by a l degree spline with equally positioned knots in the domain of x_j. By writing such a spline in the form of a linear combination of k B-Spline basis functions, B_{jk}, where k is equal to the number of knots plus the degree of the spline, we have, $f_j(x_j) = \Sigma\beta_{jk}B_{jk}$ and, in matrix notation, $\eta = \Sigma X_j\beta_j + V\gamma$. By defining a roughness penalty based on the differences of adjacent B-Spline coefficients, in order to ensure smoothness of the estimated functions, the penalized likelihood assumes the form:

$$L = l(y, \beta_1, ..., \beta_p, \gamma) - \lambda_1\Sigma(\triangle^k\beta_1)^2 - ....\lambda_p\Sigma(\triangle^k\beta_p)^2 \tag{4.1}$$

In the Bayesian set-up, β_j for $j = 1....p$ and γ are considered as random variables and assigned prior distributions. We assume independent diffuse priors for the fixed effects parameters, γ_j for $j = 1....q$. The priors for the coefficients of the non-linear functions β_j are obtained by substituting the stochastic analogues of the difference penalties. In case of first differences, we consider first order random walk and for second differences, a second order random walk. Hence, we have, $\beta_{jk} = \beta_{j,k-1} + u_{jk}$ or $\beta_{jk} = 2\beta_{j,k-1} - \beta j, k-2 + u_{jk}$ with Gaussian errors $u_{jk} \sim N(0, \tau_j^2)$ and constant diffuse priors for the initial values of β_{j1} and β_{j2}. τ_j^2 controls the smoothness of the fitted function. For Bayesian inference, τ_j^2 are also treated as random variables and simultaneously estimated with the β_j. Highly dispersed inverse gamma priors $IG(a_j, b_j)$ are assigned to the variances τ_j^2.

To the above predictor, if we add a spatial effect $f_{spatial}$, then we obtain a geoadditive model. The spatial effect may be split into a spatially correlated and uncorrelated effect, $f_{spatial} = f_{str} + f_{unstr} = X_{str}\beta_{str} + X_{unstr}\beta_{untr}$, as the spatial effect may comprise of a component that has strong spatial structure and a component that is only locally present. Following Besag et al. (1991) we assume Markov Random Field (MRF) priors for the regression coefficients β_{str}.

If $s \in 1, ...S$ are pixels of a lattice or regions of a geographical map, then the MRF prior is given as,

$$\beta_{str,s} \backslash \beta_{str,u} \sim N(\sum_{u \in \partial_s} \frac{1}{N_s} \beta_{str,u}, \frac{\tau^2_{str}}{N_s}) \tag{4.2}$$

for, $u \neq s$, where, N_s is the number of adjacent regions (pixels) and ∂_s is the neighborhood of s. This prior may be seen as an extension of a first order random walk into a two dimensional space. For the second component, β_{unstr}, we assume independent and identically distributed (i.i.d.) Gaussian random priors, $\beta_{unstr}(s) \sim N(0, \tau^2_{unstr})$, for s=1.....,S. For τ^2_{str} and τ^2_{unstr} we assume inverse gamma priors, $IG(a_{str}, b_{str})$ and $IG(a_{unstr}, b_{unstr})$.

Inference is based on the posterior and uses recent MCMC techniques. If α is a vector of the unknown parameters, assuming conditional independence of the parameters, the posterior is given by:

$$\begin{aligned} p(\alpha \backslash y) \propto L(y, \beta_1, ..., \beta_p, \beta_{str}, \beta_{unstr}, \gamma, \sigma^2) \times \prod_{j=1}^{p} (p(\beta_j \backslash \tau^2_j) p(\tau^2_j)) \\ \times p(\beta_{str} \backslash \tau^2_{str}) p(\tau^2_{str}) p(\beta_{unstr} \backslash \tau^2_{unstr}) p(\tau^2_{unstr}) p(\gamma) p(\sigma^2) \end{aligned} \tag{4.3}$$

Following Spiegelhalter et al. (2002), the Deviance Information Criteria (DIC) is used as a measure of complexity and fit for model selection. The DIC is defined as the (p. 603) "classical estimate of fit, plus twice the effective number of parameters." The unstandardized deviance is given by $-2log\{p(y \backslash \mu)\}$. Assuming that $f(y)$ as a standardizing term that is a function of the data alone, the classical estimate of fit, $D(\bar{\theta})$ is obtained from $D(\theta) = -2log\{p(y \backslash \theta)\} + 2logf(y)$, by evaluating $D(\theta)$ at the mean of the parameters $\bar{\theta}$. $D(\theta)$ is also referred to as the Bayesian deviance or the saturated deviance. For members of the exponential family with $E(Y) = \mu(\theta)$, $D(\theta)$ is obtained by setting $f(y) = p\{y \backslash \mu(\theta) = y\}$. That is, $D(\theta) = -2log\{p(y \backslash \theta)\} + 2log\{p(y \backslash \mu(\theta) = y)\}$. The measure of the effective number of parameters, p_D, is the difference between the posterior mean of the deviance $\overline{D(\theta)}$ and the deviance at the posterior means of the parameters $D(\bar{\theta})$. That is, $p_D = \overline{D(\theta)} - D(\bar{\theta})$. Then, $DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D$. Of the competing models, the specification with the least DIC is selected and reported.

Consider estimating the geoadditive model with only the spatial component.

This would show the regional patterns of start-up size without controlling for firm characteristics. However, when firm characteristics (also called fixed effects) are also introduced into the geoadditive model, the resulting spatial pattern shows the residual spatial pattern after these characteristics are controlled for. Thus, the spatial patterns estimated in this paper are the residual spatial patterns, as we simultaneously introduce firm characteristics and the spatial components in the geoadditive framework. These estimated residual spatial patterns can be explained using one of the following econometric approaches. A simple strategy is to regress the mean residual spatial effects on the regional variables. Thus, after estimating the geoadditive model, the total spatial effect of each region is explained by regressing the posterior mean of the estimated spatial residual effect on the regional variables. However, this empirical strategy does not consider the estimated posterior variance of spatial effects. In order to overcome this problem, a discrete choice model of the 95% or 80% spatial effects can be estimated. In this case, a variable is constructed that takes a value of (-1) when the region has a significant negative effect, takes a value of (0) if the effect is insignificant and takes a value of (1) if the effect is significant and positive. This leads to a straightforward multinomial specification. This variable is then regressed on the regional variables. We employ both strategies to examine the determinants of the residual spatial patterns.

4.4 Data

The main source of data for linking the geographic location of the firm with the start-up size is the Ministry of Small Scale Industries in India. We use firm level data from the third census of registered small scale firms. This census was conducted in 2001. We consider manufacturing firms that have started producing in 1998, 1999 or 2000 as new start-ups for the analysis following Audretsch and Keilbach (2004), who also consider the three year period, as new start-ups are subject to a very high degree of stochastic disturbance if only a very short period is considered. This rich dataset of entrants consists of 149,708 firms. Each such start-up was asked the set of initial conditions under which it was founded (like the original value of its plant and machinery, its year of initial production, the sector, the source of its technical knowledge, its spatial location). We use this data to test the hypothesis that directly follows from our theoretical analysis.

As the dataset is of small firms, we do not have information of large entrants. This limitation of the dataset, however, does not pose serious problems for testing our hypothesis as the theory of firm size distribution suggests that majority of entrants are small and numerous. Furthermore, if few large entrants are also present in the dataset, they would at best be outliers and introduce heterogeneity.[7]

As the descriptive tables in section 4.6 suggest 88.5% of the sample consists of firms that are started by proprietors. 6.8% of the firms are owned by two or more partners and are referred to as partnerships. Firms having other ownership structures such as co-operatives are 4.7% of the sample. 15.8% of the firms are managed by women. 73.8% are small scale industrial units. Thus, 26.2% of the firms are small scale business enterprises, primarily consisting of repairing, servicing and maintenance units. More than 14% of the firms have reported that they have technical knowledge. While only 0.94% of the firms in the sample have reported to have obtained knowledge from sources outside India, as many as 6.6% have their technical knowledge from other firms and 6.67% from universities. 20.9% of the firms are in the industrial sub-sector of apparels manufacture and 19.2% are firms dealing with food products. With 11.67% of all the firms, the next largest group comprises of firms in the industrial sub-sector of fabricated metals.

4.5 Empirical Analysis

Geoadditive models are estimated to examine the effect of the geographic location on the start-up size. Two measures of start-up size are used. In the first model, the dependent variable, start-up size, is measured using initial employment of a firm. In the second model, initial value of fixed assets is used as a measure of start-up size. The following geoadditive models are estimated:

$$\eta = \gamma_{const} + \gamma_{ProprietaryOwnership} + \gamma_{Woman} + \gamma_{TechnicalKnowledge} + \gamma_{IndustrialSector} + f_{spatial}(district) + f_{random}(district)$$

The structured spatial effects are estimated based on Markov random field

[7] There is compelling evidence that entry takes place in the form of new small firms (Audretsch, 1995; Dunne et al., 1989). This is one of the main reasons for the studies on the start-up size to use quantile regressions (Mata and Machado, 1996; Görg et al., 2000; Colombo et al., 2004).

priors and random spatial effects are estimated with gaussian priors.[8]

Table 4.4 suggests that the ownership type has an influence on the start-up employment. Firms that are started by single proprietors and women have a smaller start-up size. The estimation results suggest that start-ups by proprietary owners have a start-up size that is that is 66% smaller than the average size, and start-ups by women have a size that is 18% smaller than the average size, ceteris paribus. Firms that have a different ownership structures, such as partnerships, and firms that have technical know-how are more likely to have a higher start-up size. In particular, technical knowledge from abroad increases the start-up size by 15%, technical knowledge from other firms increases the start-up size by 7.8% and from universities by 8.3%. Thus firms that have technical knowledge at the start-up phase tend to have higher start-up size than firms that do not have any technical knowledge. Furthermore, firms that are located in urban regions are more likely to have a larger start-up size.

Table 4.5 shows that these findings are robust to an alternate specification, with initial size measured by the initial value of fixed assets. Proprietary owners are found to start with lower levels of initial assets and so are women entrepreneurs. It is also seen that technical knowledge and urban location also positively effect the start-up size.

Figure 4.1 shows a clear presence of neighborhood effects on the start-up size, measured by initial employment. The structured spatial effects plotted in Figure 4.1(a) show that start-ups in northern regions of Uttaranchal, Uttar Pradesh, Bihar, Madhya Pradesh, western regions of Gujrat and Rajasthan and southern regions of Kerala, Karnataka, and Tamil Nadu are likely to be smaller. While Uttaranchal, Uttar Pradesh, Bihar, and Madhya Pradesh are poorer states, Gujrat, Kerala, Karnataka, and Tamil Nadu are richer regions. However, as the 95% confidence map in Figure 4.1(c) suggests, the negative effect of size that is seen in the richer southern regions is insignificant. Many districts of Uttar Pradesh and Rajasthan become insignificant in the 95% confidence map, as seen in Figure 4.1(c). Figure 4.1(a) suggests that the start-ups in Maharastra, Andhra Pradesh in the south, West Bengal in the east, the northeastern states, and Punjab in the

[8]The variance components in all the cases are estimated based on inverse gamma priors with hyperparameters a=0.001 and b=0.001. The number of iterations is set to 120000 with burnin parameter set to 20000 and the thinning parameter set to 100. The autocorrelation files and the sampling paths show that the MCMC algorithm has converged. These plots are available from the author.

north have a higher start-up size. These results should be interpreted with some care, as start-up size measured by initial employment may capture some effect of labor-intensive industries. However, the problem is mitigated to a great extent, as the industry effects are controlled in the regression. The random spatial effects in Figure 4.1(b) show that local spatial effects are much less compared to the structured spatial effects.

The maps of spatial effects of the second specification, with initial value of fixed assets as the dependent variable, in Figure 4.2, demonstrate a similar spatial pattern as in the first model. While Maharastra and Andhra Pradesh form a belt of start-ups with a higher size, Uttar Pradesh, Madhya Pradesh, and Bihar, three of the poorest Indian states, have a significantly smaller start-up size (Figure 4.2(c)). The random local spatial effects in Figure 4.2(b) are relatively small when compared to the structured spatial effects. It is also seen that structured spatial effect in the state of Gujarat is insignificant.

In the next step, we examine the determinants of these spatial patterns. The empirical method is described in the last subsection of section 4.3. Fazzari et al. (1988) show that financial constraints determine firm's investment decisions. Cabral and Mata (2003) find that the firm size distribution is right skewed for start-ups but evolves over time to a more symmetric distribution. This is explainable by the presence of the financial constraints that ease out with time. They find that firms owned by young entrepreneurs are 30 percent lower in size compared to firms started by old entrepreneurs, at the start-up stage. As Cabral and Mata (2003, pp.1079-1080) note, "suppose that financing constraints are especially relevant for young firms. Then, even if the long-run size distribution for a given cohort is close to symmetric, we should observe a significant skew to the right during the first periods, that is a large mass of small firms. Among this mass of small firms, some are small because they want to be small on efficiency grounds, whereas others are small because they are financially constrained. In future periods, when financing constraints cease to be binding, the latter will grow to their optimal size, thus giving rise to a more symmetric distribution of firm size."

Holtz-Eakin et al. (1994) test the role of liquidity constraints in the formation of new enterprises. Their analysis suggests that the size of inheritance has an effect on entrepreneurial choice and also on the investment in the capital of a new enterprise. Aghion et al. (2005) predict that a country with more than a critical level of financial development converges to the growth rate of world's technology frontier. There is overwhelming evidence that financial development is an important determinant of short run growth rates (see Levine, 1997, for a survey).[9] Evans and Jovanovic (1989) and Holtz-Eakin et al. (1994) show that the probability of survival depends on assets. Under an assumption that banks lend under the security of collateral, this suggests presence of credit-rationing. Hurst and Lusardi (2004) argue that households with higher levels of wealth have a higher tolerance for risk and are most likely to be business owners.

For these reasons, we hypothesize that the level of financial development should be able to explain the spatial patterns in firm start-up size. Lucas (1978) argues that the average firm size increases with economic development. We thus introduce the net per-capita state domestic product (NSDP) as an explanatory variable. We hypothesize that a higher net per-capita state domestic product increases the start-up size in the region. We control for unemployment and literacy rates, district size, population density and agglomeration effects. The reasons for introducing these variables are as follows. High unemployment may compel people into self-employment thus making them start small micro-enterprises, as an alternative to staying unemployed.[10] Literacy rate in the region captures the ability of people to perceive opportunities and thus it is possible that it has an influence on the start-up size. As high population density is positively related to the self-employment in chapter 2, it is most likely to have an effect on the firm start-up size if the factors of production in a region are distributed to a higher number of entrepreneurs. Furthermore, high agglomeration suggests the presence of spatial spillovers that may have a positive effects on the start-up size of new firms in the region. However, it also suggests that a number of ancillary firms are present in the region that provide vertical integrated services to the few large firms and thus may have a negative effect on start-up size.

In Table 4.6, we estimate the determinants of the mean spatial effects of the first model shown in Figure 4.1(a). The results suggest that economic and finan-

[9]However, Cressy (1996) argues that it is not financial constraints that affect the survival of new firms but the human capital of the entrepreneurs. Cressy (1996) claims that the relationship between liquidity constraints and survival is spurious and vanishes when a rich vector of entrepreneurs human capital variables is introduced into the estimations.

[10]The results of chapter 2 also suggest that unemployment has a positive effect on the probability to be self-employed in India.

cial development play an important role in determining the start-up size of new firms. The financial development is measured by per-capita credit flows in the region, credit-deposit ratio and density of banking facilities. The estimates suggest that the financial-development has a positive impact on the start-up size. The per-capita net state domestic product and the literacy rate in the region have significant positive effects while unemployment has a negative effect on the start-up size. Agglomeration index is significantly negative throughout. The demographic variables are mostly insignificant. Thus, the estimation results suggest that economic and financial development are more important determinants of start-up size. Table 4.8 explains the spatial effects of the second model, shown in Figure 4.2(a). The results confirm the effects of financial and economic development on start-up size. In the estimations in Table 4.7 and Table 4.9, the 95% significant spatial effects in Figure. 4.1(c) and Figure. 4.2(c) are explained using multinomial logit models. The inferences from these tables are consistent with the estimations having the mean spatial effects as the dependent variable, though there are some deviations. For instance, start-ups in large districts have a higher initial employment in Table 4.7 and start-ups in mid-sized districts have higher initial fixed effects.

4.6 Conclusion

A growing body of literature examines the determinants of the start-up size of firms. These few studies mainly focus on the industry characteristics and personality traits of the entrepreneurs. Using a new database of entrants in India, this paper examines geography and location as determinants of start-up size.

Our contribution is threefold: First, we show that the spatial location is a micro-determinant of start-up size of entrants. In particular, spatial neighborhood effects exert strong influence on firm size at entry. Second, we show that the ownership structure and initial knowledge endowments determine the firm size distribution of new start-ups. Third, we provide first insights into the determinants of the start-up size in a developing economy. The results also suggest that financial and economic development of a region can explain, to some degree, the spatial patterns that remain after controlling for the firm level effects.

Table 4.1: Characteristics of Start-ups (Descriptives)

Log(Employment)	1.0768
Std. Dev.	(*0.8382*)
Log(Value of Plant and Machinery)	10.5329
Std. Dev.	(*1.9202*)
Proprietary	0.8850
Partnership	0.0678
Other Ownership	0.0472
Managed by Woman	0.1581
Small Scale Industry (SSI)	0.7382
Small Scale Business Enterprise (SSBE)	0.2618
Tech Knowledge (Foreign)	0.0097
Tech Knowledge (Firm)	0.0659
Tech Knowledge (University)	0.0667
Food Products	0.1922
Tobacco	0.0013
Textiles	0.0490
Apparels	0.2090
Leather	0.0226
Wood	0.0391
Paper	0.0112
Printing	0.0382
Coke	0.0046
Chemicals	0.0368
Rubber	0.0432
Minerals	0.0651
Basic Metals	0.0168
Fabricated Metals	0.1167
Machinery	0.0290
Computing Machinery	0.0021
Electric Machinery	0.0233
Communication Equipment	0.0051
Precision Instruments	0.0036
Motor Vehicles	0.0062
Transport Equipment	0.0028
Furniture	0.0814
Recycling	0.0004

Table 4.2: Model I Diagnostics

Scale Parameter(Mean)	0.398845	
Std. dev.	0.00145918	
	Unstandardized	Saturated
Deviance (Mean)	287268.13	149733.13
Std. dev	35.141	547.965
deviance($\bar{\mu}$)	286731	149195
pD	537.121	535.54219
DIC	287805.25	150271.25

Table 4.3: Model II Diagnostics

Scale Parameter (Mean)	1.83737	
Std. dev.	0.00694745	
	Unstandardized	Saturated
Deviance (Mean)	504292.27	146514.26
Std. dev.	34.018535	519.3417
deviance($\bar{\mu}$)	503744.24	145965.31
pD	548.03044	548.94972
DIC	504840.3	147063.21

Table 4.4: Determinants of Start-up Size (Model I)

(Determinants of Initial Employment)

Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Ownership Structure				
Proprietary	-0.666	0.006	-0.677	-0.655
Woman	-0.180	0.005	-0.190	-0.171
Technical Knowledge				
Tech Knowledge (Foreign)	0.152	0.017	0.119	0.185
Tech Knowledge (Firm)	0.078	0.007	0.063	0.091
Tech Knowledge (University)	0.083	0.007	0.069	0.098
Firm Type				
SSI	0.408	0.005	0.398	0.418
Urban	0.075	0.004	0.068	0.082
Industries				
Tobacco	0.309	0.047	0.222	0.401
Textiles	0.329	0.009	0.312	0.347
Apparels	0.014	0.006	0.001	0.026
Leather	-0.122	0.012	-0.145	-0.098
Wood	0.012	0.009	-0.007	0.029
Paper	0.273	0.016	0.239	0.307
Printing	-0.013	0.010	-0.031	0.006
Coke	0.498	0.025	0.449	0.547
Chemicals	0.264	0.010	0.244	0.282
Rubber	0.168	0.009	0.150	0.185
Minerals	0.530	0.008	0.515	0.545
Basic Metals	0.326	0.014	0.298	0.355
Fabricated Metals	0.038	0.007	0.026	0.051
Machinery	0.099	0.011	0.078	0.119
Computing Machinery	-0.152	0.036	-0.220	-0.084
Electric Machinery	-0.046	0.012	-0.069	-0.023
Communication Equipment	-0.010	0.023	-0.052	0.036
Precision Instruments	0.113	0.029	0.057	0.174
Motor Vehicles	0.377	0.021	0.335	0.419
Transport Equipment	0.169	0.032	0.106	0.231
Furniture	-0.041	0.007	-0.056	-0.028
Recycling	0.344	0.082	0.183	0.498
Year Controls				
Year 1999	0.016	0.004	0.009	0.024
Year 2000	0.025	0.004	0.017	0.034
Constant	1.323	0.011	1.302	1.343
Observations	149709			

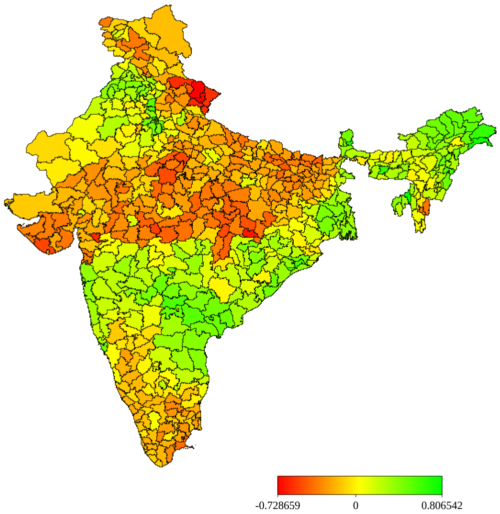
Notes: Dependent variable is log of initial employment of the firm.

Table 4.5: Determinants of Start-up Size (Model II)

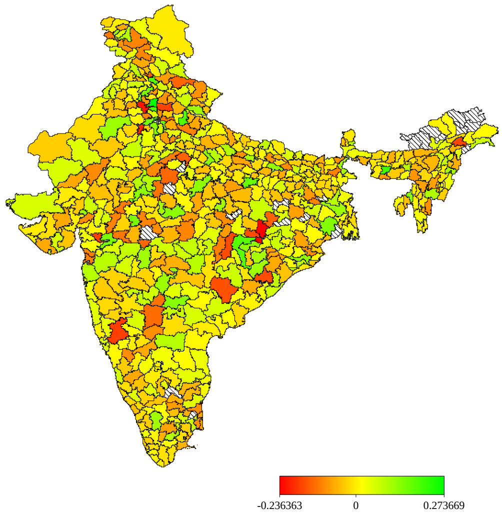
(Initial Value of Fixed Assets)

Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Ownership Structure				
Proprietary	-1.408	0.012	-1.433	-1.384
Woman	-0.547	0.011	-0.567	-0.526
Technical Knowledge				
Tech Knowledge (Foreign)	0.444	0.038	0.373	0.516
Tech Knowledge (Firm)	0.213	0.016	0.183	0.244
Tech Knowledge (University)	0.289	0.016	0.258	0.319
Firm Type				
SSI	0.508	0.011	0.486	0.529
Urban	0.194	0.008	0.177	0.211
Industries				
Tobacco	-0.986	0.102	-1.180	-0.786
Textiles	-0.465	0.021	-0.506	-0.424
Apparels	-1.067	0.014	-1.094	-1.040
Leather	-1.439	0.026	-1.492	-1.388
Wood	-1.041	0.019	-1.079	-1.001
Paper	0.048	0.037	-0.023	0.121
Printing	0.008	0.021	-0.034	0.048
Coke	0.587	0.057	0.478	0.699
Chemicals	-0.349	0.021	-0.389	-0.306
Rubber	0.133	0.020	0.095	0.171
Minerals	-0.114	0.017	-0.147	-0.082
Basic Metals	0.078	0.029	0.019	0.131
Fabricated Metals	-0.378	0.014	-0.404	-0.350
Machinery	-0.264	0.023	-0.308	-0.220
Computing Machinery	-0.609	0.076	-0.758	-0.467
Electric Machinery	-0.724	0.025	-0.772	-0.671
Communication Equipment	-0.722	0.053	-0.820	-0.617
Precision Instruments	-0.282	0.060	-0.403	-0.161
Motor Vehicles	0.243	0.047	0.153	0.337
Transport Equipment	-0.439	0.068	-0.572	-0.304
Furniture	-1.040	0.015	-1.069	-1.011
Recycling	0.069	0.177	-0.276	0.438
Year Controls				
Year 1999	0.064	0.008	0.048	0.080
Year 2000	0.086	0.009	0.068	0.104
Constant	11.867	0.027	11.813	11.921
Observations	146519			

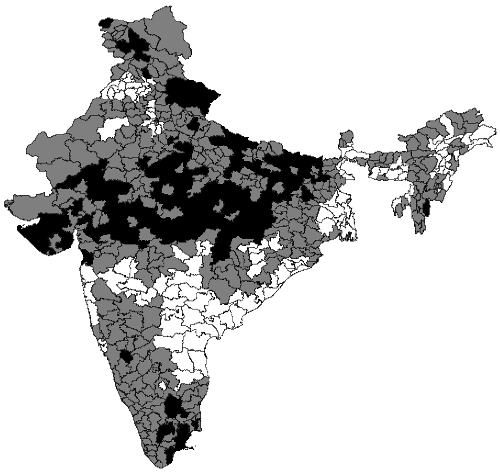
Notes: Dependent variable is log of initial value of fixed assets of the firm.



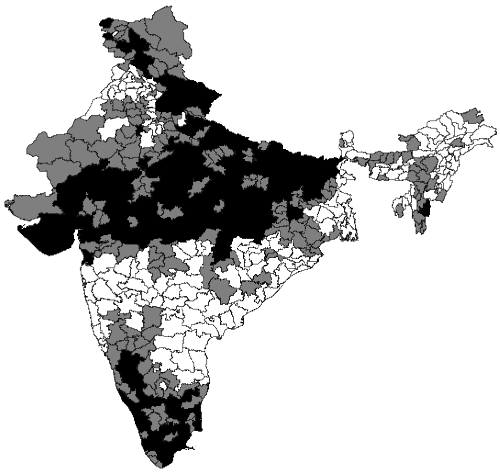
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(b) Unstructured Random Effect of 'District'. Shown are the posterior means.

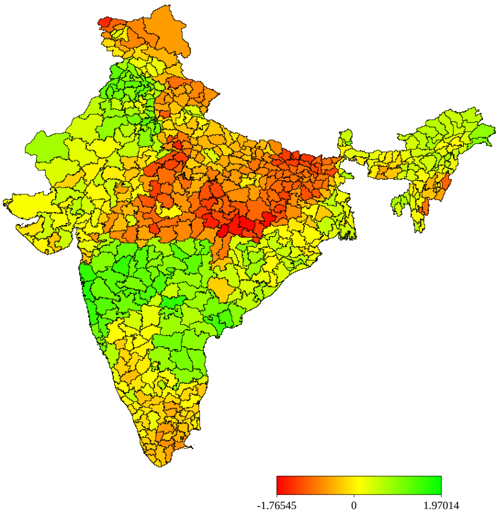


(c) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

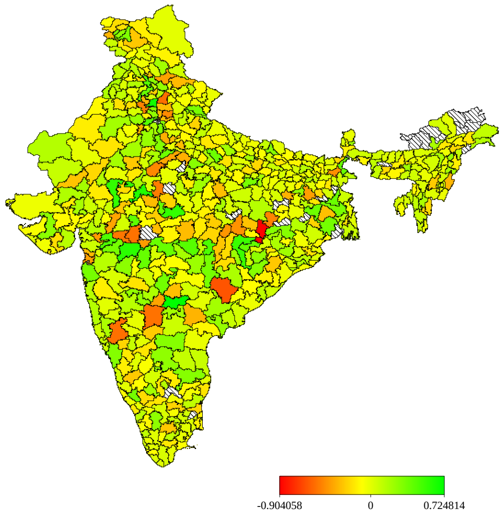


(d) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

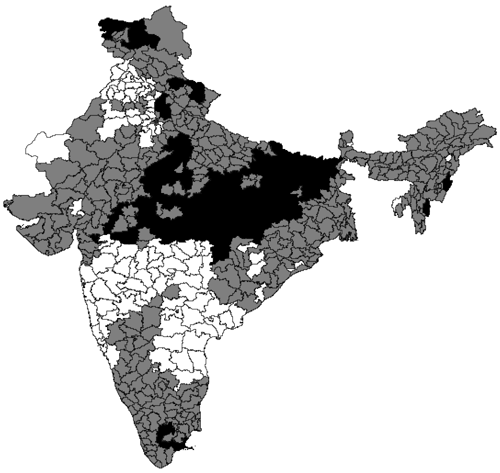
Figure 4.1: Spatial Effects in Model I



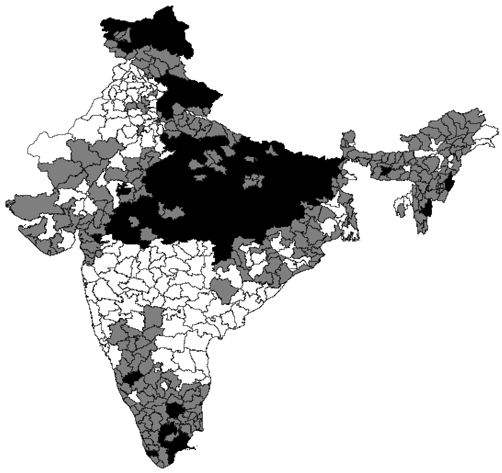
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(b) Unstructured Random Effect of 'District'. Shown are the posterior means.



(c) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(d) Non–linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

Figure 4.2: Spatial Effects in Model II

Table 4.6: Determinants of the Mean Spatial Effects in Figure 4.1

(Start-up Size given by initial employment)

Independent	Model I	Model II	Model III
Financial Development			
Per-capita Credit	0.0567***		
	(0.014)		
Credit-Deposit Ratio		0.112***	
		(0.019)	
Per-Capita Bank Offices			0.0385
			(0.027)
Economic Development			
Per-Capita NSDP	0.243***	0.231***	0.285***
	(0.035)	(0.034)	(0.033)
Unemployment	-0.0517***	-0.0726***	-0.0564***
	(0.013)	(0.013)	(0.013)
Literacy Rate	0.00331***	0.00530***	0.00404***
	(0.0011)	(0.0011)	(0.0012)
Demographics			
Mid Size District	0.0396	0.0319	0.0470*
	(0.025)	(0.024)	(0.025)
Large District	0.0740	0.0704	0.0830
	(0.075)	(0.074)	(0.076)
Population Density	-0.00749	0.00320	0.00582
	(0.013)	(0.012)	(0.012)
Agglomeration Index			
Firm Density	-0.179***	-0.173***	-0.171***
	(0.011)	(0.010)	(0.011)
Constant	-4.634***	-3.870***	-4.037***
	(0.32)	(0.33)	(0.45)
Observations	534	534	534
R^2	0.44	0.46	0.43
F	52.32	56.09	49.36
R^2 Adjusted	0.435	0.453	0.421

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is the mean spatial effect per district after estimation of the geoadditive models.

Table 4.7: Determinants of the 95% Spatial Effects in Figure 4.1

(Marginal Effects after Multinomial Logit Estimation)

	Model I			Model II			Model III		
	-ve black	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*
Financial Development									
Per-Capita Credit	-0.127	-0.00662	0.134						
	(21499)	(3108)	(18391)						
Credit-Deposit Ratio				-0.222	0.0327	0.190			
				(3515)	(2028)	(5544)			
Per-Capita Bank Offices							-0.109	0.0237	0.0855
							(0.083)	(0.038)	(0.072)
Economic Development									
Per-Capita NSDP	-0.514	0.0212	0.493	-0.499	-0.00528	0.504	-0.614***	0.0136	0.600***
	(69484)	(9944)	(79429)	(11742)	(327)	(11415)	(0.100)	(0.047)	(0.092)
Unemployment	0.0951	0.0168	-0.112	0.138	0.0148	-0.153	0.108***	0.0187	-0.126***
	(20343)	(7907)	(12436)	(3908)	(917)	(2991)	(0.039)	(0.018)	(0.035)
Literacy Rate	-0.0121	0.00206	0.0100	-0.0164	0.00255	0.0138	-0.0138***	0.00133	0.0124***
	(1071)	(968)	(2039)	(252)	(158)	(410)	(0.0035)	(0.0016)	(0.0032)
Demographics									
Mid Size District	-0.0515	-0.0808	0.132	-0.0261	-0.0908	0.117	-0.0834	-0.0730**	0.156**
	(35713)	(39303)	(3589)	(4931)	(5861)	(930)	(0.071)	(0.032)	(0.067)
Large District	-0.154	-0.348***	0.502**	-0.129	-0.353***	0.481**	-0.238	-0.344***	0.581***
	(0.22)	(0.028)	(0.22)	(0.24)	(0.029)	(0.24)	(0.16)	(0.028)	(0.16)
Population Density	0.0788	-0.0629	-0.0159	0.0513	-0.0693	0.0179	0.0427	-0.0518***	0.00916
	(10873)	(29538)	(18665)	(2209)	(4296)	(2087)	(0.038)	(0.018)	(0.035)
Agglomeration Index									
Firm Density	0.500	-0.107	-0.393	0.490	-0.121	-0.369	0.488***	-0.0950***	-0.393***
	(36533)	(50167)	(86699)	(5353)	(7514)	(12867)	(0.043)	(0.019)	(0.037)
Observations	534	534	534	534	534	534	534	534	534
Log Likelihood	-406.0			-404.0			-411.1		
$\chi^2(16)$	338.5			342.4			328.2		
Pseudo R^2	0.294			0.298			0.285		

Notes: Marginal effects after multinomial logit estimation. *Signifies $p < 0.05$; ** Signifies $p < 0.01$; *** Signifies $p < 0.001$. Standard errors are reported in parentheses. Dependent variable is the 95% spatial effect of the district, after estimation of the geoadditive models.

Table 4.8: Determinants of the Mean Spatial Effects in Figure 4.2

(Start-up Size given by initial fixed assets)

Independent	Model I	Model II	Model III
Financial Development			
Per-capita Credit	0.155***		
	(0.032)		
Credit-Deposit Ratio		0.216***	
		(0.044)	
Per-Capita Bank Offices			0.173***
			(0.060)
Economic Development			
Per-Capita NSDP	1.259***	1.274***	1.362***
	(0.079)	(0.077)	(0.075)
Unemployment	-0.0909***	-0.136***	-0.102***
	(0.029)	(0.029)	(0.029)
Literacy Rate	0.00499*	0.00986***	0.00601**
	(0.0026)	(0.0025)	(0.0026)
Demographics			
Mid Size District	0.173***	0.160***	0.201***
	(0.055)	(0.055)	(0.056)
Large District	-0.0531	-0.0608	-0.0118
	(0.17)	(0.17)	(0.17)
Population Density	-0.00409	0.0270	0.0328
	(0.028)	(0.027)	(0.028)
Agglomeration Index			
Firm Density	-0.412***	-0.393***	-0.398***
	(0.024)	(0.024)	(0.025)
Constant	-17.36***	-15.76***	-14.96***
	(0.73)	(0.76)	(1.00)
Observations	532	532	532
R^2	0.60	0.60	0.59
F	99.15	99.33	94.70
R^2 Adjusted	0.597	0.597	0.585

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is the mean spatial effect per district after estimation of the geoadditive models.

Table 4.9: Determinants of the 95% Spatial Effects in Figure 4.2

(Marginal Effects after Multinomial Logit Estimation)

	Model I			Model II			Model III		
	-ve black	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*	*-ve black*	*insig. grey*	*+ve white*
Financial Development									
Per-Capita Credit	-0.153***	0.0122	0.140***						
	(0.048)	(0.023)	(0.043)						
Credit-Deposit Ratio				-0.228	0.0377	0.191			
				(5822)	(3473)	(9295)			
Per-Capita Bank Offices							-0.168*	-0.00911	0.177**
							(0.10)	(0.044)	(0.090)
Economic Development									
Per-Capita NSDP	-1.375***	0.0813	1.294***	-1.309	0.0830	1.226	-1.466***	0.0918	1.374***
	(0.14)	(0.064)	(0.13)	(41929)	(7636)	(49564)	(0.14)	(0.061)	(0.13)
Unemployment	0.128***	0.0163	-0.145***	0.164	0.0117	-0.176	0.129***	0.0196	-0.149***
	(0.040)	(0.019)	(0.037)	(6675)	(1080)	(5593)	(0.040)	(0.019)	(0.037)
Literacy Rate	-0.00760**	0.000902	0.00670**	-0.0125	0.00182	0.0106	-0.00829**	0.00128	0.00701**
	(0.0035)	(0.0017)	(0.0034)	(333)	(168)	(501)	(0.0036)	(0.0018)	(0.0035)
Demographics									
Mid Size District	-0.0357	-0.119***	0.154**	-0.0194	-0.140	0.159	-0.0669	-0.120***	0.187**
	(0.078)	(0.034)	(0.074)	(9433)	(13558)	(4125)	(0.076)	(0.034)	(0.073)
Large District	0.235	-0.375***	0.139	0.217	-0.376***	0.159	0.0994	-0.371***	0.272
	(0.30)	(0.032)	(0.29)	(0.30)	(0.032)	(0.30)	(0.30)	(0.032)	(0.29)
Population Density	0.0227	0.00919	-0.0318	0.00663	0.00842	-0.0150	-0.0182	0.0123	0.00592
	(0.041)	(0.019)	(0.038)	(781)	(775)	(6.47)	(0.040)	(0.019)	(0.038)
Agglomeration Index									
Firm Density	0.521***	-0.128***	-0.393***	0.501	-0.157	-0.344	0.499***	-0.121***	-0.377***
	(0.046)	(0.023)	(0.039)	(8006)	(14443)	(22448)	(0.045)	(0.022)	(0.039)
Observations	532			532			532		
Log Likelihood	-363.0			-360.2			-366.5		
$\chi^2(16)$	420.4			425.9			413.4		
Pseudo R^2	0.367			0.372			0.361		

Notes: Marginal effects after multinomial logit estimation. *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the 95% spatial effect of the district, after estimation of the geoadditive models.

Chapter 5

Entrepreneurship and Welfare

We examine returns to entrepreneurship using a standard measure of welfare, the per-capita consumption expenditure of the household, in a developing country. Using simultaneous quantile regressions, we find that entrepreneurs who also employ others have the highest returns in terms of consumption, while those entrepreneurs who work for themselves, the self-employed individuals, have slightly lower returns than the salaried employees. A process of endogenous non-random selection into occupation is observed. In particular, the ablest individuals select into entrepreneurship (employers) followed by salaried employment, self-employment and casual labor, in that order.

5.1 Introduction

Although recent research postulates a positive link between entrepreneurship and economic growth in developed countries,[1] the role of entrepreneurship in the economic processes of less developed economies has received little attention.[2] Using a direct measure of welfare, this study links the occupational choice of individuals to their welfare. While most studies use income measures to examine the returns of occupations (Hamilton, 2000), this paper uses consumption measures.[3]

[1]Entrepreneurship capital, measured by quantum of new firm entry relative to the population, is positively related to the growth of the regions (Audretsch et al., 2006). Aghion et al. (2004) find that entry of firms has productivity enhancing effects on incumbents. Also see Baumol (2002) and Berkowitz and DeJong (2005) .

[2]Some theoretical studies examine the relationship between occupational choice and economic development (Banerjee and Neuman, 1993; Ghatak and Jiang, 2002).

[3]Income is usually highly correlated with consumption in a developing country. Furthermore, analyzing the consumption patterns itself has the advantage that variation is not so high as in income data. However, as people with higher incomes are likely to have greater savings, analyzing the consumption patterns for welfare comparisons may make their returns appear flattened to some extent.

In addition, the underlying process of selection into occupations and subsequent returns in terms of welfare are also examined to see whether people are compelled to opt for low-productivity self-employment or whether there is voluntary self-selection in a developing country context, in particular with respect to unobserved abilities.

The outline of the paper is as follows. The next section provides the theoretical background on the occupational choice and welfare and presents the hypotheses. The third section discusses the methodology, in particular quantile regressions (Koenker and Hallock, 2001) and selection models after multinomial logit estimation (Bourguignon et al., 2007). The fourth section presents the data and descriptive statistics. The fifth section discusses the empirical analysis linking occupation and welfare. The final section summarizes the main findings linking occupation and welfare, and concludes this chapter.

5.2 Theoretical Background

5.2.1 Occupation, Welfare and Economic Development

As Banerjee and Neuman (1993, p. 275-276) discuss in their seminal paper, *Occupational Choice and the Process of Economic Development*,

> There are several ways in which the dynamics of occupational choice influence the process of economic development. Most obvious of them is the effect on the distribution of income and wealth. In so far as the distribution can effect saving, investment, risk braining, fertility and the composition of demand and production, there is a clear link with the economy's rate of growth and hence with development in its narrow sense. . . . Conversely, the process of development also effects the structure of occupations. It alters the demand for and supply of different types of labor and, hence, the returns to and allocations of occupations. It transforms the nature of risk and the possibilities for innovations. And, of course, it changes the distribution of wealth. Since one's wealth typically effects one's incentives to enter different occupations, the effect on the wealth distribution generates a parallel effect on the occupational structure.

They show that the initial wealth distribution of a population enables or disables people from starting new enterprises and thus has an impact on the equilibrium returns to occupations and the long run distribution of wealth.[4] Occupational choice influences the capacity to save and bequest, thus leading to persisting hysteries, even though it is endogenously determined by the wealth distribution. More recently, Ghatak and Jiang (2002) argue that hysteries, long run dependence on the initial conditions, depend on the size of the threshold level of wealth needed to start enterprises relative to the productivity of the modern and the subsistence technologies.

Murphy et al. (1991, p.505) suggest that individuals move to occupations that have greatest returns to their talents. They further show that the economy grows fastest when the most able individuals become entrepreneurs. They claim that it is this process of allocation of talent that determines the growth trajectory. In their words, "which activities the most talented people choose can have significant effects on the allocation of resources. When talented people become entrepreneurs, they improve the technology in the line of business they pursue, and as a result, productivity and income grow."

A key assumption of the Banerjee and Neuman (1993) model is the inherent hierarchy of occupational choice. The most well endowed individuals become entrepreneurs, the next best self-employed, and the others become workers or subsistence workers. This is yet to be subjected to empirical validation. The possibility of self-employment being worse off in the hierarchy relative to wage workers, or at least being equal in returns, an assumption that would directly follow from the dual labor market theory hypothesis, would indicate a completely different set of convergence dynamics in the Banerjee and Neuman (1993) framework.[5]

[4]They model four feasible occupational choices. Individuals can be self-employed or entrepreneurs, work for entrepreneurs or subsist. Self-employed are equivalent to workers but are not monitored by an external agent (the entrepreneur). They provide the labor for operating their firm and own the total output. Entrepreneurs, however, hire workers and monitor them though they themselves are not monitored. Employment contracts emerge only in the presence of a initial wealth distribution that has a certain degree of inequity, otherwise everyone chooses to be self-employed.

[5]The literature on less developed countries (LDCs) traditionally identifies self-employment as a distressed residual of people rationed out of jobs in the formal sector (Ranis and Fei, 1961; Harris and Todaro, 1970), though in contemporary literature the nature of the labor market in developing countries is highly debated. More recently, it is believed that the informal sector consists of voluntarily self-selected competitive workers as well as disadvantaged individuals (Gindling, 1991; Magnac, 1991; Cunningham and Maloney, 2001; Maloney, 2004; Fields, 2005; Günther and Launov, 2006). Pratap and Quintin (2006) argue that there is no evidence of market segmentation in developing country labor markets.

5.2.2 Occupational Selection and Determinants of Welfare

Occupational choice is generally modeled as a utility maximizing decision of individuals (Lucas, 1978; Kihlstrom and Laffont, 1979).[6] While many models in the economics of entrepreneurship assume that individuals become self-employed as they expect higher returns relative to wage employment (Blau, 1987; Rees and Shaw, 1986; Parker, 1996), the labor and development literature suggests that in the LDC context, people are forced into self-employment in the absence of viable economic opportunities. However, Hamilton (2000) notes that entrepreneurs may trade lower earnings for the nonpecuniary benefits of business ownership. He finds no evidence of the earnings differential being a result of selection of low ability employees into self-employment. Further, he argues that self-employment offers significant nonpecuniary benefits, such as being one's own boss for most entrepreneurs. Evans and Leighton (1989b) suggest that individuals who prefer greater autonomy are more likely to be entrepreneurs. Blanchflower and Oswald (1998) show that business owners have greater job satisfaction than paid-employees. According to Boháček (2006), as successful firms grow over time, individuals may enter self-employment even if the returns are lower. He claims that business households may have a higher saving rate in order to relax the wealth constraints in financing entrepreneurial projects and to operate their firms at an optimal size.

Thus three compelling theories of returns to self-employment choice have emerged. First, the expected utility view claims that individuals choose self-employment when they expect higher returns in self-employment relative to wage-employment. Second, the non-pecuniary benefits view argues that individuals select into entrepreneurship even when the returns are lower, for non-pecuniary benefits such as being one's boss. Finally, the traditional low-productivity view suggests that individuals are compelled into self-employment in the absence of viable economic alternatives.

We have two compelling inter-linked hypotheses. We hypothesize that, given

[6] There are two main methods to model the returns of occupational choice. First is to estimate a mincer type wage equation for each occupation. Second is the structural probit method that estimates the reduced form probit and determines the wages corrected for selection. The sign of mill's ratio indicates the nature of selection. The predicted earnings differential are used to re-estimate the probit equation to predict self-employment choice as a function of expected utility (Rees and Shaw, 1986).

the occupational structure of individuals in an economy, the welfare returns to entrepreneurship are heterogeneous across the distribution. Entrepreneurs are either employers or solely self-employed. While employers are entrepreneurs who employ others as well, the self-employed only work for themselves. Employers are likely to have higher returns than salaried employees and self-employed people. However, the expected relative returns to self-employment compared to returns of salaried employment are unclear. This leads us to the second hypothesis. If self-employment is characterized by high skilled individuals voluntarily selecting into this occupation, the relative returns are likely to be higher than the returns to salaried employment. In the presence of segmented labor markets or if self-employment is a choice of low-skilled people, the returns to self-employment are likely to be lower. This hypothesis is tested in the empirical section using selection models. The issue of returns to occupation taking into consideration the selectivity issue, has been examined in the literature by many studies. For instance, Hamilton (2000) tests for the selectivity issue considering self-employment as a binary variable. He finds that individuals of higher abilities select into entrepreneurship (also see Rees and Shaw, 1986).[7]

We hypothesize that there are locational as well as sectoral differences in returns to the entrepreneurship choice. We also control for a number of other factors that have been found to influence the per-capita consumption of the households. Nelson (1988) shows the existence of economies of scale in all adult households. Such economies of scale are found to be more important in the consumption of shelter and less so in the consumption of clothing and transportation. Economies of scale have a range of 0 to 1, with 1 indicating no economies of scale, and the measure of welfare considering the economies of scale is equal to per-capita income of the household in this case.[8] Furthermore, a vast literature is concerned with equivalence scales in the measurements of welfare for comparisons across households. Households with the same income but different structures, in terms of the number of children and old people are likely to have different consumption pat-

[7] Our model extends these studies to more than two occupations. In our analysis, individuals can select into one of the four occupations described earlier.

[8] We, however, use the standard measure of welfare, per-capita expenditure on consumption. One of the reasons for using the standard measure in the analysis is that although we use all nonagricultural households in the beginning, we restrict the rest of the analysis to those households where the sole economically activate member is the household head. Thus, it is plausible to assume economies of scale close to 1 in such households.

terns. For instance, Lanjouw and Ravallion (1995, pp 1431-1432) suggest that the relationship between poverty and household size depends on the weight attached to child and adult welfare. They find evidence against the conventional view that household size is negatively correlated to welfare when Rothbarth method based on non-food spending is used as a measure of welfare while a measure based on child stunting indicates that larger households tend to be poor.[9] We thus control for the household demographic structure in the analysis.[10]

There are compelling reasons to hypothesize that female headed households are likely to be poorer. Dreze and Srinivasan (1997), using an earlier survey of India's National Sample Survey Organization(NSSO), also find that households that are female headed are more likely to be poor. Jenkins (2000) finds that changes in labor earnings from persons other than the household head, changes in non-labour income, changes in the earnings of the household head, and household composition are important determinants of the poverty dynamics. For these reasons, although we first analyze all nonagricultural households, we subsequently restrict the analysis to households that have only the household head economically active. Miles (1997) finds that uncertainty, education, and location matter. Using both durable and non-durable goods in the welfare measure, Glewwe (1991) finds high returns to education in urban areas compared to rural areas in Côte d'Ivory.[11] We also examine the returns to occupations in urban and rural areas separately.

5.3 Methodology

We use two empirical methods to test the hypotheses of heterogenous returns of occupation across the welfare distribution and potential non-random endogenous

[9] Browning (1992) notes though children may be endogenous to whatever we are interested in modeling, this can be circumvented by assuming that fertility is exogenous. See Browning and Crossley (2001) for recent developments in the life cycle model of consumption. More recent way of measuring poverty using perceptions of consumption adequacy are addressed in Pradhan and Ravallion (2000).

[10] In the Indian context, Dreze and Srinivasan (1997) find that the poverty head-count ratio is very robust to alternate equivalent scales. We also test the robustness of the results using adult equivalent scales. The results are not reported in the paper but are available on request from the author.

[11] Benito (2006) finds that unemployment risks leads households to defer consumption using British Household Panel. The dataset we have, however, does not allow for such controls. We control for all these factors, other than uncertainty.

selection into occupations.

5.3.1 Quantile Regressions

For testing the former, we employ quantile regressions (see Koenker and Hallock, 2001, and references therein). As Hamilton (2000) observes, superstar model of Rosen (1981) suggests that comparison of mean earnings of workers in self-employed sector and in wage sector would be highly influenced by few entrepreneurial superstars. Thus, mean earnings do not really characterize the returns of the majority of self-employed. The greatest advantage of using quantile regressions is their ability to show snapshots of relationships across different quantiles of the distribution and not only at the mean. This enables a comparison, for example, between the poorest selfemployed individual with the poorest salaried employee at the lowest quantile and the richest selfemployed individual with the richest salaried employee at the highest quantile.

5.3.2 Selection Models for Multiple Outcomes

In order to test for the selectivity issue, we employ the methods of selection bias correction based on the multinomial logit (see Bourguignon et al., 2007, for a survey). Lee (1983) and Dubin and McFadden (1984) suggested ways to extend the pioneering work of Heckman (1979) for the case of multinomial logit into a selection model. In what follows in this section, we summarize the method (refereed to as BFG) and describe the basic idea behind modeling the selection process, after multinomial logit estimation.[12]

Consider the following model:

$$y_1 = x\beta_1 + u_1, \tag{5.1}$$

$$y_j^* = z\gamma_j + \eta_j, \text{ for } j = 1, \dots M \tag{5.2}$$

where the disturbance u_i has mean 0 and variance σ^2, conditional on x and z. j is a categorical variable that describes the choice of an economic agent among M alternatives based on 'utilities' y_j^*. z determines the alternatives and x determines the outcome variable. The outcome variable y_1 is observed only if the first category

[12]This section is based on Bourguignon et al. (2007).

is chosen and this happens if $y_1^* > max(y_j^*)$ for all $j \neq 1$. By defining $\varepsilon_1 = max(y_j^* - y_1^*) = max(z\gamma_j + \eta_j - z\gamma_1 - \eta_1)$, we have $\varepsilon_1 < 0$. Assuming that the IIA hypothesis holds, that is, η_js are independent and identically Gumbel distributed, $G(\eta) = exp(-e^{-n})$ is the cumulative function and $g(\eta) = exp(-\eta - e^{-n})$ is the density function. This gives the multinomial logit model with

$$P(\varepsilon_1 < 0|z) = \frac{exp(z\gamma_1)}{\sum\limits_j exp(z\gamma_j)} \tag{5.3}$$

Maximum likelihood estimation then gives consistent estimates of γ. The selection problem involves the estimation of β_1, given that u_1 may depend on the the η_js. By defining $\Gamma = \{z\gamma_1, z\gamma_2, \ldots, z\gamma_M\}$ and generalizing the Heckman (1979), bias correction can be based on the conditional mean of u_1,

$$E(u_1|\varepsilon_1 < 0, \Gamma) = \int \int\limits_{-\infty}^{0} \frac{u_1 f(u_1, \varepsilon_1|\Gamma)}{P(\varepsilon_1 < 0|\Gamma)} d\varepsilon_1 du_1 = \lambda(\Gamma) \tag{5.4}$$

where $f(u_1|\varepsilon_1 < 0|\Gamma)$ is the conditional joint density.

If P_k be the probability of preferring an alternative k, then P_k is given by:

$$P_k = \frac{exp(z\gamma_k)}{\sum exp(z\gamma_j)} \tag{5.5}$$

As the M components of Γ and the M probabilities are invertible, a unique function exists such that $E(u_1|\varepsilon < 0, \Gamma) = \mu(P_1, P_2, \ldots, P_M) = \lambda(\Gamma)$. Then consistent estimation of β_1 can be based on

$$y_1 = x_1\beta_1 + \mu(P_1, P_2, \ldots, P_M) + w_1 \tag{5.6}$$

Bourguignon et al. (2007) survey the methods developed so far, including Lee (1983), Dubin and McFadden (1984) and Dahl (2002). They propose a variant of the Dubin and McFadden (1984) model by removing the restriction that all the correlation coefficients sum-up to zero. Using monte-carlo simulations they find that the Dubin and McFadden (1984) method and the variants of this method proposed by them have higher efficiency. Following them, we use the BFG variant of the Dubin and McFadden (1984) for the empirical analysis. Bourguignon et al. (2007) find that having the restriction in Dubin and McFadden (1984) is a source of bias when it is incorrectly specified. They recommend that as many correction parameters enter the outcome equation as the number of alternatives in the selection equation. Furthermore, they claim that system is still identified as the all the correction terms are non-linear in probabilities. Their study positively tests for the little efficiency loss when all the non-linear probabilities are introduced.

In the BFG method, the selection terms are given by,

$$E(\eta_1^*|y_1* > max(y_s^*), \Gamma) = m(P_1), \text{ for } s \neq 1 \tag{5.7}$$

$$E(\eta_j^*|y_j* > max(y_s^*), \Gamma) = m(P_1)\frac{P_j}{P_j - 1}, \text{ for all } j > 1 \tag{5.8}$$

and the outcome conditional on $j = 1$ is,

$$y_1 = x_1\beta_1 + \sigma \left[r_1^* m(P_1) + \sum_{j=2...M} r_j^* m(P_j)\frac{P_j}{P_j - 1} \right] + w_1 \tag{5.9}$$

Thus, the selection terms entering the regression may be treated as corrections for the underlying process of alternatives being chosen based on latent utilities. A positive coefficient of a selection term $m(P_j)$ in the equation estimating β_1 suggests that an upward bias is caused by the alternative j being non-randomly chosen. In our case, this suggests that people of lesser abilities have selected alternative j and this is resulting in an upward bias in the estimation of β_1 and is corrected by introducing the selection term $m(P_j)$. A negative coefficient of the selection term $m(P_j)$ in the equation estimating β_1 similarly suggests that there is a downward bias caused by people with higher abilities not choosing alternative 1 but choosing the alternative j. If the coefficient of $m(P_1)$ is positive, this suggests that there is an upward bias in the estimation of β_1 caused by people with higher abilities choosing alternative 1.

5.4 Data

The data used for the analysis comes from the 60th round employment-unemployment survey of the National Sample Survey Organization (NSSO) of India. We only consider those households where the household heads have reported to be self-employed (includes own account workers and employers), salaried employees, casual laborers, and unemployed. We restrict the sample to those who are older than 15 years but younger than 70 years. We then consider only those households who work in nonagriculture. The final sample consists of 26,485 households. In these households, 13,782 households have only the household head economically active.

Figure 5.1 shows that kernel density plots of log per-capita consumption of households with heads working as self-employed, salaried, employers and laborers. While the distribution plots of salaried employees and employers are to the right of the self-employed, the density of the laborers is centered to their left. Furthermore, the plots show that the inequality observed in the employer group is substantially higher than others.

5.5 Empirical Analysis

As mentioned in the methodology section, the hypotheses are tested through two econometric frameworks. First, heterogeneity in returns to occupations across the distribution is examined using simultaneous quantile regressions. Second, the process of endogenous non-random selection of individuals into different occupations is tested using selection models after discrete choice models with multiple outcomes.

5.5.1 Entrepreneurship and Welfare

Household Level Analysis

As Browning and Lusardi (1996, p. 1801) note, "although consumption changes are uncorrelated with anticipated income changes, the actual path of consumption may follow quite closely the actual path of income if the latter displays some persistence." Hence, the consumption and income paths are assumed to be correlated. The empirical strategy is to estimate simultaneous quantile regressions, using the log of per-capita consumption of the household as dependent variable.[13]

[13] Wodon (2000) also uses per-capita consumption. Many alternate strategies to construct welfare measures that are comparable across households exist. For instance, Lazear and Michael (1980) develop a technique that converts families of different structures into single person equivalents. Also see Muellbauer (1974) and Deaton and Muellbauer (1980, 1986) for a theory of equivalence scales. The identification of correct equivalent scales is still an unresolved issue (Deaton and Paxson, 1995).

The occupations of the members of the household enter the regression as independent variables. A series of controls that are found to influence the consumption of the household by earlier studies are introduced in the estimation. In particular, personal characteristics of the household head, demographics of the household including the proportion of children, adults and old persons, educational background of the members, urban location and land possessed are introduced as control variables.[14] State level dummies are also included to control for regional effects.

The results presented in Table 5.1 suggest that the entrepreneurship has a distinct relationship with welfare.[15] As mentioned earlier, economically active people have one of the five primary occupations. They are either employers, self-employed, salaried employees, casual laborers or unemployed. In this estimation, the left out category for the occupation variables is the proportion of economically active individuals in a household who are self-employed. As the positive coefficients suggest, households that have a higher proportion of employers and those that a higher proportion of salaried employees have higher per-capita consumption levels than self-employed households. However, households that have a higher proportion of casual laborers and unemployed people have lower welfare levels than self-employed households. This suggests the existence of a welfare hierarchy, that is determined by occupational choices of members of the household.

The coefficients of controls variables are in accordance with what might be expected. Households with older household heads are more likely to have higher consumption rates and female headed households are poorer across quantiles. Female headed households are worse off most at the lowest quantile of the distribution. Households with a higher proportion of educated individuals have higher consumption rates and the returns are increasing along the quantiles as well as along higher levels of education. The quantile regression technique enables comparisons of the returns to characteristics at different quantiles of the distribution. In particular, the quantile plots in Figure 5.2 show that the estimates based on the quantile regression are non-linear, although for the occupational variables

[14] Land variables proxy the wealth of the household. Wodon (2000) suggests that the land possessed by a household is also a determinant of the welfare. We also check for the robustness of the results with the land variables excluded from the analysis. Given that we have only nonagricultural households in the data set, the problem of endogeneity of the land variables is less severe.

[15] The estimates of the inter-quantile regressions are available from the author.

the estimates are mostly in 90% confidence intervals of the OLS estimates. As Figure 5.2 suggests, employers are increasingly better off at higher quantiles than self-employed workers. Salaried employees who are in the middle of the distribution are most different than those at the extreme quantiles relative to the self-employed. At higher quantiles, casual laborers are increasingly worse off than the self-employed, and a similar phenomena is observed for the unemployed.[16] Nonlinearities with respect to high school and university education are distinct, so OLS estimates would not have given the right picture. The returns to education are comparatively much higher at higher quantiles. Figure 5.3 shows the estimates for the other control variables that represent the demographics and the characteristics of the household.

The proportion of children less than 15 years old in the household has a significant negative effect at the lowest two quantiles, but vanishes at higher quantiles. However, the proportion of old people in a household significantly increases the per-capita consumption expenditure. A 1% increase in the proportion of elderly people, increases the per-capita consumption by 18% at the lowest quantile and 38% at the highest quantile. The proportion of females has an insignificant effect in the lower two quantiles but has a significant positive effect at higher quantiles. Thus, at median, a 1% increase in the proportion of females, increases the per-capita consumption by 4.4% and at q(.9), by 9%. The plots of the household size variables show that the relationship between household size and welfare of the household is consistent with earlier studies that households of larger size have a lower per-capita consumption expenditure. However, the household size squared term is positive and increases across quantiles, indicating that households of larger size become worse off along the quantiles, but at decreasing rates. Thus, a convex relationship exists between household size and welfare, with households in the middle of the distribution showing the greatest negative effect of size on per-capita consumption. This could be the result of higher economies of scale at the tails of the income distribution.

[16] However, the unemployed variable slightly moves upward at the highest quantile but remains significantly negative.

Analysis Restricted to Household Heads

One of the main limitations of the analysis of the household level occupation data, is the simultaneous determination of occupation of the household members leading to potential endogeneity of the occupation variables. There are three possible endogeneity problems. First, occupation of members of household may not be independent of the occupation of head of the household, in the presence of intra-household dependence of occupation choice. Second, personal characteristics such as age and educational background of the household members may determine their occupational choice. Third, current income may determine future occupational choice. However, given the cross-sectional nature of our database, we are not in a position to test this third issue. Furthermore, data on wages are available for only a small fraction of individuals in the database. However, we control for wealth of the household and thus the results are conditional upon this factor. The first two issues concerning endogeneity are addressed using the following empirical strategies.

In order to reduce the potential endogenous determination of the occupational choice of the household based on the occupational choice of the household head, we re-estimate the simultaneous quantile regressions for a restricted sample of households that have only the household head as the economically active individual in Table 5.2. This is more likely to give the pure effect of occupation, and entrepreneurship in particular, on household welfare.[17] In order to address the second issue, we estimate models with corrections for selectivity in subsection 5.5.2.

We also drop the unemployed as there are only 90 heads of household who are unemployed. Furthermore as a check for robustness of the results in Table 5.1, we also control for the industry sector of the individuals in Table 5.2 as there may be sectoral differences in returns to self-employment.[18] The base category for the occupation variables is "salaried employee". The estimation results are consistent with the estimations of the quantile regressions presented in Table 5.1. The results presented in Table 5.2 confirm the welfare hierarchy that the earlier regression

[17]An alternate strategy would be to use instrumental variables techniques and instrument for the occupation of the household members using the occupation of the household head. However, as household heads themselves are in the sample and the occupation of their parents is not known, this is not viable.

[18]As the dataset had unemployed people earlier, industry effects could not be controlled.

suggested. Households headed by employers and salaried individuals have a higher per-captia consumption than households headed by self-employed individuals and casual laborers, after controlling for other factors that influence household welfare. The magnitude of the coefficient of "employer" suggests that households headed by entrepreneurs who employ others have the highest consumption levels. Although the coefficient of salaried employees is positive, it is small, and salaried employees are only slightly better off than those who are self-employed.[19] The casual laborers are last in the hierarchy.

Table 5.2 suggests that at lower quantiles, informal education has a significant positive effect on the per-capita consumption. The returns to primary school education increase along the quantiles. It is seen that at the lowest quantile, q(.1) primary schooling increases the per-capita consumption of the household by 14%. The coefficient however is higher at the highest quantile, q(.9), where it raises the per-capita consumption of household by 19%. A similar effect is observed for other education variables. If household head has high school education, per-capita consumption expenditure increases by 23% at the lowest quantile and 36% at the highest quantile. Similarly, if the household head has university education, the per-capita consumption of the household increases by 41% at the lowest quantile and 73% increase at the highest quantile. Thus, education has a positive effect on the per-capita consumption and increases as individuals move from the lower to higher quantiles. The returns to technical degree/diploma are also positive and increasing as individuals shift from the lower to the higher quantiles.[20] The estimates of the control variables are in accordance with the hypotheses and are consistent with the estimation in Table 5.1.

Entrepreneurship, Poverty and Inequality

Per-capita consumption of individuals is predicted after estimating the quantile regression at different quantiles.[21] The cumulative distribution plots of occupation wise predicted values are shown in Figure 5.4. As the plots suggest, per-capita

[19] Hamilton (2000) postulates that lower returns to self-employment may be attributed to individuals choice of freedom leading them to select self-employment.

[20] As there are very few individuals with technical degrees or diplomas, we merge these into one variable.

[21] The log-inverse transformation of the predicted values gives the value of the normalized per-capita consumption expenditure. These transformed values are used in the poverty and inequality analysis.

consumption level is determined by occupation status. Entrepreneurs who are employers have the least probability of being under the poverty line.[22] Households headed by employers are followed by those headed by salaried employees, self-employed and the casual laborers, in that order, at all quantiles. The plot clarifies the status of the self-employed; they appear sandwiched between the salaried employees and the casual laborers. A direct implication of this observation is that, conditional on other characteristics, individuals in the informal sector, primarily comprising of the self-employed and the casual laborers, have lower returns to their occupations. Furthermore, the dataset is split into formal and informal sectors, with laborers and self-employed in the informal sector and salaried employees and employers in the formal sector, the plots suggest that in both sectors, entrepreneurship in the form of employers in the formal sector and self-employed in the informal sector entails higher relative consumption and an escape from poverty. The Lorenz curves in 5.5(a) suggest that inequality is highest amongst the households with self-employed head. As the generalized Lorenz curves in 5.5(b) suggest, the employers group has a distribution preferred by all equity respecting social welfare functions relative to the distributions of the other occupations. This is followed by the distribution of the salaried employees, self-employed people and the casual laborers.

Furthermore, we analyzed occupational choice as a determinant of poverty of households using a probit model. The poverty line was assumed to be given by half the median of per-capita consumption of the household.[23] The results suggest that households headed by employers, self-employed and salaried employees are less likely to be under the poverty line. Households headed by casual laborers are most likely to be under the poverty line, after controlling many characteristics that are likely to influence their poverty status.[24]

Rural and Urban Estimations

As we hypothesize that the returns to occupations in urban areas might be different than returns to occupations in rural areas, we estimate quantile regression

[22] The plot does away with the necessity of having a poverty line to examine the poverty status of people based on their occupation and indicates the relative positions of the various occupation groups, in which we are primarily interested.

[23] Using an alternate poverty line based on the number of adults has not significantly altered the main inferences.

[24] For brevity these results are not reported here but are available from the author.

welfare for rural and urban subsamples separately. It is seen in Table 5.3 that though the hierarchy is evident again in rural areas, the coefficients of the self-employment variable are insignificant in urban areas in Table 5.4. The difference between self-employed and salaried individuals disappears in urban areas at three of the five quantiles. The presence of several self-employed professionals in urban areas may be a reason for this. These results hold even after controlling for industrial sectors. Thus, it is possible that urban areas provide a more suitable environment for self-employment activities, while in rural areas, self-employment is primarily characterized by activities that inherently have lower returns. Thus, the results suggest that in the urban informal sector (UIS), there is no evidence that the returns to self-employment are lower than the returns to wage employment. It is also seen that in rural areas, the returns to education are lower than in urban areas. For instance, while university education increases per-capita consumption by 30% in rural households (at the lowest quantile), it increases the per-capita consumption by 46.5% in urban households. Furthermore, while university education increases the per-capita consumption by 49% in rural households (at the highest quantile), it increases the per-capita consumption by 81% in urban households. The returns to primary school education and high school education are also higher in urban areas. The returns to technical education are also higher in urban areas.

In summary, these estimations suggest that self-employment in urban areas entails returns similar to salaried employment. Thus, these results support the hypothesis that self-employment in a less developed country is a blend of low-productivity and high-productivity activities.

5.5.2 Endogenous Non-random Occupational Selection

The main assumption underlying the quantile regression analysis is that occupations are exogenous. However occupation is rarely determined exogenously. Though we control for simultaneous determination of occupation within the household (in subsection 5.5.1) by selecting only those households where the household head alone is economically active, occupation itself might be endogenously determined by individual characteristics and their cultural contexts. Hence, for analyzing occupational choice, it is also appropriate to consider the selectivity issue in order to control for endogenous non-random selection into different occupations. This approach also provides insights into the selection of individuals into different occupations based on their unobserved abilities.

We estimate consumption functions separately for each of the occupational groups. A chow test rejects the null hypothesis that the coefficients of the consumption functions are same across the occupational groups. Hence, we use the method proposed by BFG to consider occupational choice as a selectivity problem in estimating the determinants of household welfare. Using this technique, the consumption equations are re-estimated for the four types of households.[25] Once again, we use the restricted sample of households in which only the household head is economically active.[26]

The multinomial logit selection equation is given in Table 5.9. The dependent variable is the primary occupation of the household head. The Sargan test rejects the poolability of the outcomes. The set of independent variables is same as the set of variables used in chapter 2 and chapter 3. We have personal characteristics, educational background, household variables such as urban location and land possessed, religion and caste variables and regional dummies in the estimation. The estimation results of the selection equation are very similar to the estimation results in chapter 3 and hence we do not re-interpret them here. The selectivity corrected estimates for occupational groups are presented in Table 5.5, Table 5.6, Table 5.7 and Table 5.8. The following empirical approach is adopted to ensure proper identification of the models. For each occupational group, we estimate three different selection models after estimating the multinomial logit equation. In the first model, the outcome equation consists of demographics of the household and household size variables alone. Thus, all the other variables including the personal characteristics, land possessed, education, regional dummies and the religion and caste variables act as instruments for identifying the model. This is to ensure that we avoid the problem of multi-collinearity that arises as

[25] Bourguignon et al. (2007) using Monte Carlo simulations show that the selection model after multinomial logit estimation provides good correction in the outcome equation even if the IIA hypothesis is violated.

[26] As there are only 90 households in a sample of 14000 households that have an unemployed individual as a head, we drop this category in subsequent analysis. Keeping this category creates problems in the convergence of the multinomial logit selection equation, as many of the states have no such individuals. The number of households headed by employers is also very small, but they are about 250 such households and by including only those state dummies in the regressions that have such households, we are able to obtain sufficient convergence of the multinomial logit equation.

variables in the selection equation simultaneously enter the outcome equation. Hence, in the first model, there are no variables common to both the selection and outcome equations. In order to test for robustness, we estimate a second specification by introducing household characteristics such as urban location and land variables that proxy household wealth, and personal characteristics in the outcome equation. In the third specification, we introduce all the variables that form the selection equation, aside from religion and caste variables and regional dummies. Thus, religion, caste and region variables act as instruments in the third specification. We thus estimate three models for each occupational group, in order to check robustness of the selectivity coefficients.

The results given in Table 5.5 provide selectivity corrected estimates of the consumption function for households headed by self-employed people. In the first column, demographic characteristics of the household and household size variables are introduced. The negative coefficient of the selectivity coefficient m(Employer) in the estimated consumption function of the self-employed group suggests a downward bias caused by people selecting into the employers group. If consumption is assumed to be correlated with unobserved abilities, this suggests that a process of non-random selection of individuals with higher unobservable abilities into the employer category is causing a downward bias in the consumption function of the self-employed group. Similarly, the positive coefficient on the selectivity variable m(Casual) suggests that there is an upward bias caused by non-random selection of people with lesser unobservable abilities into the casual labor category. The positive coefficient of m(Salaried) suggests that a positive bias is caused by non-random selection of people with lower unobserved abilities into the salaried category. However, in contrast to the selection coefficient of the casual labor category, the selection coefficient for the salaried employee category is much smaller.

In the remaining analysis, the estimation results of the first model for the employers, salaried employees, and the laborers are only discussed for brevity. The selection coefficients in the consumption function of the employers in Table 5.6 suggest an upward bias caused by the selection of individuals with lower unobservable abilities into the self-employed and laborer categories. There is no evidence of non-random selection associated with the salaried class in this case. Similarly, the consumption function of the salaried group in Table 5.7 suggests that a downward bias is caused by people with higher unobserved abilities selecting into the employer category. Furthermore, an upward bias is caused due to the selection of individuals with lower unobservable abilities into the self-employed group and people with the lowest abilities selecting into the labor class. The results for laborer category, however, show that there is no bias caused by selection into the self-employed or the salaried class but there is a negative bias caused by non-random selection of people with higher abilities into the employer category.

The positive coefficient of a selection term on a particular category in the consumption of that category indicates that people with higher unobservable abilities have selected into that group and this is causing an upward shift in the consumption function. A negative coefficient of the selection term on a particular category in the consumption function of that specific group would suggest that people with lower abilities have selected into the group and this is causing a downward bias in the consumption function. Hence, the positive term, m(self-employed) in the self-employed equation in Table 5.5 suggests that people with higher unobservable abilities have moved into the self-employed group and this results in an upward shift of the consumption function. Thus, the table suggests non-random selection of people with lower-abilities into the casual labor and salaried employment as well as positive self-selection into self-employment. Similarly, the positive coefficient m(salaried) in the consumption equation of the salaried group in Table 5.7 suggests that people with higher unobservable abilities have selected into the salaried class and the estimation has an upward bias as low-ability people have selected into casual labor and self-employment. Thus, the selection models confirm the hypothesis that self-employment is a blend of the competitive and disadvantaged sectors (Fields, 2005; Günther and Launov, 2006).

The estimates of the second and third specifications in all the four tables consistently support the inferences drawn based on the first model though there are some deviations. For instance, in the Table 5.8, the third specification (given in the table as Model III) shows a significant positive selection term for the salaried variable, m(Salaried) in the estimated casual labor consumption function. This suggests that people with lower abilities are selected into the salaried class leading to an upward bias. However, this result is quite counter-intuitive and could be purely due to collinearity that is caused by the presence of many variables in the outcome equation that are also present in the selection equation.

In summary, this analysis gives insights into the selection process of individuals differing in their abilities into different occupational categories and its subsequent impact on the consumption functions of each group. The analysis shows the presence of non-random selectivity of individuals, based on their unobserved abilities, into different occupations. In particular, the selection corrected consumption functions of the self-employed and the salaried employees suggest that there are biases caused by people selecting into the employers group, followed by the salaried category, the self-employed group and the casual labor group based on their unobserved abilities, in that order. The selection corrected consumption function of the employer group suggests the presence of bias caused by the selection of individuals with the lowest abilities becoming self-employed and casual laborers and the selection corrected consumption function of the casual laborer group suggests the presence of bias as people with the higher unobservable abilities move into the employer category.

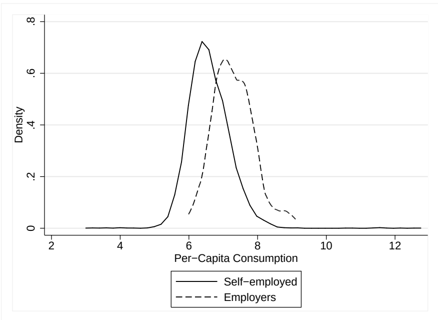
5.6 Conclusion

This paper presents important contributions to the literature on the economics of entrepreneurship. We examine extensively the welfare consequences of entrepreneurship in a developing country, an area of study that received little attention. We use recent empirical methodologies to examine returns to entrepreneurship and test for the process of endogenous non-random selection into occupations based on unobserved abilities.

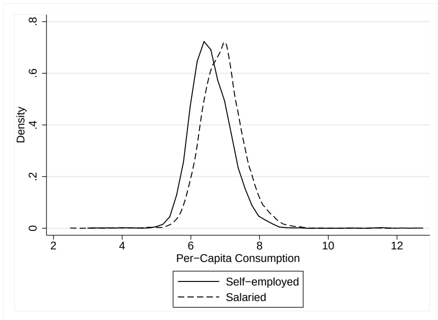
We examine the returns to entrepreneurship in India and find that entrepreneurship is characterized by different components that co-exist. Using simultaneous quantile regressions, we find that employers, those entrepreneurs who also hire others, have the highest returns in terms of consumption, while the self-employed, those entrepreneurs who work for themselves, have slightly lower returns than the salaried employees. This evidence suggests that self-employment is not a better occupational option relative to salaried employment. This contradicts a key assumption of many theoretical studies including Banerjee and Neuman (1993).

We do find evidence that the self-employed are more likely to escape poverty, along with the salaried employees and those entrepreneurs who are employers. The results are robust even after controlling for industrial sectors. The results suggest that the gap between the salaried employees and the self-employed is higher in the rural areas than in the urban areas. Lower returns to self-employment, however, do not completely support the theory that people are compelled into self-employment, as even in developed countries, it is found that self-employed have lower returns. Hamilton (2000), for instance, argues that self-employment is associated with freedom, and hence individuals might opt for it, in spite of lower returns.

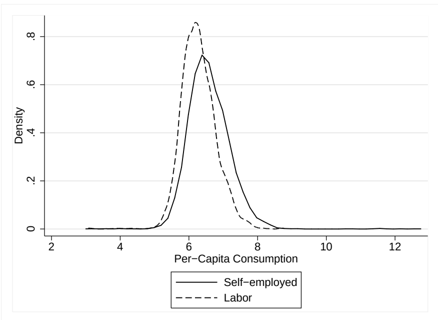
Given the potential non-random selection of individuals into different occupational categories, we also use selection models after discrete choice models with multiple outcomes, to examine the selection process and its effect on the consumption patterns in the occupation subgroups. We find evidence of endogenous non-random selection into occupation and obtain selection corrected estimates to returns to occupations. In particular we find that the ablest of individuals select into entrepreneurship and become employers, followed by salaried employment, self-employment and casual labor, in this order. Furthermore, positive self-selection into self-employment is also observed. Though this is consistent with the quantile regressions in the first part of the analysis, it is observed that the they overestimate returns to the individual characteristics, if the selectivity issue is not considered. While the quantile regression considers all individuals together and examines returns to characteristics at different quantiles, the selection model estimates separate regression curves for each of the occupational groups, at the mean of the independent variables. Extending the selectivity correction issue into the quantile regression framework is an interesting avenue for future research.



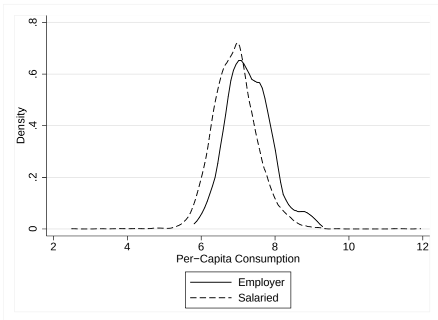
(a)



(b)



(c)



(d)

Figure 5.1: Consumption and Occupation(Un-normalised)

Table 5.1: Households, Occupation and Consumption

Estimates of Simultaneous Quantile Regression					
Independent Var.	q10	q25	q50	q75	q90
Occupation					
Prop. Employers	0.336***	0.342***	0.405***	0.454***	0.461***
	(0.038)	(0.047)	(0.039)	(0.035)	(0.045)
Prop. Salaired	0.0816***	0.0945***	0.0996***	0.0841***	0.0778***
	(0.011)	(0.0081)	(0.0077)	(0.0069)	(0.013)
Prop. Laborers	-0.148***	-0.143***	-0.158***	-0.172***	-0.184***
	(0.012)	(0.011)	(0.010)	(0.012)	(0.016)
Prop. Unemployed	-0.192***	-0.187***	-0.208***	-0.242***	-0.182***
	(0.032)	(0.017)	(0.027)	(0.020)	(0.043)
Head's Characteristics					
Age	0.0164***	0.0162***	0.0184***	0.0204***	0.0163***
	(0.0038)	(0.0019)	(0.0016)	(0.0026)	(0.0050)
Age Square	-0.0163***	-0.0156***	-0.0174***	-0.0193***	-0.0146**
	(0.0042)	(0.0022)	(0.0018)	(0.0032)	(0.0057)
Female	-0.0912***	-0.0896***	-0.0738***	-0.0801***	-0.0573**
	(0.025)	(0.025)	(0.014)	(0.021)	(0.025)
Married	0.0516*	0.0459***	0.0495***	0.0261	0.00218
	(0.028)	(0.017)	(0.016)	(0.025)	(0.031)
Divorce/Widow	-0.0382	-0.0242	-0.0285	-0.0162	-0.0205
	(0.042)	(0.026)	(0.025)	(0.030)	(0.044)
Education					
Prop. Informal Education	0.196***	0.200***	0.220***	0.214***	0.238***
	(0.022)	(0.012)	(0.010)	(0.017)	(0.033)
Prop. Primary School	0.343***	0.344***	0.365***	0.381***	0.422***
	(0.021)	(0.014)	(0.013)	(0.017)	(0.024)
Prop. High School	0.565***	0.602***	0.661***	0.704***	0.758***
	(0.024)	(0.017)	(0.018)	(0.019)	(0.028)
Prop. University Education	0.958***	1.072***	1.187***	1.335***	1.519***
	(0.019)	(0.020)	(0.020)	(0.032)	(0.031)
Prop. Technical Degree	0.190***	0.235***	0.253***	0.281***	0.305***
	(0.020)	(0.017)	(0.033)	(0.038)	(0.035)
Demographics					
Prop. Children (less 5 years)	-0.133***	-0.0732***	-0.0156	0.00982	0.0198
	(0.025)	(0.023)	(0.032)	(0.027)	(0.053)
Prop. Children (6-10 years)	-0.125***	-0.0638**	0.0116	0.0301	0.0981*
	(0.036)	(0.025)	(0.028)	(0.037)	(0.052)
Prop. Children (11-15 years)	-0.140***	-0.0941***	-0.0601*	-0.0500*	-0.0402
	(0.035)	(0.022)	(0.032)	(0.027)	(0.048)
Prop. Females (15-60 years)	0.000581	0.0323	0.0442**	0.0604**	0.0900**
	(0.020)	(0.021)	(0.018)	(0.025)	(0.039)
Prop. Old (above 60 years)	0.188***	0.196***	0.212***	0.336***	0.383***
	(0.067)	(0.041)	(0.060)	(0.082)	(0.11)
Household Characteristics					
Urban	0.232***	0.233***	0.258***	0.277***	0.281***
	(0.0078)	(0.0044)	(0.0065)	(0.0066)	(0.0100)
0.2< Land <0.4 Hectares	0.0415***	0.0341***	0.0288***	0.0230**	0.0327***

continued on next page...

Table 5.1: (*continued*)

Independent Var.	q10	q25	q50	q75	q90
	(0.0086)	(0.0059)	(0.0072)	(0.0091)	(0.013)
0.4< Land < 2 Hectares	0.0763***	0.0594***	0.0430***	0.0439***	0.0518**
	(0.015)	(0.011)	(0.013)	(0.017)	(0.021)
Land > 2 Hectares	0.127***	0.126***	0.148***	0.147***	0.173***
	(0.018)	(0.022)	(0.027)	(0.016)	(0.030)
Household Size	-0.118***	-0.140***	-0.162***	-0.184***	-0.206***
	(0.0045)	(0.0049)	(0.0048)	(0.0080)	(0.0086)
Householdsize Square	0.00447***	0.00578***	0.00686***	0.00838***	0.00985***
	(0.00029)	(0.00029)	(0.00032)	(0.00062)	(0.00064)
Region Controls					
North & East States					
Punjab	0.162***	0.109***	0.0714***	0.0571***	0.0433
	(0.013)	(0.021)	(0.015)	(0.022)	(0.037)
Delhi	0.184***	0.180***	0.135***	0.0970***	0.0604**
	(0.016)	(0.024)	(0.021)	(0.021)	(0.030)
Rajasthan	0.0802***	0.0535***	-0.00930	-0.0596***	-0.102***
	(0.019)	(0.012)	(0.015)	(0.012)	(0.036)
Uttar Pradesh	-0.0687***	-0.0729***	-0.103***	-0.130***	-0.149***
	(0.011)	(0.0096)	(0.0073)	(0.014)	(0.018)
Bihar	-0.171***	-0.197***	-0.257***	-0.281***	-0.330***
	(0.018)	(0.016)	(0.016)	(0.019)	(0.019)
Manipur	0.0381	-0.0538***	-0.126***	-0.195***	-0.265***
	(0.032)	(0.018)	(0.013)	(0.019)	(0.034)
Assam	-0.0702***	-0.0766***	-0.111***	-0.159***	-0.221***
	(0.025)	(0.019)	(0.014)	(0.012)	(0.021)
West Bengal	-0.0712***	-0.0617***	-0.106***	-0.132***	-0.160***
	(0.012)	(0.013)	(0.0079)	(0.0080)	(0.020)
Orissa	-0.310***	-0.328***	-0.324***	-0.343***	-0.352***
	(0.020)	(0.013)	(0.015)	(0.020)	(0.018)
Central & West & South States					
Chhattisgar	-0.163***	-0.202***	-0.254***	-0.231***	-0.243***
	(0.028)	(0.015)	(0.019)	(0.028)	(0.051)
Madhya Pradesh	-0.218***	-0.209***	-0.227***	-0.262***	-0.292***
	(0.023)	(0.019)	(0.012)	(0.018)	(0.028)
Gujrat	0.118***	0.124***	0.0822***	0.0212*	-0.0526***
	(0.022)	(0.017)	(0.011)	(0.013)	(0.014)
Maharastra	-0.0118	-0.0174	-0.0281**	-0.0335*	-0.0493**
	(0.015)	(0.013)	(0.012)	(0.020)	(0.022)
Karnataka	-0.0671***	-0.0749***	-0.117***	-0.130***	-0.150***
	(0.018)	(0.015)	(0.012)	(0.014)	(0.026)
Kerala	0.0381	0.0830***	0.0664***	0.0711***	0.0981***
	(0.026)	(0.019)	(0.016)	(0.018)	(0.032)
Tamil Nadu	-0.143***	-0.126***	-0.154***	-0.148***	-0.146***
	(0.014)	(0.017)	(0.012)	(0.011)	(0.020)
Constant	5.726***	5.963***	6.181***	6.443***	6.807***
	(0.069)	(0.030)	(0.038)	(0.041)	(0.094)
Observations	26485	26485	26485	26485	26485

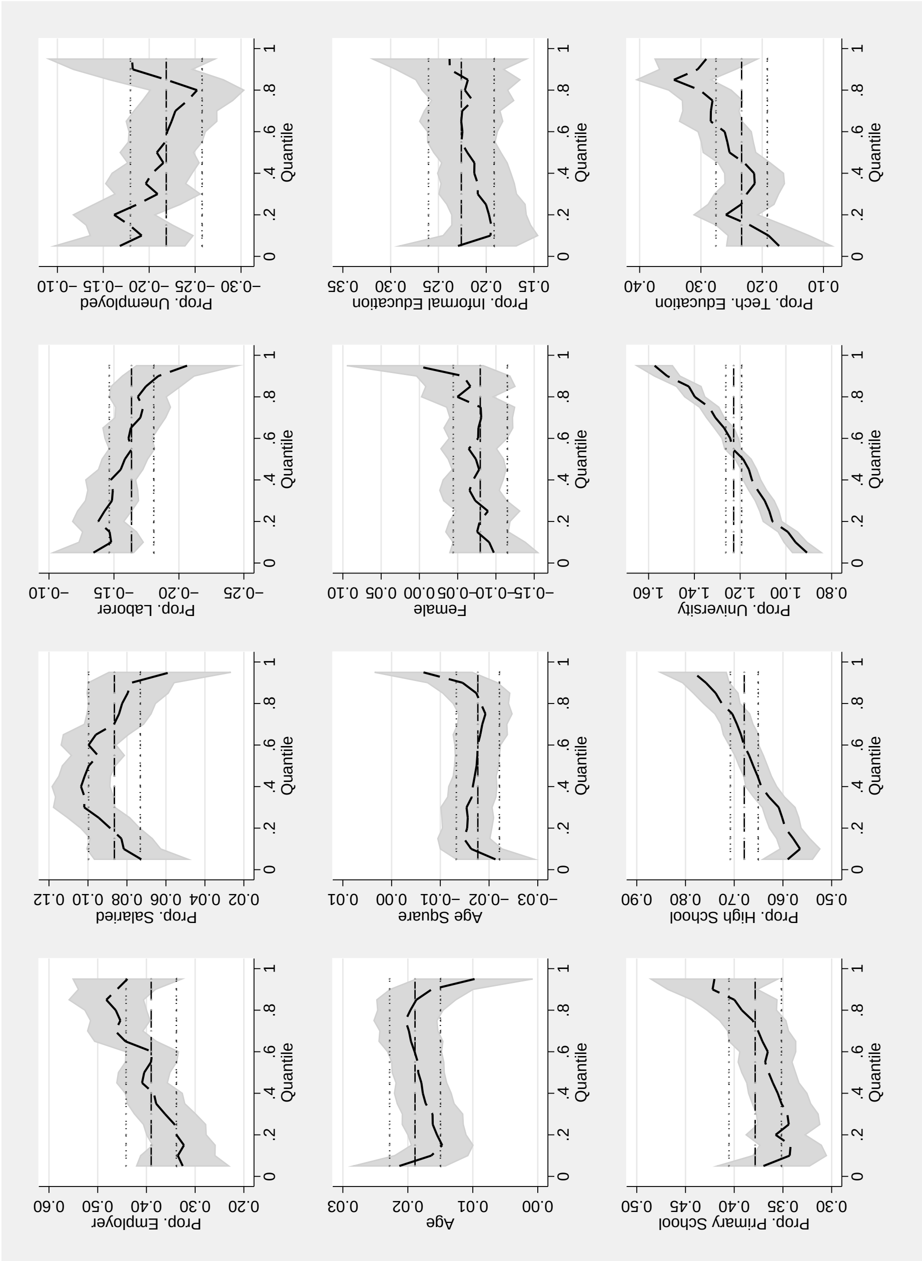


Figure 5.2: Quantile Plots-Households

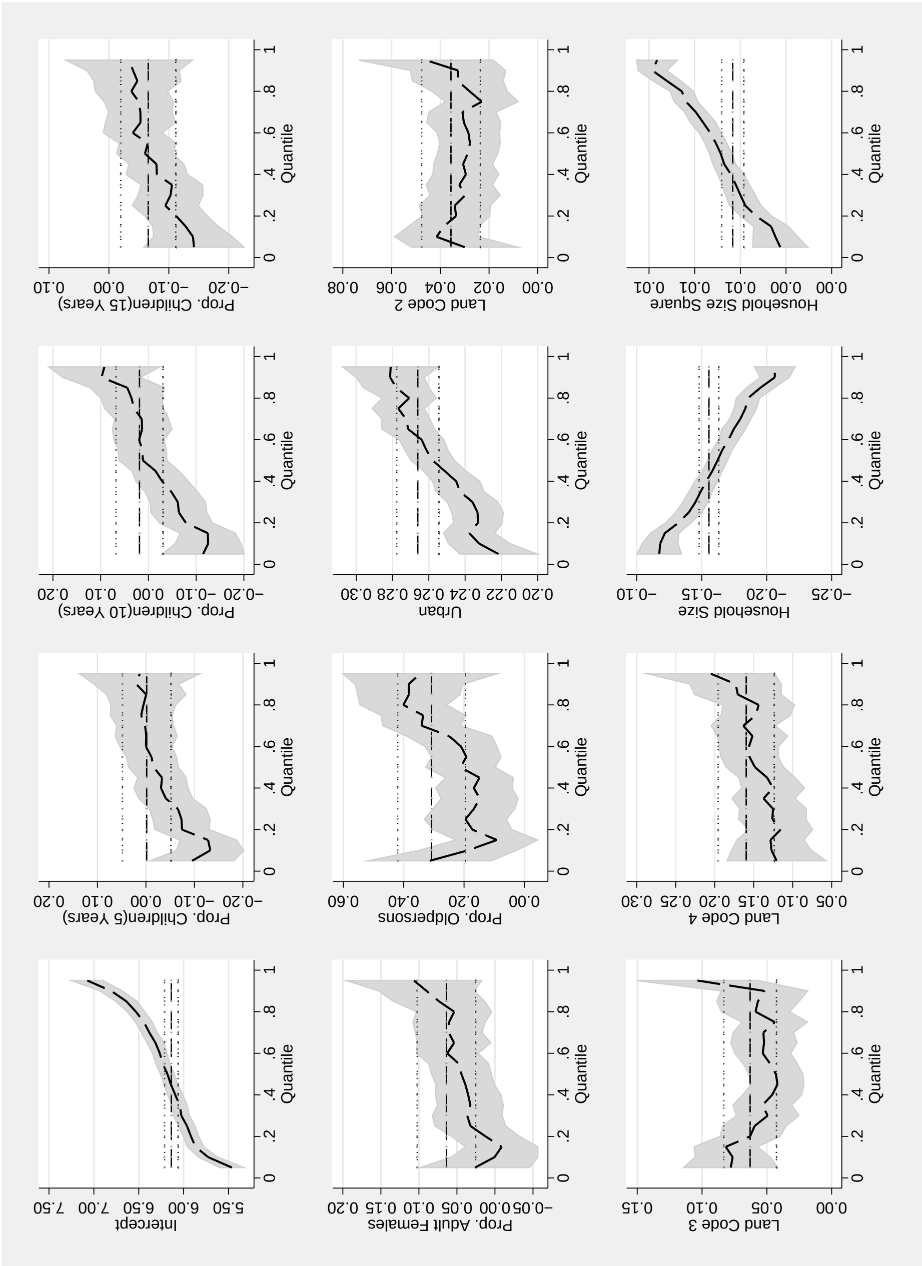


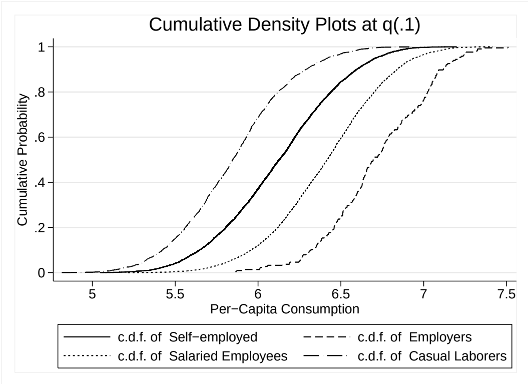
Figure 5.3: Quantile Plots-Households (continued)

Table 5.2: Household Heads, Occupation and Consumption

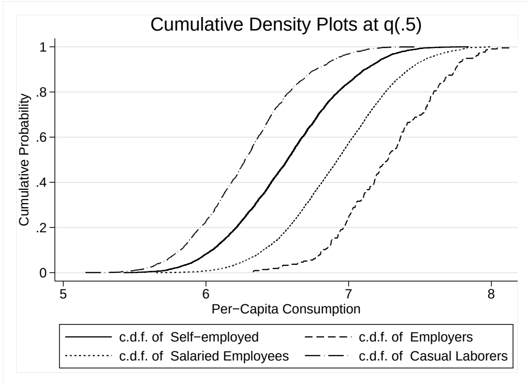
Estimates of Simultaneous Quantile Regression

Independent Var.	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0491***	-0.0579***	-0.0631***	-0.0564***	-0.0225
	(0.013)	(0.012)	(0.012)	(0.012)	(0.019)
Employer	0.224***	0.226***	0.258***	0.252***	0.306***
	(0.058)	(0.044)	(0.037)	(0.077)	(0.069)
Laborer	-0.228***	-0.229***	-0.246***	-0.225***	-0.203***
	(0.016)	(0.017)	(0.012)	(0.019)	(0.018)
Personal Characteristics					
Age	0.0340***	0.0324***	0.0395***	0.0405***	0.0282***
	(0.0047)	(0.0039)	(0.0039)	(0.0043)	(0.0066)
Age Square	-0.0371***	-0.0329***	-0.0409***	-0.0399***	-0.0240***
	(0.0061)	(0.0050)	(0.0048)	(0.0051)	(0.0083)
Female	-0.0144	-0.0296	-0.0653	0.0125	0.0811
	(0.035)	(0.031)	(0.043)	(0.041)	(0.060)
Married	-0.0301	-0.0312	-0.0321	-0.0658***	-0.0435
	(0.037)	(0.021)	(0.029)	(0.022)	(0.053)
Divorce/Widow	-0.212***	-0.233***	-0.176***	-0.220***	-0.184**
	(0.037)	(0.034)	(0.042)	(0.034)	(0.075)
General Education					
Informal Education	0.0479*	0.0390**	0.0219	0.0339*	0.0233
	(0.027)	(0.019)	(0.025)	(0.018)	(0.024)
Primary School	0.142***	0.146***	0.137***	0.172***	0.191***
	(0.018)	(0.013)	(0.018)	(0.018)	(0.016)
High School	0.235***	0.268***	0.292***	0.341***	0.361***
	(0.017)	(0.014)	(0.016)	(0.015)	(0.017)
University Education	0.413***	0.483***	0.559***	0.640***	0.732***
	(0.025)	(0.015)	(0.019)	(0.023)	(0.022)
Technical Degree or Diploma	0.170***	0.180***	0.169***	0.191***	0.235***
	(0.021)	(0.015)	(0.016)	(0.017)	(0.024)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Sector Controls	YES				
Constant	5.773***	6.081***	6.237***	6.478***	6.923***
	(0.085)	(0.071)	(0.072)	(0.068)	(0.12)
Observations	13692	13692	13692	13692	13692

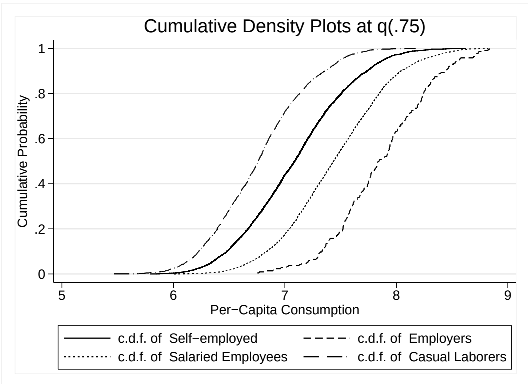
Notes: *Signifies p< 0.05; ** Signifies p<0.01;*** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for general/technical education is no general/technical education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.



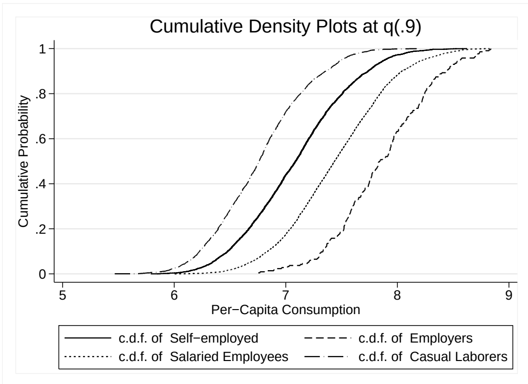
(a) Normalised Poverty Plots



(b) Normalised Poverty Plots



(c) Normalised Poverty Plots



(d) Normalised Poverty Plots

Figure 5.4: Occupation and Poverty Plots

(a)

(b)

Figure 5.5: Occupation and Inequality Plots at Median

Table 5.3: Occupation and Consumption in Rural Areas

Estimates of Simultaneous-Quantile Regression

	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0737***	-0.0885***	-0.122***	-0.125***	-0.0771**
	(0.022)	(0.022)	(0.019)	(0.026)	(0.031)
Employer	0.266***	0.152*	0.165*	0.185*	0.263
	(0.088)	(0.078)	(0.087)	(0.10)	(0.23)
Laborer	-0.207***	-0.232***	-0.242***	-0.250***	-0.224***
	(0.028)	(0.024)	(0.024)	(0.027)	(0.025)
Personal Characteristics					
Age	0.0219***	0.0286***	0.0367***	0.0397***	0.0415***
	(0.0059)	(0.0034)	(0.0057)	(0.0045)	(0.0092)
Age Square	-0.0258***	-0.0315***	-0.0399***	-0.0430***	-0.0442***
	(0.0069)	(0.0035)	(0.0067)	(0.0056)	(0.011)
Female	0.0635	0.0552	0.00795	-0.00698	0.0296
	(0.044)	(0.039)	(0.040)	(0.058)	(0.075)
Married	0.0426	-0.0211	-0.0747**	-0.0877*	-0.101
	(0.064)	(0.053)	(0.036)	(0.048)	(0.092)
Divorce/Widow	-0.164**	-0.244***	-0.242***	-0.220***	-0.254**
	(0.082)	(0.070)	(0.040)	(0.042)	(0.11)
General Education					
Informal Education	0.0503	0.0319	0.0242	0.0316	0.00941
	(0.032)	(0.029)	(0.020)	(0.027)	(0.031)
Primary School	0.153***	0.159***	0.136***	0.164***	0.158***
	(0.029)	(0.016)	(0.015)	(0.020)	(0.022)
High School	0.178***	0.199***	0.208***	0.255***	0.275***
	(0.027)	(0.023)	(0.020)	(0.030)	(0.034)
University Education	0.303***	0.313***	0.384***	0.449***	0.491***
	(0.028)	(0.015)	(0.027)	(0.029)	(0.064)
Technical Degree or Diploma	0.169***	0.204***	0.179***	0.180***	0.191***
	(0.045)	(0.028)	(0.024)	(0.041)	(0.062)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Constant	6.135***	6.289***	6.461***	6.671***	6.929***
	(0.11)	(0.10)	(0.093)	(0.097)	(0.15)
Observations	5202	5202	5202	5202	5202

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for education is no education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.

Table 5.4: Occupation and Consumption in Urban Areas

Estimates of Simultaneous-Quantile Regression

	q10	q25	q50	q75	q90
Occupation					
Self-employed	-0.0173	-0.0367**	-0.0370*	-0.0141	0.0244
	(0.025)	(0.016)	(0.020)	(0.015)	(0.023)
Employer	0.244***	0.223***	0.249***	0.269***	0.312***
	(0.055)	(0.059)	(0.045)	(0.078)	(0.060)
Laborer	-0.215***	-0.240***	-0.278***	-0.244***	-0.228***
	(0.034)	(0.025)	(0.015)	(0.024)	(0.028)
Personal Characteristics					
Age	0.0338***	0.0307***	0.0392***	0.0309***	0.0172**
	(0.0058)	(0.0054)	(0.0039)	(0.0058)	(0.0080)
Age Square	-0.0359***	-0.0297***	-0.0389***	-0.0274***	-0.0100
	(0.0076)	(0.0070)	(0.0043)	(0.0072)	(0.0098)
Female	-0.0805	-0.0726	-0.0653	0.0120	0.135
	(0.073)	(0.049)	(0.076)	(0.049)	(0.083)
Married	-0.0362	-0.0389	-0.0159	-0.0142	0.0168
	(0.042)	(0.032)	(0.033)	(0.032)	(0.047)
Divorce/Widow	-0.196***	-0.208***	-0.160**	-0.153**	-0.118
	(0.074)	(0.056)	(0.066)	(0.062)	(0.088)
General Education					
Informal Education	0.0529	0.0463**	0.0268	0.0532	0.0305
	(0.034)	(0.023)	(0.027)	(0.037)	(0.039)
Primary School	0.160***	0.159***	0.148***	0.187***	0.211***
	(0.029)	(0.025)	(0.022)	(0.022)	(0.028)
High School	0.286***	0.314***	0.349***	0.399***	0.405***
	(0.029)	(0.025)	(0.021)	(0.024)	(0.026)
University Education	0.465***	0.562***	0.633***	0.717***	0.812***
	(0.040)	(0.025)	(0.024)	(0.031)	(0.031)
Technical Degree or Diploma	0.175***	0.160***	0.152***	0.201***	0.215***
	(0.026)	(0.017)	(0.025)	(0.030)	(0.032)
Demographics	YES				
Household Characteristics	YES				
Region Controls	YES				
Constant	5.974***	6.337***	6.446***	6.826***	7.293***
	(0.13)	(0.092)	(0.073)	(0.11)	(0.15)
Observations	8490	8490	8490	8490	8490

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is log per-capita consumption expenditure. Base categories for occupation is salaried employee, for marital status is unmarried, for education is no education. Full set of state level regional dummies are also included in the regression with the excluded state being Andhra Pradesh.

Table 5.5: Occupational Selection and Consumption (*Self-employed*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0333***	0.0394***
		(0.0063)	(0.0060)
Age Square		-0.0420***	-0.0430***
		(0.0077)	(0.0069)
Female		-0.0645	0.0890
		(0.078)	(0.077)
Married		0.151***	0.0142
		(0.052)	(0.054)
Divorce/Widow		0.0854	-0.102
		(0.078)	(0.089)
Education			
Informal Education			0.110***
			(0.036)
Primary School			0.273***
			(0.049)
High School			0.419***
			(0.083)
University Education			0.633***
			(0.11)
Technical Degree or Diploma			0.0949**
			(0.044)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.214***	0.327***
		(0.018)	(0.029)
Land Variables		YES	YES
Household Size	-0.0956***	-0.140***	-0.141***
	(0.023)	(0.020)	(0.019)
Household Size Square	0.000948	0.00449***	0.00457***
	(0.0020)	(0.0017)	(0.0015)
Selection Coefficients			
m(Self-employed)	0.655***	0.703***	0.368***
	(0.087)	(0.079)	(0.12)
m(Employer)	-2.338***	-1.714***	-1.696***
	(0.40)	(0.33)	(0.30)
m(Salaried Employee)	0.697***	0.829***	1.017***
	(0.22)	(0.20)	(0.23)
m(Labor)	1.715***	1.559***	0.586**
	(0.20)	(0.18)	(0.27)
Constant	7.286***	6.492***	6.181***
	(0.12)	(0.15)	(0.17)
Observations	5047	5047	5047

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 5.6: Occupational Selection and Consumption (*Employers*)

Selection after Multinomial Logit			
Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0152	0.0142
		(0.036)	(0.037)
Age Square		-0.0254	-0.0148
		(0.039)	(0.040)
Female		-0.628**	-0.419
		(0.27)	(0.38)
Married		-0.0853	-0.437
		(0.35)	(0.43)
Divorce/Widow		-0.0214	-0.384
		(0.40)	(0.51)
Education			
Informal Education			-0.0861
			(0.23)
Primary School			0.157
			(0.23)
High School			0.552*
			(0.32)
University Education			0.883**
			(0.40)
Technical Degree or Diploma			0.333**
			(0.16)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.176	0.408***
		(0.13)	(0.15)
Land Variables		YES	YES
Household Size	-0.0477	-0.0865	-0.0504
	(0.12)	(0.13)	(0.13)
Household Size Square	-0.00906	-0.00448	-0.00814
	(0.011)	(0.013)	(0.014)
Selection Coefficients			
m(Self-employed)	1.917***	1.630**	-1.225
	(0.70)	(0.74)	(1.27)
m(Employer)	-0.168	-0.214*	-0.317**
	(0.11)	(0.11)	(0.13)
m(Salaried Employee)	1.448	1.203	0.332
	(0.98)	(1.12)	(1.31)
m(Labor)	1.623**	1.303	-1.078
	(0.80)	(0.86)	(1.26)
Constant	9.855***	9.380***	7.102***
	(0.82)	(1.12)	(1.37)
Observations	215	215	215

Notes: *Signifies $p < 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 5.7: Occupational Selection and Consumption (*Salaried*)

Selection after Multinomial Logit

Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0323***	0.0382***
		(0.0066)	(0.0064)
Age Square		-0.0365***	-0.0388***
		(0.0081)	(0.0080)
Female		-0.193**	-0.0298
		(0.087)	(0.21)
Married		0.113***	-0.0332
		(0.040)	(0.040)
Divorce/Widow		0.0544	-0.146**
		(0.071)	(0.066)
Education			
Informal Education			0.0795*
			(0.045)
Primary School			0.246***
			(0.049)
High School			0.375***
			(0.070)
University Education			0.595***
			(0.086)
Technical Degree or Diploma			0.184***
			(0.029)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.174***	0.273***
		(0.018)	(0.023)
Land Variabels		YES	YES
Household Size	-0.0841***	-0.165***	-0.158***
	(0.022)	(0.018)	(0.017)
Household Size Square	-0.00282	0.00397**	0.00348**
	(0.0020)	(0.0017)	(0.0015)
Selection Coefficients			
m(Self-employed)	1.844***	1.960***	0.614***
	(0.17)	(0.15)	(0.24)
m(Employer)	-0.930***	-0.311	-0.750***
	(0.31)	(0.26)	(0.23)
m(Salaried Employee)	0.748***	0.649***	0.414***
	(0.099)	(0.090)	(0.11)
m(Labor)	2.609***	2.165***	0.763***
	(0.18)	(0.16)	(0.24)
Constant	8.115***	7.378***	6.227***
	(0.058)	(0.13)	(0.19)
Observations	6391	6391	6391

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 5.8: Occupational Selection and Consumption (*Laborers*)

Selection after Multinomial Logit			
Independent Var.	Model I	Model II	Model III
Personal Characteristics			
Age		0.0157*	0.0210**
		(0.0090)	(0.0084)
Age Square		-0.0226*	-0.0225**
		(0.012)	(0.011)
Female		0.0532	0.167
		(0.094)	(0.11)
Married		0.0401	-0.0531
		(0.057)	(0.062)
Divorce/Widow		-0.106	-0.254**
		(0.086)	(0.11)
Education			
Informal Education			0.0928**
			(0.041)
Primary School			0.257***
			(0.053)
High School			0.329***
			(0.100)
University Education			0.451***
			(0.16)
Technical Degree or Diploma			0.127
			(0.080)
Demographics	YES	YES	YES
Household Characteristics			
Urban		0.113***	0.217***
		(0.022)	(0.034)
Land Variables		YES	YES
Household Size	-0.147***	-0.170***	-0.178***
	(0.027)	(0.032)	(0.031)
Household Size Square	0.00590**	0.00760***	0.00831***
	(0.0026)	(0.0028)	(0.0028)
Selection Coefficients			
m(Self-employed)	0.162	0.190	0.249
	(0.20)	(0.20)	(0.22)
m(Employer)	-3.378***	-3.075***	-2.065***
	(0.50)	(0.49)	(0.50)
m(Salaried Employee)	-0.116	0.153	1.021***
	(0.21)	(0.22)	(0.30)
m(Labor)	0.0897	0.133*	0.104
	(0.074)	(0.072)	(0.090)
Constant	6.962***	6.738***	6.860***
	(0.11)	(0.18)	(0.18)
Observations	2036	2036	2036

Notes: *Signifies $p < 0.05$; ** Signifies $p<0.01$;*** Signifies $p<0.001$. Dependent variable is log per-capita consumption expenditure. State dummies are included only in the selection equation.

Table 5.9: Occupational Selection

Base Multinomial Selection Equation

Independent Var.	Self-employed	Employer	Laborer
Personal Characteristics			
Age	-0.0294**	0.0536	-0.0497**
	(0.015)	(0.057)	(0.021)
Age Square	0.0225	-0.0300	-0.00269
	(0.018)	(0.064)	(0.026)
Female	-0.685***	-1.460**	-0.970***
	(0.14)	(0.68)	(0.19)
Married	0.807***	0.484	0.619***
	(0.10)	(0.43)	(0.13)
Divorce/Widow	1.042***	0.411	1.178***
	(0.17)	(0.73)	(0.22)
Education			
Informal Education	-0.345***	-0.488	-0.626***
	(0.10)	(0.58)	(0.11)
Primary School	-0.744***	-0.175	-1.541***
	(0.078)	(0.39)	(0.086)
High School	-1.301***	-0.238	-2.919***
	(0.079)	(0.39)	(0.10)
University Education	-1.772***	-0.351	-4.496***
	(0.086)	(0.40)	(0.20)
Technical Degree or Diploma	-0.535***	-0.339	-0.903***
	(0.083)	(0.24)	(0.22)
Household Characteristics			
Urban	-0.376***	0.486**	-0.874***
	(0.046)	(0.20)	(0.063)
$0.2 <$ Land < 0.4 Hectares	0.342***	0.619***	0.180***
	(0.046)	(0.18)	(0.067)
$0.4 <$ Land < 2 Hectares	0.0457	0.984***	-0.0912
	(0.087)	(0.31)	(0.13)
Land > 2 Hectares	-0.127	0.813	-0.986**
	(0.17)	(0.55)	(0.38)
Hindu	-0.255***	-0.368**	0.0340
	(0.049)	(0.16)	(0.073)
Backward	-0.181***	-0.740***	0.340***
	(0.044)	(0.17)	(0.069)
Region Variables	YES	YES	YES
Constant	1.147***	-6.354***	2.104***
	(0.28)	(1.18)	(0.38)
Observations	13700	13700	13700
LR $\chi^2(81)$	4193.61		
Log likelihood	-12593.207		
Pseudo R-squared	0.1427		

Notes: Marginal effects after multinomial logit estimation. Dependent variable is primary occupation of the household head. Set of state level regional dummies are also included in the regression.

Chapter 6

The Dynamics of Entrepreneurship

This chapter analyzes the spatio-temporal dynamics of entrepreneurship in India. The estimation results of the geoadditive models for the repeated cross sections of the years 2000 and 2004 suggest that while education predicts self-employment choice positively in 2000, it has a negative effect on self-employment choice in 2004. The spatial patterns are persistent across both the years and we do not observe much change in the four year period. Using three different cross-sectional databases collected over 1994-2004, we also construct pseudo panels of individuals based on 5 year bands of birth cohorts interacted with regions. The results of the pseudo panel analysis support the findings of the repeated cross section analysis. In particular, it is seen that education reduces the likelihood of self-employment in nonagriculture and increases it in agriculture.

6.1 Introduction

The aim of this chapter is to study the evolution of occupational choice over space and time in India. The years after liberalization in early nineties have unleashed many entrepreneurial opportunities. This has simultaneously resulted in a tremendous increase in employment opportunities for qualified and unqualified members of the work force. There are compelling reasons to hypothesize that the dynamic economic environment is influencing the occupational behavior of individuals. Under the assumption that entrepreneurship is riskier than providing professional services, Iyigun and Owen (1999) show that as an economy develops, individuals invest time in accumulating professional skills through education than accumulating entrepreneurial human capital. Lucas (1978) predicts that entrepreneurship decreases with economic development. Many empirical studies also find that the per-captia GNP is negatively related to the self-employment rates (Acs et al., 1994; Fölster, 2002). Acs et al. (1994) argue that self-employment increases at later stages of development when the importance of service sector increases. Furthermore, recent studies suggest a U-shaped relationship between economic development and entrepreneurship (Wennekers and Thurik, 1999; Wennekers et al., 2005).

However, the dynamics of entrepreneurship in developing countries has not been analyzed in the literature.[1] We test the claim of Iyigun and Owen (1999) that individuals invest in professional human capital in early stages of economic development and in entrepreneurial abilities in later stages of development, using three large databases of individuals in India, collected by the National Sample Survey Organization (NSSO) in 1994-1995, 1999-2000 and 2004. Thus, by analyzing the spatio-temporal dynamics of entrepreneurship in India, we make important contributions to the understanding of the nature of entrepreneurship in a developing economy.

The outline of the paper is as follows. In the next section, we briefly discuss the methodology employed, in particular geoadditive models used for repeated cross sections and the pseudo panel techniques. In the third section, the datasets used in the analysis are described. The fourth section presents the empirical findings relating to the determinants of self-employment choice over space and time. In particular, robustness of the insights that emerge from the repeated cross section study are tested using pseudo panel analysis. The final section summarizes the main results and discusses the limitations of the analysis.

6.2 Methodology

6.2.1 Repeated Cross Section Analysis

We replicate the analysis of chapter 2 for the year 2000.[2] Thus, semiparametric regression techniques based on Bayesian P-Splines and geoadditive models are used for comparing two cross-sections of NSSO data, in the analysis. The methodology is extensively discussed in section 2.3. In addition, we also employ the within estimator on a pseudo panel constructed from three sample surveys. The pseudo-panel method is discussed in subsection 6.2.2.

[1] There are a few exceptions. See, for instance, Mohapatra et al. (2007) for an analysis of the changing nature of self-employment in China.

[2] We analyze the 2004 employment-unemployment survey in chapter 2.

6.2.2 Pseudo Panel Approach

In the absence of genuine panel data, repeated cross-sectional data can be used to construct synthetic or pseudo panels, as suggested by the seminal paper of Deaton (1985). While cross-sectional data are collected over many years in developing countries, genuine panel data are very rare. A pseudo panel based on, for instance, age cohorts, gender, or education levels can be used to control for at least cohort fixed level effects. Such methods are similar to instrumental variable methods where group dummies are used as instruments.[3]

Consider the following linear model with individual effects,

$$y_{it} = x_{it}\beta + \alpha_i + e_{it}, \tag{6.1}$$

for i=1,. . . ,N and t=1,. . . ,T

For simplicity, we assume that observations on N individuals are available for all the time periods. When the individual fixed effects α_i are uncorrelated with x_{it}, it is possible to pool the cross sections to consistently estimate the regression coefficients β. In most situations, the correlation between the individual effects and some of the explanatory variables implies that the K moment condition given by $E\{(y_{it} - x_{it}\beta)x_{it}\} = 0$ is violated, in which case the cross sections are not poolable. In case the data is genuine panel data, the fixed effects approach can be used to treat α_is as unknown fixed parameters. However, if the data on the same individual are not available for each year, this cannot be used.

Following Deaton (1985), the observations are aggregated to cohort levels, where cohorts represent people of similar characteristics. In this case, the model assumes the following form,

$$\overline{y}_{ct} = \overline{x}_{ct}\beta + \overline{\alpha}_c + \overline{e}_{ct}, \tag{6.2}$$

for c=1. . . C and t=1. . . T,

where the variables are aggregated to cohort level averages. This pseudo panel, however, does not allow consistent estimation of β as $\overline{\alpha}_{ct}$ is likely to be correlated with the $\overline{x}_{ct}$. Under an assumption that $\overline{\alpha}_{ct}$ is a term fixed over time, the above equation can be consistently estimated. This is very likely to be the case when the average cohort size, $n_c \to \infty$. In such a case, the natural estimator for β is

[3] This section is based on Verbeek (2006).

the within estimator given by,

$$\hat{\beta}_W = (\sum_{c=1}^{C}\sum_{t=1}^{T}(\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)')^{-1}\sum_{c=1}^{C}\sum_{t=1}^{T}(\overline{x}_{ct} - \overline{x}_c)(\overline{y}_{ct} - \overline{y}_c) \tag{6.3}$$

As described in Verbeek (2006), the asymptotic behavior of pseudo panel estimators can be derived for the following alternative asymptotic sequences. First, when $N \to \infty$, with C fixed, so that $n_c \to \infty$. Second, when $N \to \infty$ and $C \to \infty$, with n_c fixed. Third, $T \to \infty$, with N, C and n_c fixed. While Moffitt (1993) and Verbeek and Vella (2005) employ the asymptotics of the first type, Deaton (1985), Verbeek and Nijman (1993) employ the second type.

In this paper, we also assume asymptotics of the first type. In this case, the fixed effects estimator is consistent estimator for β, when

$$plim\frac{1}{CT}\Sigma\Sigma(\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)' \tag{6.4}$$

is finite and invertible and

$$plim\frac{1}{CT}\Sigma\Sigma(\overline{x}_{ct} - \overline{x}_c)\alpha_{ct} = 0 \tag{6.5}$$

As $n_c \to \infty$ the above conditions are automatically satisfied as the cohort fixed effects converge to a constant over time, that is, $\alpha_{ct} \to \alpha_c$ (Moffitt, 1993).[4] Deaton (1985) relies on asymptotics of the first type and does away with the necessity to have large numbers of observations in each cohort. This is achieved by considering the cohort averages as error-ridden measurements of the population averages of the cohorts. By assuming that measurement errors are distributed with zero mean, the moment matrices of the within estimator are adjusted to correct for the measurement error. McKenzie (2004) shows that when cohorts are based on age groups, the asymptotics of the second type seldom get satisfied, as the number of cohorts is fixed. For this reason, as the cohort sizes in the sample are very large, we assume asymptotics of the first type, to consistently estimate β.

[4] Verbeek and Nijman (1992) show that even when the cohort sizes are large, the bias may be present.

6.3 Data

The data used for the repeated cross-sectional analysis are the 55th round and 60th round employment-unemployment surveys of the National Sample Survey Organization (NSSO) of India. More than 500,000 thousand individuals in 120,000 thousand households were questioned about their economic status and personal background in 2000.[5] The households were selected based on a stratified sampling methodology. Since the focus of this paper is on economically active individuals, we use the same selection criteria as in chapter 2 to produce a comparable sample. Only those individuals who have reported being self-employed (including own account workers and employers), salaried employees, casual laborers, or unemployed are included in the sample. For similar reasons, we restrict the sample to those who are older than 15 years but younger than 70 years. We thus exclude from our analysis family members who assist household enterprises, children and the elderly, and people classified into other miscellaneous occupational categories. The final sample for the year 2000 consists of 169,147 individuals.

For the pseudo panel analysis, the 50th round of the employment-unemployment survey of the NSSO conducted in 1994 is also considered. It is pertinent to mention that the number of states and districts has increased in India in the decade following the collection of this survey. Hence, we harmonize the datasets based on the geographic boundaries in the year 1995 for the pseudo panel analysis. We construct cohorts of men based on age and their state of residence for this purpose. We do not consider women for two reasons. First, only 10%-14% of women in the datasets are economically active. Many female cohorts based on the states and 5 year bands have very few observations. Hence, we only analyze the pseudo panel of male cohorts generated in each region in 5 year bands. We only consider those cohorts that have at least 500 observations in each of the surveys, for the asymptotic reasons described in the previous section.

[5] For a description of the 2004 sample, please see chapter 2.

6.4 Empirical Analysis

6.4.1 Repeated Cross Sections

We estimate the following geoadditive model for the year 2000 and compare the results with the estimation results in chapter 2.

$\eta = \gamma_{const} + \gamma_{female} + \gamma_{marital\_status} + \gamma_{education\_general} + \gamma_{education\_technical} + \gamma_{wealth} + \gamma_{urban} + \gamma_{agri} + \gamma_{hindu} + \gamma_{backward} + f_{age} + f_{spatial}(district) + f_{random}(district)$

As described in chapter 2, this methodology allows for the estimation of non-linear effects of the continuous variables and the neighborhood effects of spatial units on the probability of individuals selecting self-employment. We also estimate sector specific models, separately for agriculture and nonagriculture.

$\eta = \gamma_{const} + \gamma_{female} + \gamma_{marital\_status} + \gamma_{education\_general} + \gamma_{education\_technical} + \gamma_{wealth} + \gamma_{urban} + \gamma_{hindu} + \gamma_{backward} + f_{age} + f_{spatial}(district) + f_{random}(district)$

The evidence from the estimation of the empirical models for the year 2000 show consistency in the determinants of self-employment. Age shows a similar non-linear effect in Figure 6.1(a), with probability of self-employment increasing at decreasing rates until the age of 55 and remarkably increasing after 55.[6] The effect is mostly attributable to such a phenomena in nonagriculture, where retirement effects may be playing an important role. As Figure 6.1(e) suggests the age effect in agriculture is close to being linear.

Table 6.3 presents the estimates of the fixed effects (those variables that entered parametric part of the model) in the first estimation when all individuals are included in the analysis. The estimation results suggest that the effects are very similar in two years. The only startling difference is in the effect of university education on self-employment. While the effect is positive in 2000, it is significant and negative in 2004. Furthermore, high school education also predicts the choice of self-employment in 2000 more strongly than in 2004. Hindus and members of backward classes have a lower likelihood to be self-employed in year 2000 as well.

[6]As in chapter 2, the non-linear effect of age is modeled as third degree P-Spline with second order random walk penalty. The number of equidistant knots is assumed to be 20. The structured spatial effects are estimated based on Markov random field priors and random spatial effects are estimated with gaussian priors. The variance component in all the cases are estimated based on inverse gamma priors with hyperparameters a=0.001 and b=0.001. The number of iterations is set to 110000 with burnin parameter set to 10000 and the thinning parameter set to 100. The autocorrelation files and the sampling paths show that the MCMC algorithm has converged. These plots are available from the author.

Table 6.4 shows the estimates for nonagriculture. It is seen that though education reduces the probability of self-employment in the year 2004, it has a significantly positive effect in the year 2000. This finding supports the arguments of Lucas (1978) and Iyigun and Owen (1999) that people move from self-employment to paid employment as economy develops. However, Table 6.5 suggests that in the agricultural sector, education increases the probability of self-employment in both years. Thus, a startling result of the analysis is that while educated people were more likely to be self-employed in nonagriculture in the year 2000, they were less likely to be self-employed in 2004. The estimates of other variables such as gender and marital status are remarkably similar.

Furthermore, while the coefficient of urban location variable is negative in year 2000, it is positive in 2004. This suggests that individuals located in urban areas are less likely to be self-employed in the year 2000 and more likely to be self-employed in the year 2004, conditional on other factors. This could be a result of globalization affecting urban areas in India more directly than rural areas, leading to more self-employment opportunities for individuals living in urban areas. It is also possible that increased migration from rural areas into urban areas is compelling people to enter entrepreneurship in urban regions. However, these speculations need to be empirically validated.

As the spatial maps in the plots in Figure 6.3 suggest, the spatial dynamics are predominantly observed in the North and South, with the structured spatial effects in some districts of Rajasthan becoming insignificant and those in some districts of Bihar becoming positive and significant. In the southern regions, it is seen that some more districts of Karnataka have become black over the four year period, suggesting that the probability to self-employment in these regions has decreased over time. Figure 6.4 suggests that in nonagriculture, spatial effects have remained stable, although spatial effects in some districts of Andhra Pradesh have become positive in 2004. The spatial patterns in Figure 6.5 show that agricultural self-employment has decreased in the southern regions and even in the northern parts of Rajasthan and Uttar Pradesh.

In Table 6.6, the determinants of the spatial variation are estimated using the set of regional indicators as in subsection 2.3.3.[7] The estimation results are very similar to the results in Table 2.7. Per-capita net state domestic product (NSDP) is significantly negative throughout. The main inferences about the credit-deposit variable, the unemployment variable and the demography variables are qualitatively same as in section 2.5. While population density increases self-employment in nonagriculture, it decreases the self-employment in agriculture. This also holds for the unemployment and credit-deposit ratio variables. However, a startling difference is observed in the coefficient of the per-capita credit variable, that is significantly negative in Table 6.6 but insignificant in Table 2.7. One reason could be lagged effect of credit. As the spatial maps suggest that people have relatively higher propensity to be self-employed in districts of Uttar Pradesh and Bihar, two regions that have lowest levels of financial development, the estimated inverse relationship is plausible. In Table 6.7 and Table 6.8, the multinomial logit models are estimated for the 95% spatial effects. The results are consistent with the analysis of the mean spatial effects. As in subsection 2.5.3, however, the credit-deposit variable does not have a positive effect on self employment in 2000 as well.

6.4.2 Pseudo Panel Analysis

The estimation results of pseudo panel regression are presented in Table 6.9. The standard OLS model is biased as the F test in all the fixed effects regressions is significant, suggesting that cohort fixed effects are correlated with the exogenous variables. The cohorts are constructed on five year bands from 15 years to 70 years. These five year bands are interacted with the state regions to define cohorts. In the first column, estimation is based on cohorts of all men in the sample.[8] For the estimation in the second column, we construct a psuedo panel of men working in nonagriculture, to analyze more homogenous cohorts. In the third estimation, we similarly construct a psuedo panel of men working in agriculture.

[7]See Table 2.7 for the estimation of the determinants of spatial effects in the year 2004. The same variables are used as determinants of regional patterns here, except for the per-capita net state domestic product(NSDP) variable. In table Table 2.7, the per-capita NSDP in 2003-2004 is used and in Table 6.6, the per-captia NSDP in 1997-1998 is used. The other variables are assumed to be stable over time over different spatial units. However, this could cause some bias in the results and they should be interpreted keeping this in view.

[8]There are 492 cohorts that have at least 500 individuals in each of the cross sections when individuals of agricultural as well as nonagricultural sectors are considered in the construction of the pseudo panel.

We introduce the average age of the cohort, proportion of married individuals, proportion of people in urban regions, along with a series of variables that indicate the proportion of individuals in each education category. We also introduce the religion and caste variables, as these have been found to play an important role in previous chapters.

The estimation results in the first column suggest that 1% increase in the share of people with informal education leads to an increase in the transition rate into self-employment by 0.24%. The higher education variables and technical education variables are insignificant. This suggests that people entering self-employment over the period 1995-2005 are mostly individuals with informal education. Surprisingly, however, the analysis suggests that individuals with primary education are less likely to transition into self-employment, in contrast to the evidence from the cross-sectional studies. This could be a result of the high heterogeneity within cohorts, as individuals of both agricultural and nonagricultural sectors are considered in the pseudo-panel construction.

The 'Urban' variable is also insignificant, suggesting that a rise in the urban share of the population in a cohort has no influence on the proportion of self-employed people. The variable measuring the proportion of people in agriculture is positive and strongly significant, and suggests that a 1% increase in the share of people in agriculture increases the self-employment transition rate by 0.49%. Furthermore, a 1% increase in the proportion of people belonging to the scheduled castes and scheduled tribes decreases the self-employment transition rate by 0.3%. The variable 'Hindu', however, is insignificant and suggests that the relationship remained stable over time.

In the second column, we analyze the pseudo panel of cohorts of individuals in nonagriculture. The effects of the age and gender are consistent with the theoretical predictions. However, we see that education of all types, other than informal education, reduces transition into self-employment in the nonagricultural sector. A 1% increase in the share of individuals with university education, for instance, is found to decrease the self-employment transition rate by 0.3%. The effect of having a 'technical diploma' is also negative and significant. The coefficient of the backward caste variable suggests that belonging to such castes has a negative influence in nonagriculture. In particular, it is seen that a 1% rise in the proportion of individuals belonging to the scheduled castes/scheduled tribe groups in a cohort, reduces the self-employment transition rate by 0.55%.

In the third column, the analysis is done on a pseudo panel of cohorts in agriculture. In contrast to results of the nonagriculture estimation, it is seen that education has a significant positive effect on the transition into self-employment in agriculture. This is consistent with the results of the repeated cross-sectional analysis in Table 6.5, where coefficients of the education variables are higher in the year 2004. The effect of education on self-employment is highest for those with informal education. This effect keeps decreasing as education rises but remains positive and significant. However, the coefficient of "University" is almost half the coefficient of "Informal" education, suggesting that education has a positive effect that is non-linear and decreasing. Thus, in the agricultural sector, while a 1% increase in the proportion of people with informal education increases the self-employment transition rate by 0.42%, a similar increase in the proportion of people with university education increases the self-employment rate by only 0.23%. This suggests that educated individuals who stay in agriculture choose self-employment over paid employment.

In Table 6.10, the land variables are also introduced in the estimation. While the coefficients of nonagriculture estimation in the second column are similar to the coefficients of the estimation without land variables, the coefficients of education variables in the agriculture equation in the third column have shifted downward. This suggests that the education variables captured the positive effect of the land variables in the earlier estimation. While in nonagriculture equation, land is positive and significant in the lowest category, in the agriculture equation, the highest land variables are positive and significant. Thus, while small amounts of land enable individuals to procure capital for entering self-employment in nonagriculture, individuals with large amounts of land choose self-employment in agriculture. Moreover, the urban variable remains insignificant in both estimations. This result is unexpected as one would predict an increase in the share of urban population in the cohort to have a positive influence on the self-employment transition rate. It is possible that rural-urban migration increases the share of people working as self-employed and share of people working as paid employees proportionately, therefore leading to an insignificant effect on movement into self-employment.

Finally, we construct a pseudo panel of cohorts of individuals working in nonagriculture and living in rural areas as well as a pseudo panel of cohorts of individuals working in nonagriculture and living in urban areas. Table 6.11 presents the estimation results. In the first column, we consider cohorts of rural men working in nonagriculture. Once again, it is seen that education has a significant negative effect, which is more pronounced at the university level. The land variables are insignificant, suggesting that in rural areas, possession of land does not lead people into self-employment. One reason could be that in rural areas, the level of financial development is comparatively low and price of the land is cheaper. This can reinforce credit constraints for such individuals.

The results of the estimation on cohorts of urban males is presented in the second column. The results suggest that the negative effect of education on self-employment choice is also present in the urban cohorts. However, the coefficients are much smaller than coefficients of the rural estimation. In particular, the results suggest that while a 1% increase in the proportion of university educated individuals in rural areas reduces the transition into self-employment by 0.61%, an increase in the share of university educated individuals in urban areas decreases the transition rate into self-employment by 0.24%.[9] Furthermore, the informal education variable is positive and significant in the urban equation, suggesting that, ceteris paribus, an increase in the share of people with informal education by 1% in the cohort, increases the self-employment rate by 0.65%. The lowest level land variable is positive and significant, suggesting that the overall effect of land (seen in the earlier estimation with all the workers in nonagriculture) is primarily due to such an effect in urban areas. The coefficient of the scheduled castes/scheduled tribes (SC/ST) variable is also significantly negative, and the effect is higher in urban areas.[10] Thus, it is seen that a 1% increase in the proportion of the SC/ST people in a cohort reduces the transition rate into self-employment by 0.51% in rural areas and by 0.7% in urban areas. This could be a result of government's reservation policy that sets aside public sector jobs for individuals from these castes. The concentration of such jobs in urban areas could explain this result to some extent.

[9] There are many reasons why this might be the case. If the credit constraints are relatively more severe for education in rural areas than in urban areas, this result is plausible. Moreover, it should also be noted that the estimations are based on two separate pseudo panels constructed for the rural and urban cohorts. Thus the coefficients are not strictly comparable.

[10] This is also consistent with the findings in chapter 2 (Table 2.10).

Table 6.1: Results for Agriculture

	2000	2004	Pseudo
Informal Education	+	+	+
Primary Education	+	+	+
High School	+	+	+
University	+	+	+
Technical Education	*insig.*	*insig.*	*insig.*
Technical Diploma	*insig.*	*insig.*	*insig.*
Urban	+	+	*insig.*
Hindu	*insig.*	*insig.*	*insig.*
Backward	-	-	*insig.*

Table 6.2: Results for Non-agriculture

	2000	2004	Pseudo	Pseudo Rural	Pseudo Urban
Informal Education	+	+	+	*insig.*	+
Primary Education	+	+	-	-	-
High School	+	-	-	-	-
University	+	-	-	-	-
Technical Education	-	-	*insig.*	*insig.*	*insig.*
Technical Diploma	-	-	-	*insig.*	-
Urban	-	+	*insig.*		
Hindu	-	-	*insig.*	*insig.*	-
Backward	-	-	-	-	-

6.4.3 Reconciling the Results

Table 6.1 summarizes the results of the repeated cross section analysis and the pseudo panel analysis for the agricultural sector. The first column summarizes the results for the year 2000, the second column for the year 2004 and the third column, results of the pseudo panel analysis.

The coefficient of informal education is positive in all three columns. The pseudo panel estimation supports the finding of the repeated cross sectional analysis in Table 6.5, that the coefficient has increased, suggesting that informal education increased the probability of people choosing self-employment over time. This holds for the other educational variables as well. The technical education variables are, however, insignificant in all three columns, suggesting that these factors do not influence self-employment in agricultural sector. However, the backward class variable is negative and significant in both cross sections but is insignificant in the pseudo panel analysis, suggesting that the relationship with self-employment has remained somewhat constant over time. These results should be compared with care, as the pseudo panel analysis includes data from three cross sections and captures the dynamics of self-employment over a longer period of time. However, the results are broadly consistent with the repeated cross sectional analysis of the years 2000 and 2004.

Table 6.2 summarizes the main results for nonagriculture. The first two columns summarize the results of the repeated cross-sectional analysis and the next three columns summarize the results of the pseudo panel analysis.

The coefficient of informal education is positive in the first three columns, suggesting that informal education is associated with an increase in the probability of self-employment in nonagriculture as well. This is supported by the results in Table 6.4 which show that the coefficient of informal education is higher in the year 2004. The coefficient for rural areas is insignificant in the fourth column and is positive for urban areas in the fifth column. This suggests that the positive effect is attributable to the role of informal education in urban areas, as a determinant of self-employment.

Though the coefficient of primary schooling is positive in the first two columns, the estimated effect is negative in the pseudo panel analysis. This suggests that individuals with primary education have become less likely to be self-employed over time. This result is corroborated by the lower coefficient of the primary education variable in the year 2004, relative to the coefficient for the year 2000, in the repeated cross-sectional analysis (in Table 6.4). The pseudo panel analysis further suggests that the negative effect over time is seen in rural as well as urban areas.

In contrast to the variables of lower education, the variables of higher education switch signs over the years 2000 and 2004. Both high school and university education reduced the likelihood of individuals choosing self-employment over time, as the negative sign seen in the third column (pseudo panel analysis) suggests. This result is consistent with the repeated cross-section analysis, which shows that coefficients of higher education variables have decreased over time.

This negative effect is observed in rural as well as urban areas.

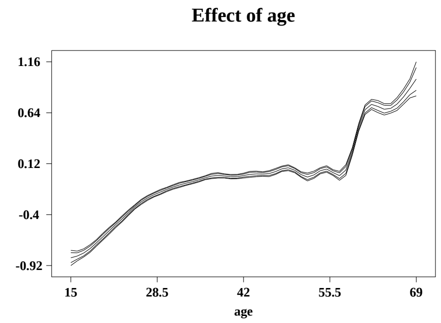
The effect of technical education is negative in the year 2000 and as well as in the year 2004. The pseudo panel suggests that the nature of the relationship has not changed over time, at least in the case of technical degree. For the technical diploma variable, the effect is negative in all columns except in the rural pseudo panel. This suggests that in urban areas, individuals with technical diplomas have become less likely over time to be self-employed. The results of the repeated cross-sectional analysis in Table 6.4 show that the absolute value of the coefficient of technical diploma variable is smaller in 2004. This is captured by the negative sign of the coefficient estimated by the pseudo panel analysis.

The effect of urban location also switched signs, suggesting that while individuals located in urban areas were less likely to be self-employed in the year 2000, they were more likely to be self-employed in the year 2004, conditional on other factors. However, the pseudo panel analysis does not capture this change as the value in the third column suggests that the relationship has remained stable. This could be an artifact of analyzing this relationship over a longer period of time in the pseudo panel analysis than the cross-sectional analysis. With urban areas experiencing the effects of globalization directly, it is plausible that this relationship is unstable over time.

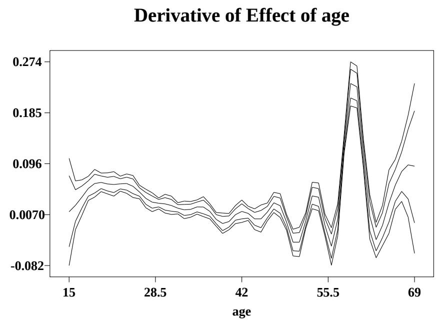
The signs of the cultural context variables, Hindu and SC/ST suggest that the relationship has remained stable over time. While Hindus were less likely to be self-employed in the years 2000 and 2004, the insignificant coefficient in the pseudo panel analysis suggests that this relationship has remained stable over time. This is corroborated by the repeated cross-section analysis in Table 6.4, where the coefficient of the Hindu variable is almost equal in both years. However, the SC/ST variable is negative throughout, indicating that individuals in these castes have become less likely over time to be self-employed. This result is supported by an increase in the absolute value of the coefficients of this variable in the repeated cross sectional analysis in Table 6.4.

6.5 Conclusion

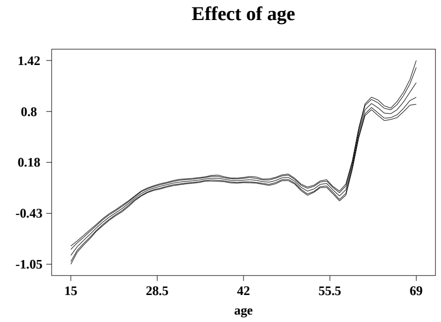
This chapter uses two different empirical methods to analyze the spatio-temporal dynamics of entrepreneurship in India. The results of the repeated cross section analysis for the years 2000 and 2004 suggest that while education predicts self-employment choice positively in 2000, it has a negative effect on self-employment choice in 2004. The spatial dynamics are almost non-existent and we do not observe much change in self employment patterns in the four year period of analysis. Using three different cross-sectional databases collected over 1994-2004, we also construct pseudo panels of individuals based on 5 year bands of birth cohorts, regions and sectors. Thus, the pseudo panel analysis tracks the dynamics of entrepreneurship over a longer period of time. The results also suggest that higher education reduced the self-employment rate in the nonagricultural sector but increased the self-employment rate in the agricultural sector. The results of the repeated cross-sectional analysis and the pseudo panel analysis are consistent with each other. The analysis also suggests that while individuals with lower levels of wealth (land) have moved into self-employment in nonagriculture, individuals with higher levels of wealth (land) have entered self-employment in agriculture.



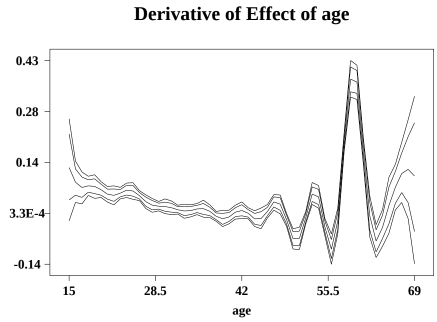
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in the Aggregate Model.



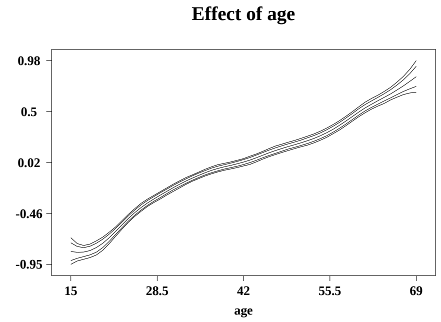
(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in the Aggregate Model.



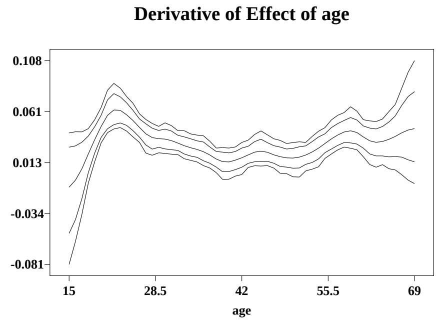
(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture.



(d) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Nonagriculture.

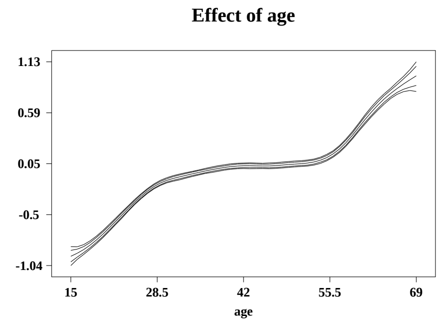


(e) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Agriculture.

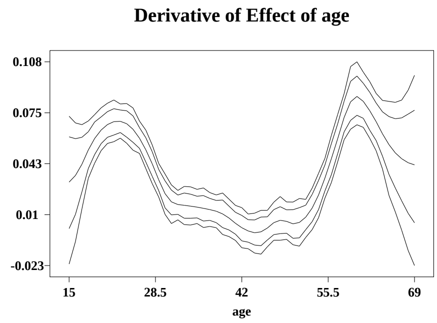


(f) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Agriculture.

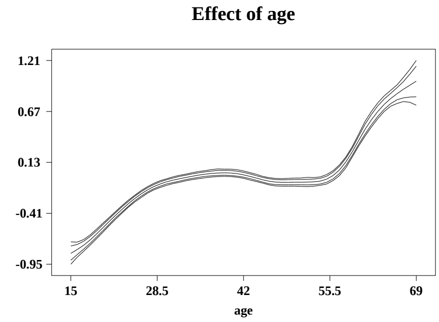
Figure 6.1: Non-linear Effect of Age on Self-employment (2000)



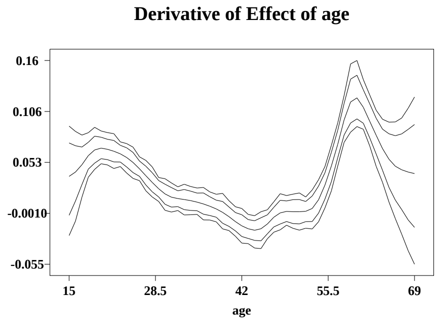
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in the Aggregate Model.



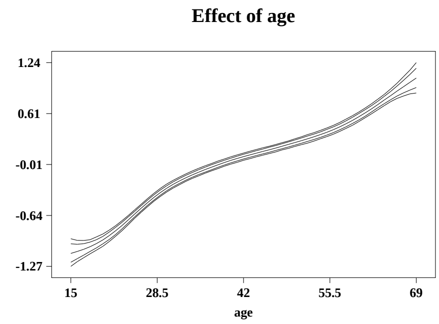
(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in the Aggregate Model.



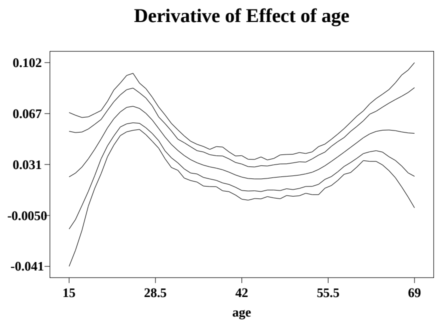
(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture.



(d) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Nonagriculture.



(e) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Agriculture.



(f) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Agriculture.

Figure 6.2: Non-linear Effect of Age on Self-employment (2004)

Table 6.3: Determinants of Self-employment

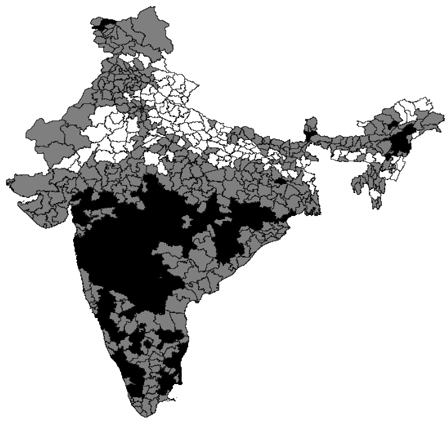
	2000		2004	
Variable	Coefficient	Std. Dev.	Coefficient	Std. Dev.
Personal Characteristics				
Female	-0.369	0.010	-0.398	0.014
Married	0.214	0.013	0.175	0.018
Divorced	0.295	0.021	0.317	0.029
General Education				
Informal	0.222	0.054	0.265	0.019
Primary School	0.307	0.012	0.332	0.014
High School	0.386	0.009	0.193	0.016
University	0.180	0.010	-0.181	0.020
Technical Education				
Technical Degree	-0.215	0.041	-0.127	0.057
Technical Diploma	-0.392	0.020	-0.117	0.026
Land Possessed				
0.2< Land <0.4 Hectares	0.143	0.012	0.149	0.014
0.4< Land <2 Hectares	0.919	0.015	0.791	0.017
Land >2 Hectares	1.438	0.017	1.180	0.024
Location				
Urban	0.197	0.009	0.253	0.013
Agriculture	0.113	0.009	0.336	0.013
Religion Social Group				
Hindu	-0.197	0.010	-0.205	0.014
Backward	-0.180	0.008	-0.183	0.012
Constant	-0.638	0.021	-0.545	0.027
N	169147		86140	
Deviance(Mean)	182825.52		93422.587	
Std. Dev.	34.42		36.196992	
deviance($\bar{\mu}$)	182367.35		92973.92	
pD	458.16973		448.66642	
DIC	183283.69		93871.253	

Table 6.4: Determinants of Self-employment (Nonagriculture)

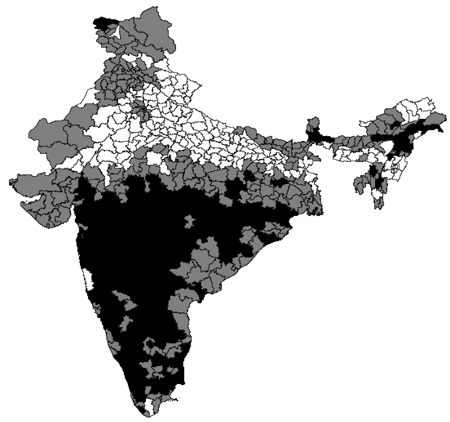
	2000		2004	
Variable	Coefficient	Std. Dev.	Coefficient	Std. Dev.
Personal Characteristics				
Female	-0.196	0.014	-0.256	0.018
Married	0.257	0.015	0.203	0.019
Divorced	0.256	0.027	0.218	0.042
General Education				
Informal	0.209	0.078	0.141	0.028
Primary School	0.259	0.016	0.130	0.021
High School	0.244	0.011	-0.039	0.022
University	0.075	0.011	-0.349	0.024
Technical Education				
Technical Degree	-0.213	0.043	-0.109	0.057
Technical Diploma	-0.370	0.019	-0.134	0.025
Land Possessed				
0.2< Land <0.4 Hectares	0.145	0.013	0.151	0.015
0.4< Land <2 Hectares	0.100	0.019	0.112	0.022
Land >2 Hectares	0.131	0.025	0.160	0.033
Location				
Urban	-0.087	0.011	0.029	0.015
Religion Social Group				
Hindu	-0.182	0.011	-0.180	0.016
Backward	-0.101	0.010	-0.150	0.014
Constant	-0.350	0.024	-0.222	0.031
N	97153		51674	
Deviance(Mean)	113415.26		60166.724	
Std. Dev:	32.91		34.978124	
deviance($\bar{\mu}$)	113011.95		59807.524	
pD	403.31415		359.20045	
DIC	113818.58		60525.925	

Table 6.5: Determinants of Selfemployment (Agriculture)

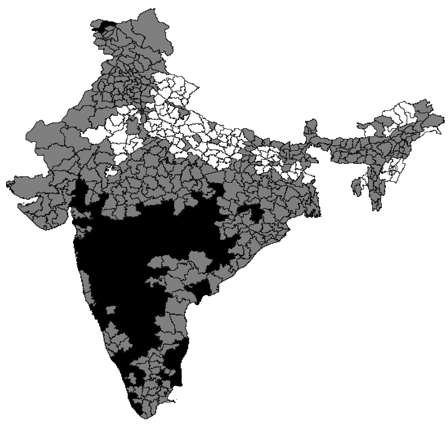
	2000		2004	
Variable	Coefficient	Std. Dev.	Coefficient	Std. Dev.
Personal Characteristics				
Female	-0.495	0.018	-0.540	0.027
Married	0.175	0.029	0.206	0.042
Divorced	0.282	0.040	0.447	0.058
General Education				
Informal	0.156	0.088	0.233	0.032
Primary School	0.244	0.021	0.435	0.025
High School	0.496	0.018	0.758	0.035
University	0.742	0.027	0.862	0.076
Technical Education				
Technical Degree	-0.202	0.204	0.157	0.274
Technical Diploma	0.034	0.102	0.193	0.114
Land Possessed				
0.2< Land <0.4 Hectares	0.543	0.041	0.533	0.042
0.4< Land <2 Hectares	2.070	0.042	1.986	0.042
Land >2 Hectares	3.158	0.047	2.787	0.050
Location				
Urban	0.371	0.030	0.459	0.044
Religion Social Group				
Hindu	-0.032	0.022	-0.015	0.035
Backward	-0.358	0.018	-0.286	0.027
Constant	-1.243	0.053	-1.031	0.064
N	71994		34466	
Deviance(Mean)	47457.96		22493.237	
Std. Dev:	32.90		35.860231	
deviance($\bar{\mu}$)	47020.57		22042.36	
pD	437.39442		450.87693	
DIC	47895.359		22944.114	



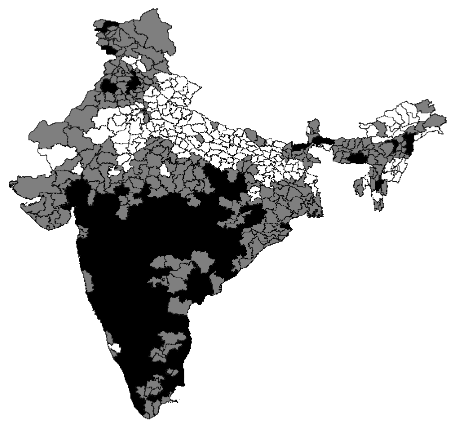
(a) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(b) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

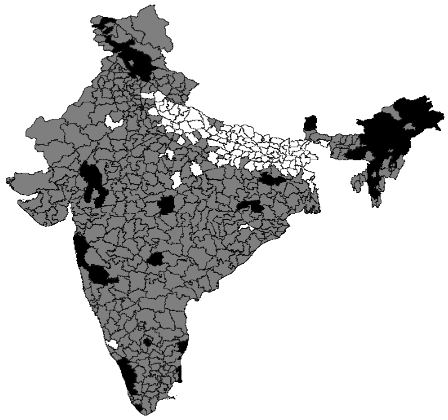


(c) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

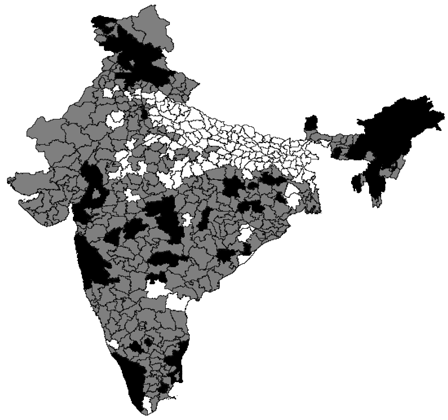


(d) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

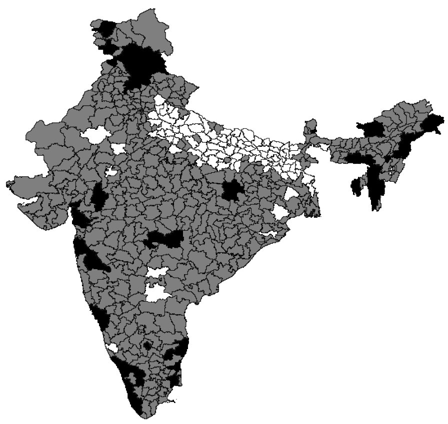
Figure 6.3: Spatial Effects on Self-employment Choice



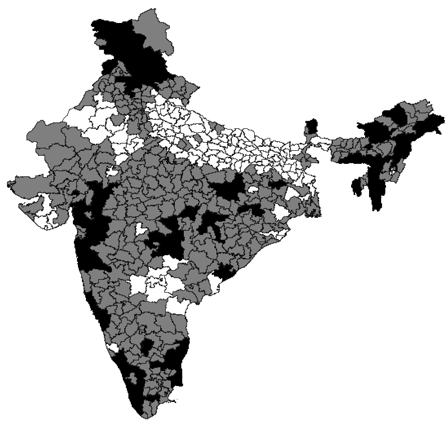
(a) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(b) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

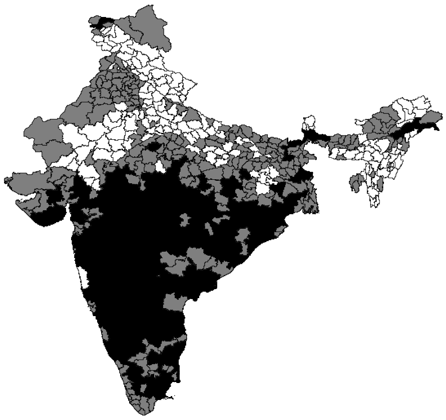


(c) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

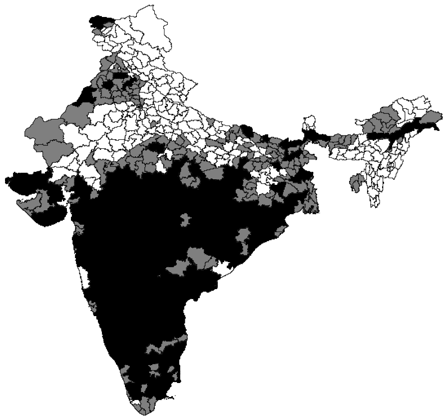


(d) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

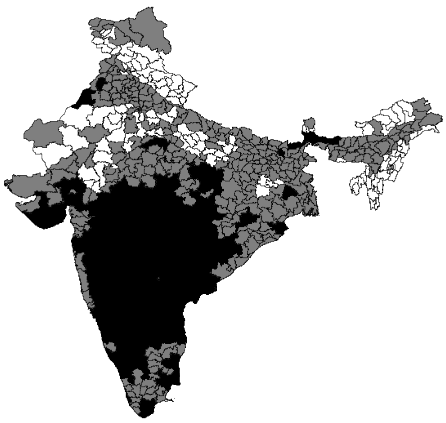
Figure 6.4: Spatial Effects in 'Nonagriculture'



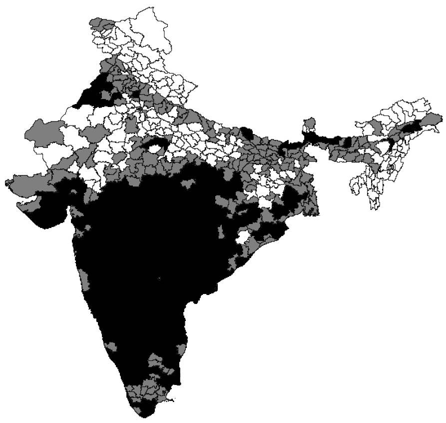
(a) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(b) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(c) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



(d) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

Figure 6.5: Spatial Effects in 'Agriculture'

Table 6.6: Determinants of Spatial Patterns in Figure 6.3, Figure 6.4 and Figure 6.5

	All		Nonagriculture		Agriculture	
Financial Development						
Per-Capita Credit	-0.0412**		-0.0457***		-0.0278	
	(0.019)		(0.016)		(0.044)	
Credit-Deposit Ratio		-0.138***		0.0411*		-0.445***
		(0.028)		(0.025)		(0.064)
Economic Development						
Per-Capita NSDP	-0.240***	-0.230***	-0.304***	-0.380***	-0.393***	-0.230***
	(0.045)	(0.039)	(0.038)	(0.034)	(0.10)	(0.088)
Unemployment	-0.0471***	-0.0242	0.0683***	0.0679***	-0.272***	-0.209***
	(0.018)	(0.018)	(0.016)	(0.016)	(0.042)	(0.041)
Demographics						
Mid Size District	-0.00474	0.0124	0.0744***	0.0678**	-0.194**	-0.136*
	(0.033)	(0.032)	(0.028)	(0.028)	(0.077)	(0.073)
Large District	0.0562	0.0578	0.0850	0.0819	-0.104	-0.0958
	(0.098)	(0.096)	(0.084)	(0.084)	(0.23)	(0.22)
Population Density	-0.00574	-0.0208	0.0909***	0.0785***	-0.0741*	-0.0895**
	(0.018)	(0.016)	(0.015)	(0.014)	(0.041)	(0.037)
Agglomeration Index						
Firm Density	0.0295**	0.0251*	0.0141	0.00405	0.0561*	0.0633**
	(0.014)	(0.013)	(0.012)	(0.012)	(0.033)	(0.030)
Constant	2.988***	2.311***	2.936***	3.125***	4.901***	2.734***
	(0.43)	(0.44)	(0.37)	(0.39)	(1.00)	(1.01)
Observations	471	471	471	471	470	470
R^2	0.13	0.17	0.37	0.36	0.14	0.22
F	10.05	13.20	38.17	37.02	10.86	18.89
R^2 Adjusted	0.119	0.154	0.356	0.349	0.128	0.211

Notes: *Signifies $p< 0.05$; ** Signifies $p<0.01$; *** Signifies $p<0.001$. Standard errors are reported in parentheses. Dependent variable is the mean spatial effect after estimation of the geoadditive models. The coefficients of the constant term are not reported.

Table 6.7: Determinants of 95% Spatial Patterns in Figure 6.3, Figure 6.4 and Figure 6.5

	All			Nonagriculture			Agriculture		
	-ve *black*	*insig.* *grey*	*+ve* *white*	*-ve* *black*	*insig.* *grey*	*+ve* *white*	*-ve* *black*	*insig.* *grey*	*+ve* *white*
Financial Development									
Per-Capita Credit	0.0462	0.0207	-0.0669**	0.0508*	0.0106	-0.0614**	0.0326	0.0498	-0.0825**
	(0.031)	(0.031)	(0.032)	(0.027)	(0.032)	(0.027)	(0.035)	(0.033)	(0.032)
Economic Development									
Per-Capita NSDP	0.406***	-0.0566	-0.350***	0.271***	0.148*	-0.419***	0.286***	-0.0841	-0.202***
	(0.079)	(0.075)	(0.075)	(0.074)	(0.087)	(0.077)	(0.084)	(0.074)	(0.075)
Unemployment	0.0876***	-0.0240	-0.0636**	-0.0803***	0.0104	0.0699**	0.214***	-0.0421	-0.172***
	(0.031)	(0.030)	(0.029)	(0.026)	(0.032)	(0.030)	(0.037)	(0.033)	(0.031)
Demographics									
Mid Size District	0.121**	-0.196***	0.0749	-0.0371	-0.149***	0.186***	0.260***	-0.269***	0.00890
	(0.056)	(0.050)	(0.054)	(0.048)	(0.057)	(0.054)	(0.058)	(0.046)	(0.052)
Large District	0.0419	-0.0806	0.0387	-0.0954	-0.258	0.354	0.324**	-0.242***	-0.0817
	(0.17)	(0.14)	(0.19)	(0.12)	(0.17)	(0.22)	(0.16)	(0.058)	(0.15)
Population Density	-0.0514*	0.0720**	-0.0206	-0.0366	-0.0533*	0.0899***	-0.133***	0.207***	-0.0743**
	(0.028)	(0.029)	(0.029)	(0.026)	(0.031)	(0.027)	(0.034)	(0.035)	(0.030)
Agglomeration Index									
Firm Density	0.0137	-0.0693***	0.0555**	0.00503	-0.00901	0.00398	0.000734	-0.0756***	0.0749***
	(0.023)	(0.023)	(0.023)	(0.021)	(0.023)	(0.019)	(0.025)	(0.025)	(0.024)
Observations	471			471			470		
Log Likelihood	-461.0			-390.2			-437.0		
$\chi^2(14)$	109.8			205.8			150.2		
Pseudo R^2	0.106			0.209			0.147		

Table 6.8: Determinants of 95% Spatial Patterns in Figure 6.3, Figure 6.4 and Figure 6.5

	All			Nonagriculture			Agriculture		
	-ve *black*	*insig.* *grey*	*+ve* *white*	*-ve* *black*	*insig.* *grey*	*+ve* *white*	*-ve* *black*	*insig.* *grey*	*+ve* *white*
Financial Development									
Credit-Deposit Ratio	0.232***	-0.0376	-0.194***	-0.111**	0.106**	0.00467	0.321***	-0.0739	-0.247***
	(0.049)	(0.048)	(0.046)	(0.045)	(0.049)	(0.038)	(0.057)	(0.051)	(0.050)
Economic Development									
Per-Capita NSDP	0.376***	-0.0256	-0.350***	0.378***	0.139*	-0.517***	0.192**	-0.00184	-0.190***
	(0.072)	(0.068)	(0.067)	(0.069)	(0.080)	(0.072)	(0.077)	(0.066)	(0.069)
Unemployment	0.0545*	-0.0242	-0.0303	-0.0728***	-0.00943	0.0822***	0.187***	-0.0508	-0.136***
	(0.031)	(0.031)	(0.029)	(0.027)	(0.032)	(0.030)	(0.037)	(0.033)	(0.031)
Demographics									
Mid Size District	0.0859	-0.185***	0.0990*	-0.0142	-0.176***	0.190***	0.232***	-0.266***	0.0339
	(0.057)	(0.052)	(0.056)	(0.050)	(0.057)	(0.054)	(0.062)	(0.048)	(0.055)
Large District	0.0204	-0.0691	0.0487	-0.0997	-0.332**	0.432**	0.343**	-0.258***	-0.0852
	(0.18)	(0.15)	(0.19)	(0.11)	(0.15)	(0.20)	(0.16)	(0.057)	(0.14)
Population Density	-0.0334	0.0743***	-0.0409	-0.0190	-0.0448	0.0638***	-0.120***	0.216***	-0.0962***
	(0.027)	(0.028)	(0.028)	(0.026)	(0.029)	(0.024)	(0.034)	(0.034)	(0.032)
Agglomeration Index									
Firm Density	0.0165	-0.0630***	0.0465**	0.0169	-0.00898	-0.00789	-0.00446	-0.0581**	0.0626***
	(0.023)	(0.023)	(0.022)	(0.021)	(0.023)	(0.018)	(0.025)	(0.024)	(0.024)
Observations	471			471			470		
Log Likelihood	-449.0			-390.5			-420.0		
$\chi^2(14)$	133.6			205.3			184.2		
Pseudo R^2	0.130			0.208			0.180		

Table 6.9: Pseudo Panel Estimation

	All	Nonagri.	Agri.
Age	0.00996***	0.0139***	0.0111***
	(0.0013)	(0.0023)	(0.0013)
Married	0.168***	0.103***	0.0865***
	(0.025)	(0.032)	(0.032)
General Education			
Informal Education	0.240***	0.146	0.425***
	(0.070)	(0.13)	(0.086)
Primary School	-0.109**	-0.330***	0.346***
	(0.048)	(0.073)	(0.071)
High School	0.00299	-0.309***	0.299***
	(0.054)	(0.071)	(0.073)
Diploma/University Education	0.0286	-0.313***	0.229*
	(0.067)	(0.092)	(0.13)
Technical Education			
Technical Degree	-0.0884	0.109	-0.811
	(0.28)	(0.31)	(1.19)
Technical Diploma	-0.262	-0.413**	-0.379
	(0.17)	(0.19)	(0.53)
Urban	-0.0380	0.103	0.0856
	(0.060)	(0.068)	(0.18)
Agriculture	0.490***		
	(0.063)		
Hindu	0.0571	0.0110	-0.0845
	(0.044)	(0.074)	(0.16)
SC/ST	-0.296***	-0.556***	-0.0246
	(0.055)	(0.11)	(0.084)
Constant	-0.212***	0.0356	-0.0765
	(0.082)	(0.11)	(0.15)
Observations	492	375	279
Number of Cohorts	164	125	93
R-squared	0.76	0.54	0.84
F	82.22	25.95	86.28
Test F(u_i=0)	8.829	3.201	23.61

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the proportion of individuals who are self-employed in a cohort.

Table 6.10: Pseudo Panel Estimation

	All	Nonagri.	Agri.
Age	0.0123***	0.0106***	0.0150***
	(0.0015)	(0.0029)	(0.0016)
Married	0.199***	0.111***	0.0981***
	(0.024)	(0.034)	(0.030)
General Education			
Informal Education	0.0763	0.234*	0.212**
	(0.075)	(0.14)	(0.087)
Primary School	-0.175***	-0.261***	0.156**
	(0.049)	(0.077)	(0.073)
High School	-0.100*	-0.252***	0.179**
	(0.055)	(0.075)	(0.072)
Diploma/University Education	-0.0507	-0.310***	0.156
	(0.065)	(0.094)	(0.12)
Technical Education			
Technical Degree	-0.314	0.232	-1.187
	(0.27)	(0.32)	(1.09)
Technical Diploma	-0.450***	-0.503**	-0.252
	(0.17)	(0.20)	(0.49)
Urban	0.111*	0.0890	0.110
	(0.062)	(0.075)	(0.17)
Agriculture	0.354***		
	(0.063)		
Hindu	0.0226	0.0402	-0.139
	(0.043)	(0.075)	(0.14)
SC/ST	-0.197***	-0.571***	-0.0398
	(0.054)	(0.11)	(0.077)
0.2< Land <0.4 Hectares	0.0233	0.124**	-0.106
	(0.045)	(0.055)	(0.095)
0.4< Land <2 Hectares	0.525***	0.0645	0.310***
	(0.088)	(0.14)	(0.11)
Land > 2 Hectares	0.151**	-0.0625	0.289**
	(0.061)	(0.079)	(0.12)
Constant	-0.424***	0.0414	-0.262
	(0.097)	(0.14)	(0.17)
Observations	492	375	279
Number of Cohorts	164	125	93
R-squared	0.79	0.56	0.87
F	76.90	21.35	84.54
Test F(u_i=0)	8.020	3.140	14.05

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the proportion of individuals who are self-employed in a cohort.

Table 6.11: Pseudo Panel Estimation: Rural-Urban Areas (Nonagriculture)

	Rural	Urban
Age	0.00901***	0.00475
	(0.0032)	(0.0031)
Married	0.211***	0.177***
	(0.040)	(0.032)
General Education		
Informal Education	0.114	0.656***
	(0.20)	(0.16)
Primary School	-0.464***	-0.179**
	(0.099)	(0.078)
High School	-0.165*	-0.285***
	(0.095)	(0.076)
Diploma/University Education	-0.614***	-0.248***
	(0.14)	(0.088)
Technical Education		
Technical Degree	0.499	0.295
	(0.40)	(0.29)
Technical Diploma	0.0720	-0.440**
	(0.24)	(0.19)
Hindu	-0.0545	-0.219**
	(0.14)	(0.095)
SC/ST	-0.513***	-0.695***
	(0.16)	(0.10)
0.2< Land <0.4 Hectares	0.129	0.153***
	(0.093)	(0.047)
0.4< Land <2 Hectares	0.0131	0.140
	(0.14)	(0.24)
Land > 2 Hectares	0.0697	-0.139*
	(0.15)	(0.075)
Constant	0.275	0.406***
	(0.18)	(0.12)
Observations	138	264
Number of Cohorts	46	88
R-squared	0.71	0.71
F	14.90	30.43
Test F(u_i=0)	6.529	4.151

Notes: *Signifies p< 0.05; ** Signifies p<0.01; *** Signifies p<0.001. Standard errors are reported in parentheses. Dependent variable is the proportion of individuals who are self-employed in a cohort.

Chapter 7

Conclusion

This chapter analyzes entrepreneurship in light of the previous chapters of this dissertation, which provide insights into the nature of entrepreneurship in a developing country. We summarize the role of exogenous constraints such as religion and caste system on the occupational choice, in an Indian context. We also link the occupational choice of individuals and entrant firms' start-up size in the spatial context that simultaneously determines the two processes. This link is established using a simple model of occupational choice, firm start-up size and the spatial location. Furthermore, we summarize the results on returns to entrepreneurship and the underlying selection process. Finally, we link the dynamics of entrepreneurship to the process of economic development.

7.1 Exogenous Constraints and Entrepreneurship

Recent studies (Iannaccone, 1998; McCleary and Barro, 2006a) suggest that there is a definitive link between religion and economics. Contributing to this growing literature, the chapter *Religion and Entrepreneurship*, shows the role of exogenous constraints, such as religion and the caste system, in the occupational choice of individuals, using discrete choice models. The persistence of occupational decisions of Hindus based on the caste system is tested using a large dataset. It is found that the presence of occupational mobility restrictions continue to inhibit occupational choices of Hindus. In particular, the analysis of this chapter suggests that Hindus of backward classes are least likely to be self-employed. Hence, elements of religion and the caste system need to be explicitly considered in understanding what influences important economic phenomena, such as entrepreneurship.

7.2 The Dual Theory of Entrepreneurship

The chapter 2, *The Geography and Determinants of Entrepreneurship*, builds on the literature on the economics of entrepreneurship (Parker, 2004) and shows that spatial location is an important micro-determinant of self-employment choice. Using geoadditive models and Bayesian semiparametric techniques, the probability of self-employment choice is estimated as a function of personal characteristics, educational background and spatial location.

Using a database of about 90,000 people in India collected in 2004, the analysis suggests that the effects of the standard determinants such as age and gender are consistent with the literature on developed countries. However, it is seen that while education decreases the likelihood of self-employment choice in nonagriculture, it positively predicts the self-employment choice in agriculture.

The estimation shows simultaneous co-existence of neighborhoods that induce people into self-employment, and neighborhoods that decrease the probability of self-employment. One striking observation of the estimated spatial patterns of self-employment is that the likelihood of individuals to be self-employed is high in regions that are relatively poor. While this can be explained to some extent by the dual labor market hypothesis, that considers parallel existence of formal and informal sub-markets in less developed countries due to job rationing, it is compelling to observe that these spatial neighborhood patterns of self-employment choice are related to the geographic patterns of the start-up size of new firms.

The emerging literature on the determinants of start-up size shows that industry characteristics and human capital of entrepreneurs play an important role in determining the start-up size of firms (Mata and Machado, 1996; Colombo et al., 2004). The chapter 4, *The Geography of Start-up Size*, contributes to this growing body of literature by highlighting that the start-up size is not independent of the geographic location. Using Bayesian semiparametric techniques and geoadditive models, the effect of spatial location is estimated as a micro-determinant of the start-up size. Using a different database of 150,000 start-ups in the manufacturing sector, the analysis shows that the start-up size of firms is spatially skewed, with spatial effects crossing boundaries of the districts and states. The analysis suggests the presence of distinct spatial patterns of start-up size. It is further seen that financial and economic development positively predict start-up size in the region.

These two chapters together give startling revelations about the nature of entrepreneurship across the space. It is seen that many regions that have higher likelihood of self-employment, as suggested by chapter 2, are also the regions that have smaller start-up size. This, in essence, suggests that self-employment in the region is linked to start up size. In particular, regions that give birth to firms of higher size absorb the labor force that would have been compelled to choose self-employment otherwise. This leads to a reduction in self-employment levels.

While a growing body of literature examines the returns to entrepreneurship (Hamilton, 2000) and the role of entrepreneurship in economic growth in developed countries (Audretsch et al., 2006), very little is known about the returns to entrepreneurship in developing countries. The chapters *The Geography and Determinants of Entrepreneurship* and *The Geography of Start-up Size* suggest that the returns to self-employment in a developing country may be lower than returns to salaried employment, as the propensity to be self-employed is higher in poorer regions. In chapter 5, *Entrepreneurship and Welfare*, the returns to entrepreneurship are estimated using a direct measure of welfare, the per-capita consumption expenditure of the household. The returns of being an employer, a self-employed worker, a salaried employee, or a casual laborer are estimated using quantile regressions. Although it is hypothesized that returns to occupations are heterogenous across quantiles, a hierarchy of welfare amongst individuals of different occupations is discovered that is persistent across quantiles. It is seen that while households headed by employers have highest per-capita consumption expenditure, households headed by self-employed individuals are slightly poorer than those headed by salaried employees. Households headed by casual laborers are the poorest. This hierarchy is found to persist even after controlling for the industrial sectors to which the individuals belong. However, the quantile regression estimates for urban and rural regions show that while the difference between the self-employed and salaried employees increases in rural regions, it vanishes at three of the five quantiles in the urban regions.

Further whether or not a process of endogenous non-random selection determines the occupations of individuals is tested using an extension of Dubin and McFadden (1984) proposed by Bourguignon et al. (2007). The results of the extended selection models after multinomial logit estimation suggest that a process of non-random selection affects the welfare returns in each occupational group.

The selection corrected estimates of the consumption functions of different occupational groups suggest a hierarchy in the selection of people into occupations based on their unobserved abilities. It is observed that while people with higher unobserved abilities become employers, the others select into salaried employment, self-employment and casual labor, in that order. However, the analysis gives evidence of positive self-selection into self-employment and simultaneous existence of disadvantaged and competitive self-employed individuals.

In chapter 6, *The Dynamics of Entrepreneurship*, the spatio-temporal dynamics of the self-employment choice are analyzed using three waves of India's National Sample Survey Organization (NSSO) employment-unemployment surveys. First, the repeated cross-section analysis compares the determinants of self-employment choice in the years 2000 and 2004. In nonagriculture, education positively predicts self-employment in the year 2000 and negatively in the year 2004. However, in agriculture, education positively predicts self-employment in both years. The spatial dynamics are almost non-existent and the spatial patterns are persist over this time period. Pseudo panels of male cohorts based on age, region and sector are constructed using the 1994, 1999-2000, and the 2004 surveys of the NSSO, to study the dynamics of the relationship between education and self-employment (Deaton, 1985; Verbeek, 2006). The results of the pseudo-panel study are consistent with the findings of the repeated cross-sections study. In particular, it is seen that higher education decreases the likelihood of self-employment choice in nonagriculture over time. Furthermore, the individuals of backward castes have become less likely to be self-employed and the effect of religion remained stable.

7.2.1 Entrepreneurship, Start-Up Size, and the Spatial Location

We formalize some of these arguments in a simple model of occupational choice, firm start-up size and the geographic location. For simplifying the representation of the spatial context in the model, we define the *Entrepreneurial Climate* of a region as the constellation of region specific factors that determine entrepreneurship.[1]

7.2.2 A Simple Model

The model builds on the model of regional entrepreneurship of Georgellis and Wall (2000), that can be summarized in the following two equations:

$$\begin{aligned} P[s_{rj}=1] &= P[{U^{se}}_r - {U^{pe}}_r + \delta_r > 0] \\ \Rightarrow S_r &= \frac{1}{L_r}\sum_{j=1}^{L_r} P[s_{rj}=1] = F[{U^{se}}_r - {U^{pe}}_r + \delta_r] \end{aligned} \tag{7.1}$$

Here the probability of choosing self employment in a region is modeled in terms of the difference between the utilities of choosing self-employment and paid employment. S_r is the proportion of people who opt for self-employment in region r, L_r is labor force of the region, s_{rj} refers to the individual j in region r, U_r^{se} is the utility of self employment in the region r, and U_r^{pe} is the utility of paid employment. Thus, their model predicts the self employment rate in a region as a function of the differences between both utilities and a random term that differentiates a average person in region r from the rest of the average person of the whole economy.

It is assumed that there is only one industry in which firms can enter. Furthermore, it is assumed that there are n regions that are similar in every manner except in their entrepreneurial climates. The model builds on the model of Georgellis and Wall (2000) by first considering $S_r * L_r$ as a function of time and then defining net entry rate as a differential of $S_r * L_r$ with respect to time. Let X_t^r be entrepreneurial climate in region r at time t, then:

$$S_r * L_r = f(X_t^r) \tag{7.2}$$

Considering e_t^r as the net entry rate in region r at time t, the net entry rate is given by

$$e_t^r = \frac{\partial f}{\partial t} \tag{7.3}$$

[1] A region defined by its institutional factors that include its technical and research institutions, financial markets, and industrial structure is expected to influence both the rate of entry and its quality.

In order to keep the model simple, there is no exit in the regions. This assumption implies that the above equation gives the number of entrants in a region at time t.

For notational simplification, assume that each x_k is such that $\frac{\partial e_t^r}{\partial x_k} > 0, x_k \in X$. Then $\|X_t^r\|$ can be used to order regions on the strength of entrepreneurial climates, where $\|.\|$ is a simple distance metric.

Thus, $\|X_t^p\| > \|X_t^q\|$ implies that the entrepreneurial climate in region p is better than the entrepreneurial climate in region q. Let the initial start-up size of a firm j entering region r at time t, be given by V_{jt}^r.

There are two types of entry. Let T be the threshold such that, if $V_{jt}^r < T$, the firm is found on very poor initial conditions or is a micro firm.

Let the set of entrants in a region be E_t^r. This set can be partitioned in two subsets E_{1t}^r and E_{2t}^r based on the threshold condition $V_{jt}^r < T$. The first subset, E_{1t}^r consists of firms that violate the threshold condition. Then, $e_{mt}^r = |E_{mt}^r|$ for $m = 1, 2$ (where $|.|$ refers to the cardinality of the set), gives the number of entrants of each type at time t.

The mean start-up size of all entrants at time t in region r is given by,

$$\mu_t^r = \frac{1}{e_t^r} \int\limits_{j \in E_t^r} V_{jt}^r \tag{7.4}$$

The papers in this thesis suggest that regions that have a poor entrepreneurial climate increase the probability of entry but entrants are very small in size. The following two propositions are derived based on the empirical analysis in the chapters.[2]

The *proposition of cardinality*: Self-employment is likely to be higher in regions with poorer entrepreneurial climates ($\|X_t^p\| > \|X_t^q\| \Rightarrow e_t^p < e_t^q$).

The *proposition of initial conditions*: If entrepreneurial climate in a region p is better than the entrepreneurial climate in a region q, then the mean start-up size of entrants in region p should be higher than the mean start-up size of firms in region q ($\|X_t^p\| > \|X_t^q\| \Rightarrow \mu_t^p > \mu_t^q$).

These two propositions suggest that self-employment is inversely related to

[2]See chapter 2 for estimated spatial patterns, chapter 6 for a study of entrepreneurial dynamics and chapter 4 for the spatial patterns of start-up size.

start-up size of new firms.[3] Also, if the labor force is stable, in region p people move out of self-employment into wage employment, while in region q people are compelled to become self-employed, as new firms do not absorb the excess labor force. Thus, regions with superior entrepreneurial climates give birth to superior entrants who absorb the excess labor force that would be self-employed otherwise. This leads to a reduction in self-employment over time.

The empirical results in this dissertation supports these propositions. The simplest proxy of entrepreneurial climate of a region is its financial development. This can be measured by credit constraints in the region. An alternate proxy is per-capita income, as this is likely to be highly correlated with entrepreneurial climate. It is observed that the level of financial development and economic development, are be able to explain the spatial patterns of self-employment choice and the firm start-up size. While these measures of entrepreneurial climate are, surprisingly, negatively related to the self-employment choice in chapter 6, they are positively related to the firm size in chapter 4.[5] Thus, the empirical analyses in chapter 2, chapter 4 and chapter 6 suggest that in regions with poor entrepreneurial climates, entrepreneurship measured by the relative number of self-employed people is high, and the firm start-up size is low.

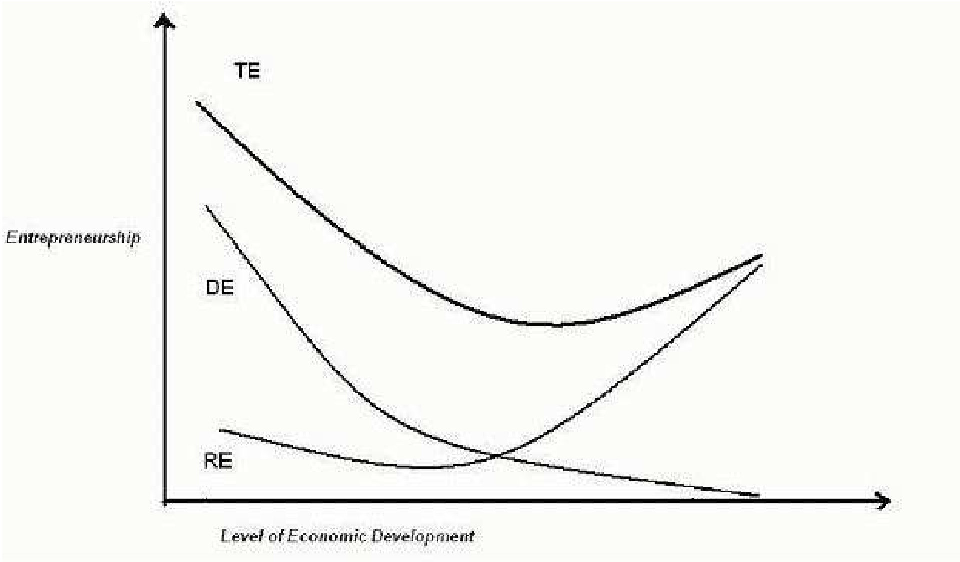
7.2.3 Entrepreneurship and Economic Development: The Dual Curve

The dual curve in Figure 7.1 captures the temporal dimensions of entrepreneurship and economic development. The chapter 6 shows that while higher education is positively related to self-employment in the year 2000, it is negatively related in the year 2004. Thus, the process of development pulls educated individuals out of self-employment and pushes them into salaried employment. Given the findings in chapter 5, this is plausible, as individuals get a signal about the relative welfare afforded by different occupations, a hierarchy in which self-employment comes lower than salaried employment. Furthermore, a superior entrepreneurial climate resulting from the process of development induces entrepreneurs to enter

[3]Under certain assumptions, it can be analytically shown that entry of type 1 (where the threshold condition is not satisfied and the start-up size is higher than T), e^p_{1t}, will be higher than e^q_{1t}, suggesting that regions with better entrepreneurial climates induce better firms to enter.[4] Conversely, it can also be shown that e^p_{2t} would be lower than e^q_{2t}.

[5]The financial development measures are insignificant in chapter 2.

Figure 7.1: Entrepreneurship and Economic Development



with a higher firm size (chapter 4). These firms absorb the excess labor force that would be compelled to choose self-employment otherwise, leading to a decrease in self-employment. Thus, while in the beginning there is high level of self-employment, the level drops with time and with economic development. This is shown by the curve of distressed entrepreneurship (DE). Simultaneously, a different type of entrepreneurship emerges in the form of employers and self-employed professionals who gradually increase in number. This is shown by the curve of real entrepreneurship (RE). This is supported by the selection models in chapter 5, that show positive self-selection into self-employment in the year 2004. Thus, entrepreneurship in a developing country assumes a variety of forms.[6] The curve TE shows the aggregate entrepreneurship and the curves DE and RE are components of TE.

7.3 Conclusion

This chapter formalizes the results of the chapters in this dissertation. It builds a coherent theory linking self-employment choice with the start-up size of new firms

[6]The simplest example is the simultaneous presence of employers and self-employed individuals, who have distinct personalities.

and the spatial location. In particular, this chapter argues that self-employment should not be viewed independently of the firm size of entrants in the region. Higher levels of self-employment are mostly correlated inversely with the size of entrants. A high level of entrepreneurship, given by a high degree of self-employment in the region, should go hand-in-hand with a higher initial size of new firms, for entrepreneurship to make an impact on the economy. The chapter also defines the entrepreneurial climate of the region for linking the individual, the firm, and the region in a simple theoretical model and suggests that superior entrepreneurial climates entail better entrants in the region. Such entrants absorb the excess labor force and lead to a reduction in the level of people who are compelled to opt self-employment. It is also argued that educated individuals who are self-employed move into salaried employment with the process of economic development, but at later stages of development, may come back to entrepreneurship as employers. Finally, the presence of exogenous constraints such as religion and caste system are found to influence the economic decision making of the individuals.

Bibliography

Acs, Zoltan J. and David B. Audretsch, "Small Firm Entry in US Manufacturing," *Economica*, 1989, *56* (222), 256–66.

_ , _ , and David S. Evans, "Why does Self Employment Rate vary Across Countries and Over Time," CEPR Discussion Paper No. 871 1994.

Agarwal, Rajshree and David B. Audretsch, "Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival," *Journal of Industrial Economics*, 2001, *49*, 21–43.

Aghion, Philippe, Peter Howitt, and David Mayer-Foulkes, "The Effect of Financial Development on Convergence: Theory and Evidence," *The Quarterly Journal of Economics*, 2005, *120* (1), 173–222.

_ , Richard Blundell, Rachel Griffith, Peter Howitt, and Susanne Prantl, "Effect of Entry on Incumbent Innovation and Production," *Journal of the European Economic Association*, 2004, *2*, 265–276.

Anderson, Gary M., "Mr. Smith and the Preachers: The Economics of Religion in the Wealth of Nations," *Journal of Political Economy*, 1988, *96* (5), 1066–1088.

Arauzo-Carod, Josep-Maria and Agusti Segarra-Blasco, "The Determinants of Entry are not Independent of Start-Up Size: Some Evidence from Spanish Manufacturing," *Review of Industrial Organization*, 2005, *27*, 147–165.

Armington, Catherine and Zoltan Acs, "The Determinants of Regional Variation in New Firm Formation," *Regional Studies*, 2002, *36*, 33–45.

Astebro, Thomas and Irwin Bernhardt, "The Winner's Curse of Human Capital," *Small Business Economics*, 2005, *24*, 63–78.

Audretsch, David B., *Innovation and Industry Evolution*, Cambridge: MIT Press, 1995.

_ and Max Keilbach, "Entrepreneurship Capital and Economic Performance," *Regional Studies*, 2004, *38*, 949–959.

— , **Enrico Santarelli, and Marco Vivarelli**, "Start-Up Size and Industrial Dynamics: Some Evidence from Italian Manufacturing," *International Journal of Industrial Organization*, 1999, *17*, 965–983.

— , **Max Keilbach, and Erik Lehman**, *Entrepreneurship and Economic Growth*, New York: Oxford University Press, 2006.

Banerjee, Abhijit V. and Andrew F. Neuman, "Occupational Choice and the Process of Economic Development," *Journal of Political Economy*, 1993, *101*, 274–298.

Barro, Robert J and Rachel McCleary, "Religion and Economic Growth Across Countries," *American Sociological Review*, 2003, *68* (3), 760–781.

Bates, Timothy, "Entrepreneur Human Capital Inputs and Small Business Longevity," *Review of Economics and Statistics*, 1990, *72* (4), 551–559.

Baumol, William J., "Entrepreneurship in Economic Theory," *The American Economic Review, Vol. 58, No 2, Papers and Proceedings of the Eightieth Annual Meeting of the American Economic Association, 64-71*, 1968.

— , *Free Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*, Princeton: Princeton University Press, 2002.

Benito, Andrew, "Does Job Insecurity Affect Household Consumption?," *Oxford Economic Papers*, January 2006, *58* (1), 157–181.

Berkowitz, Daniel and David N. DeJong, "Entrepreneurship and Post socialist Growth," *Oxford Bulletin of Economics and Statistics*, 2005, *67* (1), 21–46.

Besag, Julian, Jeremy York, and Annie Mollié, "Bayesian Image Restoration with Two Applications in Spatial Statistics," *Annals of Institute of Statistical Mathematics*, 1991, *43*, 1–59.

Bhidé, Amar, *The Origin and Evolution of New Businesses*, Oxford University Press, 2000.

Blanchflower, David G., "Self-Employment in OECD Countries," *Labour Economics*, 2000, *7* (5), 471–505.

— **and Andrew Oswald**, "What Makes an Entrepreneur?," *Journal of Labor Economics*, 1998, *16* (1), 26–60.

— **and Bruce D. Meyer**, "A Longitudinal Analysis of Young Entrepreneurs in Australia and the United States," *Small Business Economics*, 1994, *6* (1), 1–19.

— **, Andrew Oswald, and Alois Stutzer**, "Latent Entrepreneurship Across Nations," *European Economic Review*, 2001, *45* (4-6), 680–691.

Blau, David M., "Self-Employment and Self-Selection in Developing Country Labor Markets," *Southern Economic Journal*, 1985, *52* (2), 351–363.

— , "Self-employment, Earnings, and Mobility in Peninsular Malaysia," *World Development*, 1986, *14* (7), 839–852.

— , "A Time Series Analysis of Self-Employment in United States," *Journal of Political Economy*, 1987, *95* (3), 445–467.

Boháček, Radim, "Financial Constraints and Entrepreneurial Investment," *Journal of Monetary Economics*, 2006, *53* (8), 2195–2212.

Borjas, George J., "The Self-Employment Experience of Immigrants," *The Journal of Human Resources*, 1986, *21* (4), 485–506.

— **and Stephen G. Bronars**, "Consumer Discrimination and Self Employment," *Journal of Political Economy*, 1989, *97*, 581–605.

Bourguignon, Francois, Martin Fournier, and Marc Gurgand, "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons," *Journal of Economic Surveys*, 2007, *21* (1), 174–205.

Brezger, Andreas and Stefan Lang, "Generalised Structured Additive Regression Based on Bayesian P-Splines," *Computational Statistics and Data Analysis*, 2005, *50* (4), 967–991.

Browning, Martin, "Children and Household Economic Behavior," *Journal of Economic Literature*, September 1992, *30* (3), 1434–75.

— **and Annamaria Lusardi**, "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, December 1996, *34* (4), 1797–1855.

— **and Thomas F. Crossley**, "The Life-Cycle Model of Consumption and Saving," *Journal of Economic Perspectives*, Summer 2001, *15* (3), 3–22.

Cable, J. and J. Schwalbach, "International Comparisons of Entry and Exit," in J. Cable and P. A. Geroski, eds., *Entry and Market Contestability: an international comparison*, Blackwell, 1991, pp. 257–281.

Cabral, Luis M.B. and Jose Mata, "On the Evolution of the Firm Size Distribution: Facts and Theory," *The American Economic Review*, 2003, *93* (4), 1075–1090.

Calvo, Guillermo.A. and Stanislaw Wellisz, "Technology, Entrepreneurs, and Firm Size," *The Quarterly Journal of Economics*, 1980, *95* (4), 663–677.

Carrasco, Raquel, "Transitions to and from Self-Employment in Spain: An Empirical Analysis," *Oxford Bulletin Of Economics And Statisticschis*, 1999, *61*, 315–341.

Carroll, Christopher D., "Portfolios of the Rich," NBER Working Paper 2000.

Caves, Richard E., "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, 1998, *36* (4), 1947–1982.

Chiswick, Carmel Ullman, "On Estimating Earnings Functions for LDCs," *Journal of Development Economics*, 1976, *3* (1), 67–78.

Clark, Kenneth and Stephen Drinkwater, "Ethnicity and Self-Employment in Britain," *Oxford Bulletin of Economics and Statistics*, 1998, *60* (4), 383–407.

Colombo, Massimo G. and Luca Grilli, "Start-Up Size: The Role of External Financing," *Economic Letters*, 2005, *88*, 243–250.

_ **, Marco Delmastro, and Luca Grilli**, "Entrepreneurs' Human Capital and the Start-Up Size of New Technology-Based Firms," *International Journal of Industrial Organization*, 2004, *22*, 1183–1211.

Cressy, Robert, "Are Business Startups Debt-Rationed?," *The Economic Journal*, 1996, *106* (438), 1253–1270.

Cunningham, Wendy V. and William F. Maloney, "Heterogeneity in the Mexican Micro-enterprise Sector," *Economic Development and Cultural Change*, 2001, *50*, 131–156.

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70* (6), 2367–2420.

Deaton, Angus S., "Panel data from Time Series of Cross-Sections," *Journal of Econometrics*, 1985, *30* (1-2), 109–126.

_ , "The Analysis of Household Surveys: A Microeconometric Approach to Development Policy," 1997.

_ **and Christina H. Paxson**, "Measuring Poverty Among the Elderly," *NBER Working Paper*, 1995, *5296.*

_ **and John Muellbauer**, *Economics and Consumer Behavior*, Cambridge University Press, 1980.

_ **and** _ , "On Measuring Child Costs: With Applications to Poor Countries," *The Journal of Political Economy*, 1986, *94* (4), 720–744.

Dixit, Avinash, "Entry and Exit Decisions Under Uncertainty," *Journal of Political Economy*, 1989, *97*, 620–638.

Dreze, Jean and P. V. Srinivasan, "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data," *Journal of Development Economics*, December 1997, *54* (2), 217–234.

Dubin, Jeffrey A. and Daniel L. McFadden, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 1984, *52* (2), 345–362.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "The Growth and Failure of U.S. Manufactoring Plants," *The Quarterly Journal of Economics*, 1989, *104* (4), 671–698.

Eilers, Paul H. C. and Brian D. Marx, "Flexible Smoothing Using VB-Splines and Penalized Likelihood," *Statistical Science*, 1996, *11* (2), 89–121.

Eisenstadt, Shmuel. N., "The Protestant Ethic Thesis in an Analytical and Comparative Framework," in S. N. Eisenstadt, ed., *The Protestant Ethic and Modernization*, Basic Books Inc., New York, 1968, pp. 3–45.

Evans, David S. and Boyan Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, 1989, *97* (4), 808–827.

_ **and** _ , "Small Business Formation by Unemployed and Employed Workers," *Small Business Economics*, 1990, *2*, 319–330.

_ **and Linda S. Leighton**, "The Determinants of Changes in US Self-Employment, 1968-1987," *Small Business Economics*, 1989, *1*, 111–119.

_ **and** _ , "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, 1989, *79* (3), 519–535.

Fahrmeir, Ludwig and Gerhard Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer, 2001.

_ **and Stefan Lang**, "Bayesian Inference for Generalised Additive Mixed Models Based on Markov Random Field Priors," *Journal of the Royal Statistical Society C(Applied Statistics)*, 2001, *50*, 201–220.

_ **and** _ , "Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data," *Annals of the Institute of Statistical Mathematics*, 2001, *53*, 10–30.

Fairlie, Robert W. and Bruce D. Meyer, "Ethnic and Racial Self-Employment: Differences and Possible Explanations," *Journal of Human Resources*, 1996, *31* (4), 757–793.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1988, *1988* (1), 141–206.

Fields, Gary S., "A Guide to Multisector Labor Market Models," Social Protection Discussion Paper Series No. 0505, World Bank 2005.

Fölster, Stefan, "Do Lower Taxes Stimulate Self-Employment?," *Small Business Economics*, 2002, *19* (2), 135–145.

Fritsch, Michael and Pamela Müller, "Effects of New Business Formation on Regional Development over Time," *Regional Studies*, 2004, *38* (8), 961–975.

Fuchs, Victor R., "Self-Employment and Labor Force Participation of Older Males," *The Journal of Human Resources*, 1982, *17* (3), 339–357.

Fujita, Masahisa and Paul Krugman, "The New Economic Geography: Past, Present and the Future," *Papers in Regional Science*, 2003, *83* (1), 139–164.

Georgellis, Yannis and Howard J. Wall, "What Makes a Region More Entrepreneurial? Evidence from Britain," *Annals of Regional Science*, 2000, *34*, 385–403.

Geroski, Paul A., "What Do We Know About Entry?," *International Journal of Industrial Organization*, 1995, *13*, 421–440.

Ghatak, Maitreesh and Neville Nien-Huei Jiang, "A Simple Model of Inequality, Occupational Choice and Development," *Journal of Development Economics*, 2002, *69* (1), 205–226.

Gindling, T. H., "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal and Informal Sectors in San-Jose, Costa-Rica," *Economic Development and Cultural Change*, 1991, *39*, 585–605.

Glewwe, Paul, "Investigating the Determinants of Household Welfare in Côte d'Ivoire," *Journal of Development Economics*, April 1991, *35* (2), 307–337.

Günther, Isabel and Andrey Launov, "Competitive and Segmented Informal Labor Markets," IZA Discussion Papers 2349, Institute for the Study of Labor (IZA) September 2006.

Goedhuys, Micheline and Leo Sleuwaegen, "Entrepreneurship and Growth of Entrepreneurial Firms in Cote d' Ivoire," *Journal of Development Studies*, 2000, *36*, 122–145.

Görg, Holger and Erik Strobl, "Multinational Companies and Entrant Start-Up Size: Evidence from Quantile Regressions," *Review of Industrial Organization*, 2002, *20*, 15–31.

_ , _ , **and Frances Ruane**, "Determinants of Firm Start-Up Size: An Application of Quantile Regressions for Ireland," *Small Business Economics*, 2000, *14*, 211–222.

Guimaraes, Paulo, Jose Mata, and Pedro Portugal, "The Survival of New Plants: Start-Up Conditions and Post-Entry Evolution," *International Journal of Industrial Organization*, 1995, *13*, 459–481.

Guisa, Luigi, Paolo Sapienza, and Luigi Zingales, "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 2006, *20* (2), 23–48.

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns of Self-Employement," *Journal of Political Economy*, 2000, *108* (3), 604–631.

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two Sector Analysis," *American Economic Review*, 1970, *60* (1), 126–142.

Hastie, Trevor and Robert Tibshirani, *Generalised Additive Models*, London: Chapman and Hall, 1990.

_ **and** _ , "Bayesian Backfitting," *Statistical Science*, 2000, *15*, 193–223.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–162.

Holmes, Thomas J. and James A. Schmitz Jr., "A Theory of Entrepreneurship and Its Application to the Study of Business Transfers," *Journal of Political Economy*, 1990, *98* (2), 265–294.

Holtz-Eakin, Douglas and Chihwa Kao, "Entrepreneurship and Economic Growth: The Proof is in the Productivity.," Center for Policy Research, Syracuse University 2003.

_ **, David Joulfaian, and Harvey S. Rosen**, "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints," *The Journal of Political Economy*, 1994, *102* (1), 53–75.

Hurst, Erik and Annamaria Lusardi, "Liquidity Constraints, Household Wealth, and Entrepreneurship," *Journal of Political Economy*, 2004, *112* (2), 319–347.

Iannaccone, Laurence R., "Introduction to the Economics of Religion," *Journal of Economic Literature*, 1998, *36* (3), 1465–1495.

Iyigun, Murat F. and Ann L. Owen, "Entrepreneurs, Professionals, and Growth," *Journal of Economic Growth*, 1999, *4*, 213–232.

Jenkins, Stephen P., "Modelling Household Income Dynamics," *Journal of Population Economics*, 2000, *13* (4), 529–567.

Johnson, William, "A Theory of Job Shipping," *Quarterly Journal of Economics*, 1978, *May 1978* (22), 261–78.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *The Journal of Political Economy*, 1979, *87* (5), 972–990.

_ , "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670.

_ , "Firm Formation with Heterogenous Management and Labor Skills," *Small Business Economics*, 1994, *6* (3), 185–191.

Kammann, E. E. and Matthew P. Wand, "Geoadditive Models," *Journal of the Royal Statistical Society C*, 2003, *52*, 1–18.

Kandala, Ngianga Bakwin, Ludwig Fahrmier, and Stephan Klasen, "Geo-Additive Models of Childhood Undernutrition in Three Sub-Saharan African Countries," *SFB Discussion Paper No. 287*, 2002.

_ , Stefan Lang, Stephan Klasen, and Ludwig Fahrmeir, "Semiparametric Analysis of the Socio-Demographic Determinants of Undernutrition in Two African Countries," *Research in Official Statistics, EUROSTAT*, 2001, *4*, 81–100.

Kihlstrom, Richard E. and Jean Jacques Laffont, "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy*, 1979, *87*, 719–748.

Koenker, Roger and Kevin F. Hallock, "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 2001, *15* (4), 143–156.

Krugman, Paul, "Increasing Returns and Economic Geography," *The Journal of Political Economy*, 1991, *99* (3), 483–499.

Lang, Stefan and Andreas Brezger, "Bayesian P-Splines," *Journal of Computational and Graphical Statistics*, 2004, *13*, 183–212.

Lanjouw, Peter and Martin Ravallion, "Poverty and Household Size," *Economic Journal*, November 1995, *105* (433), 1415–34.

Lazear, Edward P., "Entrepreneurship," Working Paper No. 9109, National Bureau of Economic Research, Cambridge, MA. 2002.

_ , "Entrepreneurship," *Journal of Labour Economics*, 2005, *23* (4), 649–680.

_ **and Robert T. Michael**, "Family Size and the Distribution of Real Per Capita Income," *American Economic Review*, March 1980, *70* (1), 91–107.

Le, Anh T., "Empirical Studies of Self-Employment," *Journal of Economic Surveys*, 1999, *13* (4), 381–416.

Lee, Lung-Fei, "Generalized Econometric Models with Selectivity," *Econometrica*, 1983, *51* (2), 507–512.

Levine, R., "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature*, 1997, *35* (2), 688–726.

Lewis, Arthur W., *Economic Development with Unlimuted Supplies of Labour* 1954.

Lin, X. and D. Zhang, "Inference in Generalised Additive Mixed Models by Using Smoothing Splines," *Journal of the Royal Statistical Society B,*, 1999, *61*, 381–400.

Lindh, T. and H. Ohlsson, "Self-Employment and Windfall Gains: Evidence from the Swedish Lottery," *The Economic Journal*, 1996, *106* (439), 1515–1526.

Lucas, Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9*, 508–523.

Magnac, Th., "Segmented or Competitive Labor Markets," *Econometrica*, 1991, *59*, 165–187.

Majumdar, Dipak, *The Urban Labor Market and Income Distribution - A Study of Malaysia*, Oxford University Press, 1981.

Maloney, William F., "Informality Revisited," *World Development*, 2004, *32* (7), 1159–1178.

Mata, Jose, "Market, Entrepreneurs and the Size of New Firms," *Economic Letters*, 1996, *52*, 89–94.

_ **and Jose A.F. Machado**, "Firm Start-Up Size: A Conditional Quantile Approach," *European Economic Review*, 1996, *40*, 1305–1323.

McCelland, David, *The Achieving Society*, Princeton, NJ: D. Van Nostrand, 1964.

McCleary, Rachel M. and Robert J. Barro, "Religion and Economy," *Journal of Economic Perspectives*, 2006, *20* (2), 49–72.

_ **and** _ , "Religion and Political Economy in an International Panel," *Journal of Scientific Study of Religion*, 2006, *45* (2), 149–175.

McKenzie, David, "Asymptotic Theory for Heterogeneous Dynamic Pseudo-panels," *Journal of Econometrics*, 2004, *120* (2), 235–262.

Medhora, Phiroze B., "Entrepreneurship in India," *Political Science Quarterly, Vol 80, No 4, 558-580*, 1965.

Miles, David, "A Household Level Study of the Determinants of Incomes and Consumption," *Economic Journal*, January 1997, *107* (440), 1–25.

Miller, Robert, "Job Matching and Occupational Choice," *Journal of Political Economy*, 1984, *December 1984* (92), 1086–120.

Minns, Chris and Marian Rizov, "The Spirit of Capitalism? Ethnicity, Religion, and Self-Employment in Early 20th Century Canada," *Explorations in Economic History*, 2005, *42*, 259–281.

Moffitt, Robert, "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections," *Journal of Econometrics*, 1993, *59* (1), 99–124.

Mohapatra, Sandeep, Scott Rozelle, and Rachael Goodhue, "The Rise of Self-Employment in Rural China: Development or Distress?," *World Development*, 2007, *35* (1), 163–181.

Muellbauer, John, "Household Composition, Engel Curves and Welfare Comparisons Between Households," *European Economic Review*, 1974, *5*, 103–122.

Munshi, Kaivan and Mark Rosenzweig, "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 2006, *96* (4), 1225–1252.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics*, 1991, *106* (2), 503–530.

Nelson, Julie A., "Household Economies of Scale in Consumption: Theory and Evidence," *Econometrica*, November 1988, *56* (6), 1301–14.

Nurmi, Satu, "Sectoral Differences in Plant Start-Up Size in the Finnish Economy6," *Small Business Economics*, 2006, *26*, 39–59.

Parker, Simon C., "A Time Series Model of Self-employment Under Uncertainty," *Economica*, 1996, *63*, 469–475.

— , *The Economics of Self-Employment and Entrepreneurship*, Cambridge: Cambridge University Press, 2004.

— **, Yacine Belghitar, and Tim Barmby**, "Wage Uncertainty and the Labor Supply of Self-Employed Workers," *Economic Journal*, 2005, *115*, 190–207.

Pradhan, Menno and Martin Ravallion, "Measuring Poverty Using Qualitative Perceptions Of Consumption Adequacy," *The Review of Economics and Statistics*, August 2000, *82* (3), 462–471.

Pratap, Sangeeta and Erwaan Quintin, "Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach," *European Economic Review*, 2006, *50*, 1817–1841.

Premi, Mahendra, "Religion in India: A Demographic Perspective," *Economics and Political Weekly*, September 2004, pp. 4292–4297.

Quinn, Joseph F., "Labor-force Participation Patterns of Older Self-employed Workers," *Social Security Bulletin*, 1980, *43* (4), 17–28.

Ranis, Gustav and John Fei, "A Theory of Economic Development," *American Economic Review*, 1961, *51*, 533–565.

Rees, H. and Anup Shaw, "An Empirical Analysis of Self-Employment in UK," *Journal of Applied Econometrics*, 1986, *1*, 95–108.

Reynolds, Paul D., David J. Storey, and Paul Westhead, "Cross National Comparison of the Variation in New Firm Formation Rates," *Regional Studies*, 1994, *28*, 443–456.

Rosen, Sherwin, "The Economics of Superstars," *The American Economic Review*, 1981, *71* (5), 845–858.

Rosenzweig, Mark R., "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labour Supply in a Developing Country," *Quarterly Journal of Economics*, 1980, *94* (1), 31–55.

Sen, Amartya K., "Peasants and Dualism with or without Surplus Labor," *Journal of Political Economy*, 1966, *74*, 425–450.

Singer, Milton, "Religion and Social Change in India: The Max Weber Thesis Phase Three," *Economic Development and Cultural Change*, 1966, *14*, 497–505.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde, "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2002, *64* (2), 583–639.

Storey, David J., "The Birth of New Firms - Does Unemployment Matter? A Review of the Evidence," *Small Business Economics*, 1991, *3* (3), 167–178.

_ **and A. M. Jones**, "New Firm Formation - A Labour Market Approach to Industrial Entry," *Scottish Journal of Political Economy*, 1987, *34*, 37–51.

Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, 1997, *35* (1), 40–59.

Taylor, Mark P., "Earnings, Independence or Unemployment: Why Become Self-Employed?," *Oxford Bulletin of Economics and Statistics*, 1996, *58*, 253–266.

Uppal, J. S., "Hinduism and Economic Development in South Asia," *International Journal of Social Sciences*, 2001, *13* (3), 20–33.

van der Sluis, Justin, Mirjam van Praag, and Wim Vijverberg, "Entrepreneurship Selection and Performance: A Meta-Analysis of the Impact of Educaton in Developing Economies," *World Bank Economic Review*, 2005, *2*, 225–261.

van Praag, C. Miriam and J. S. Cramer, "The Roots of Entrepreneurship and Labour Demand: Individual Ability and Low Risk Aversion," *Economica*, 2001, *68* (269), 45–62.

Verbeek, Marno, "Pseudo Panels and Repeated Cross Sections," in L Matyas and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of the Theory with Applications*, 2006.

_ **and Francis Vella**, "Estimating Dynamic Models from Repeated Cross-Sections," *Journal of Econometrics*, 2005, *127*, 83–102.

_ **and Theo Nijman**, "Can Cohort Data be Treated as Genuine Panel Data?," *Empirical Economics*, 1992, *17* (1), 9–23.

_ **and** _ , "Minimum MSE Estimation of a Regression Model with Fixed Effects and a Series of Cross Sections," *Journal of Econometrics*, 1993, *59*, 125–136.

Weber, Max, *The Religion of India*, New York: The Free Press, 1958.

Wennekers, Sander and Roy Thurik, "Linking Entrepreneurship and Economic Growth," *Small Business Economics: An International Journal*, 1999, *13*, 27–55.

_ , **Andre van Stel, Roy Thurik, and Paul Reynolds**, "Nascent Entrepreneurship and Level of Economic Development," *Small Business Economics*, 2005, *24*, 293–309.

White, Lwrence J., "The Determinants of the Relative Importance of Small Business," *Review of Economics and Statistics*, 1982, *64*, 42–49.

Wodon, Quentin T., "Microdeterminants of Consumption, Poverty, Growth, and Inequality in Bangladesh," *Applied Economics*, 2000, *32* (10), 1337–1352.

Yamada, Gustavo, "Urban Informal Employment and Self-Employment in Developing Countries: Theory and Evidence," *Economic Development and Cultural Change*, 1996, *44*, 289–314.

Zucker, Lynne G., Micheal R. Darby, and Marilynn B. Brewer, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 1998, *88*, 290–306.

Curriculum Vitae

October 2007: PhD (Economics), University of Göttingen

September 2004- September 2007: PhD Candidate, Entrepreneurship, Growth and Public Policy Group, Max Planck Institute of Economics, Jena, Germany

April 2004- September 2007: Doctoral Fellow, Chair for Development Economics, University of Göttingen, Göttingen, Germany

July 2002- October 2003: Research Scholar and Tutor, Indian Institute of Technology, New Delhi, India

2002: Master of Science (Mathematics), Sri Sathya Sai Institute of Higher Learning, Prashanti Nilayam, Andhra Pradesh, India

2000: Bachelor of Science (Honours), Sri Sathya Sai Institute of Higher Learning, Prashanti Nilayam, Andhra Pradesh, India

1997: Senior Secondary School Certificate, Rama Krishna Mission Residential College, Narendrapur, West Bengal, India

1995: Secondary School Certificate, Assembly of God Church School, Calcutta, West Bengal, India