

Tamvada, Jagannadha Pawan

**Working Paper**

## The geography and determinants of entrepreneurship

Jena Economic Research Papers, No. 2008,026

**Provided in Cooperation with:**

Max Planck Institute of Economics

*Suggested Citation:* Tamvada, Jagannadha Pawan (2008) : The geography and determinants of entrepreneurship, Jena Economic Research Papers, No. 2008,026, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/25704>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2008 – 026

## **The Geography and Determinants of Entrepreneurship**

**by**

**Jagannadha Pawan Tamvada**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [m.pasche@wiwi.uni-jena.de](mailto:m.pasche@wiwi.uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

# The Geography and Determinants of Entrepreneurship

March 2008

**Jagannadha Pawan Tamvada**  
Max Planck Institute of Economics

## Abstract

Entrepreneurship literature (Parker 2004) has rarely considered spatial location as a *micro-determinant* of occupational choice, although there are compelling reasons to posit that spatial location influences economic behavior. Using Bayesian semiparametric methodologies and geoadditive techniques, we examine spatial location as a *micro-determinant* of self-employment choice of Indians, in addition to standard determinants such as age, gender and education. The empirical analysis suggests the presence of spatial occupational neighborhoods and a clear north-south divide in self-employment choice in India when individuals of agricultural and nonagricultural sectors are considered together; however, such spatial patterns are less pronounced when individuals in nonagriculture alone are considered in the analysis. These residual spatial patterns are found to be inversely related to the per-capita GDP of the region. The results further suggest nonlinear relationships between age, wealth and the probability of self-employment.

JEL-classification: J24, J43, J44, L26

Keywords: Entrepreneurship, Self-employment, Developing Countries, Bayesian Semiparametric Methods, Geoadditive Models

Acknowledgement: Discussions with Stephan Klasen have been very useful in the preparation of this manuscript. I am grateful to David Audretsch, Michael Grimm, Adam Lederer, Mili Shrivastava, participants at Max Planck IISc International Conference and seminar participants in Goettingen for helpful discussions and comments.

Contact: Jagannadha Pawan Tamvada, Entrepreneurship, Growth and Public Policy Group, Max Planck Institute of Economics and Faculty of Economics, University of Goettingen. Address for Correspondence: Max Planck Institute of Economics, Kahlaische Strasse 10, 07745 Jena, Germany. Telephone: +49-3641-686730, Fax: +49-3641-686710, Email: tamvada@econ.mpg.de

# 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.<sup>1</sup> Parallel to this body of literature linking entrepreneurship to the economic progress,<sup>2</sup> 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 of an individual has rarely been considered as a *micro-determinant* of the entrepreneurship choice, although there are compelling reasons to assume that it plays an important role in shaping the occupational choice of people. Existing studies examine the influence of region on self-employment choice either by using dummy variables (Blanchflower and Oswald, 1998; Borjas and Bronars, 1989; Fairlie and Meyer, 1996) or by introducing regional characteristics in the estimations (Evans and Jovanovic, 1989; Blanchflower and Oswald, 1998; Glaeser, 2007). Some other studies are based on aggregate self-employment data (Georgellis and Wall, 2000; Fölster, 2002; Torrini, 2005). None of these approaches consider spatial location as a *micro-determinant* of self-employment choice. Further, they do not allow the estimation of neighborhood effects, although, as the results of this paper suggest, neighborhood effects play an important role in determining self-employment choice. This study identifies a new approach to analyze the determinants of entrepreneurship using recent advances in Bayesian semipara-

---

<sup>1</sup>For instance, 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.

<sup>2</sup>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.

metric geoaddivitive models that allow incorporation of spatial location as an additional *micro-determinant* of self-employment choice. Thus, the estimated spatial patterns reflect the propensity of people to be self-employed in a region, after controlling for other individual-level effects.

Furthermore, the entrepreneurship literature, until recently, has largely ignored the labor markets of Less Developed Countries (LDCs). An assumption 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, and therefore choice was assumed away.<sup>3</sup> 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 suggesting that choice is relevant (Maloney, 2004; Mohapatra et al., 2007). Given these new insights, it is pertinent to examine the determinants and the role of entrepreneurship in less developed countries. This paper meets this need, by examining the determinants of self-employment choice in one such growing economy, India. In recent years, India has experienced substantial leaps in both its entrepreneurial activity and growth rates. Hence this study makes a timely and important contribution to this neglected area of research.

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 geoaddivitive models. While earlier studies mostly considered nonagricultural self-employment, we analyze self-employment in the agriculture as well. Household level data collected by the National Sample Survey Organization in 2004 are used for the empirical analysis. The

---

<sup>3</sup>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.”

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. Further, strong residual spatial patterns are observed that remain after controlling for the personal, demographic and education characteristics of individuals. The spatial patterns are less pronounced when individuals in nonagriculture alone are considered in the analysis suggesting that personal, demographic and education variables explain most of the variation in self-employment choice in nonagriculture. Further, nonagricultural self-employment patterns are not correlated with agricultural patterns, which suggests considering the two separately. Consistent with earlier empirical studies on the determinants of entrepreneurship, the results show that 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 geoaddivitive modeling techniques and the dataset. The fifth section presents the empirical results. The final section provides conclusions and discusses possible avenues for future research.

## 2 Theoretical Background

### Entrepreneurship 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 mar-

kets along the formal-informal lines (also see Sen, 1966; Ranis and Fei, 1961; Harris and Todaro, 1970).<sup>4</sup> 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).<sup>5</sup> 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).<sup>6</sup>

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. Thus, examining the determinants of self-

---

<sup>4</sup> 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.

<sup>5</sup>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.

<sup>6</sup>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.

employment choice, in light of the formal-informal sector debate, remains an interesting avenue for future research.

## **Determinants of Entrepreneurship**

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 traits such as drive for achievement (McClland, 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, "[t]he 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) 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.<sup>7</sup>

Though there are compelling reasons to posit that there are sectoral differences in self-employment choice, male, married and older individuals are more likely to be self-employed in general.<sup>8</sup> 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.<sup>9</sup> 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

---

<sup>7</sup>Other examples of studies analyzing the determinants of entrepreneurship include Evans and Jovanovic (1990) and Parker et al. (2005).

<sup>8</sup>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.

<sup>9</sup>Goedhuys and Sleuwaegen (2000), for instance, find this to be true for individuals in Côte d'Ivoire.

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 significant 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, “[a]s 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.<sup>10</sup> As Boháček (2006, p.2196) notes, “[i]n 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

---

<sup>10</sup>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).

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.<sup>11</sup> 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 affects 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.

## **The Role of Spatial Location**

There are three standard approaches to analyzing the role of region on self-employment choice. The first approach controls for the influence of the region through dummy variables (Blanchflower and Oswald, 1998; Borjas and Bronars, 1989; Fairlie and Meyer, 1996). The obvious limitation here is that a large number of dummy variables should be introduced in the estimation, if the role of the spatial location at a micro level is to be analyzed. The second approach simultaneously estimates individual as well as regional

---

<sup>11</sup>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).

characteristics as micro-determinants of self-employment choice (Evans and Jovanovic, 1989; Blanchflower and Oswald, 1998; Glaeser, 2007). This technique, however, does not allow estimation of spatial location as a continuous variate. Thus, spatial dependence of the self-employment choice on neighboring regions cannot be examined. The third approach analyzes determinants of region-level self-employment figures, based on aggregated data, and estimates the quantity of entrepreneurial activity as a function of regional variables such as unemployment and tax rates (Georgellis and Wall, 2000; Fölster, 2002; Torrini, 2005).<sup>12</sup> However, this approach ignores the underlying structure of the labor force. For instance, regressing the aggregate level of self-employment on regional indicators such as unemployment or the per-capita GDP would not account for the individual level factors that determine the self-employment choice. In addition, this approach does not allow the estimation of neighborhood effects, although, as the results of this paper suggest, they play an important role in determining self-employment choice.

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 compels people to select into self-employment in such neighborhoods.

---

<sup>12</sup>Another strand of literature examines the determinants (Reynolds et al., 1994; Armington and Acs, 2002) and the effect of new firm formation on the region (Fritsch and Müller, 2004).

### 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.<sup>13</sup>

#### 3.1 Geoadditive Models

Let  $(y_i, x_i, v_i)$  for  $i$  in  $\{1, 2, \dots, N\}$  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.<sup>14</sup> Assume  $y_i$  are independent and Gaussian with mean  $\eta_i = f_1(x_{i1}) + \dots + f_p(x_{ip}) + v_i\gamma$ , and a common variance  $\sigma^2$ . 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.<sup>15</sup> Furthermore, Lang

<sup>13</sup>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.

<sup>14</sup>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.

<sup>15</sup>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.

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.<sup>16</sup> 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,  $f_j(x_j) = \sum \beta_{jk} B_{jk}$  and, in matrix notation,  $\eta = \sum X_j \beta_j + V \gamma$ . 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:

$$L = l(y, \beta_1, \dots, \beta_p, \gamma) - \lambda_1 \sum (\Delta^k \beta_1)^2 - \dots - \lambda_p \sum (\Delta^k \beta_p)^2 \quad (1)$$

In the Bayesian framework,  $\beta_j$  for  $j = 1 \dots p$  and  $\gamma$  are considered as random variables and assigned prior distributions. Independent diffuse priors are assumed for the fixed effects parameters,  $\gamma_j \propto \text{const}$  for  $j = 1 \dots q$ . The priors for the coefficients of the non-linear functions,  $\beta_j$ , 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  $\beta_{j1}$  and  $\beta_{j2}$ .  $\tau_j^2$  controls the smoothness of the fitted function. For Bayesian inference,  $\tau_j^2$  are also treated as random variables and simultaneously estimated with the  $\beta_j$ . Highly dispersed inverse gamma priors  $IG(a_j, b_j)$  are assigned to the variances  $\tau_j^2$ .

The geoadditive model is obtained if a spatial effect,  $f_{\text{spatial}}$ , is added to the above predictor. The spatial effect may be split into spatially correlated and uncorrelated

---

<sup>16</sup>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).

effects,  $f_{spatial} = f_{str} + f_{unstr} = X_{str}\beta_{str} + X_{unstr}\beta_{unstr}$ , 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  $\beta_{str}$ . If  $s \in 1, \dots, S$  are pixels of a lattice or regions of a geographical map, the MRF prior is given as,

$$\beta_{str,s} \setminus \beta_{str,u} \sim N\left(\sum_{u \in \partial_s} \frac{1}{N_s} \beta_{str,u}, \frac{\tau_{str}^2}{N_s}\right) \quad (2)$$

for,  $u \neq s$ , where,  $N_s$  is the number of adjacent regions (pixels) and  $\partial_s$  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,  $\beta_{unstr}$ , independent and identically distributed (i.i.d.) Gaussian random priors,  $\beta_{unstr}(s) \sim N(0, \tau_{unstr}^2)$ , are assumed for  $s=1, \dots, S$ . For  $\tau_{str}^2$  and  $\tau_{unstr}^2$  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  $\alpha$  is a vector of the unknown parameters, assuming conditional independence of the parameters, the posterior is given by:

$$\begin{aligned} p(\alpha \setminus y) &\propto L(y, \beta_1, \dots, \beta_p, \beta_{str}, \beta_{unstr}, \gamma, \sigma^2) \times \prod_{j=1}^p (p(\beta_j \setminus \tau_j^2) p(\tau_j^2)) \\ &\times p(\beta_{str} \setminus \tau_{str}^2) p(\tau_{str}^2) p(\beta_{unstr} \setminus \tau_{unstr}^2) p(\tau_{unstr}^2) p(\gamma) p(\sigma^2) \end{aligned} \quad (3)$$

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  $\Phi$  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  $U_i$ .



### 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  $-2\log\{p(y|\mu)\}$ . 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  $D(\theta) = -2\log\{p(y|\theta)\} + 2\log f(y)$ , 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  $f(y) = p\{y|\mu(\theta) = y\}$ . That is,  $D(\theta) = -2\log\{p(y|\theta)\} + 2\log\{p(y|\mu(\theta) = y)\}$ . 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  $D(\bar{\theta})$ . 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.

### 3.3 Explaining the Residual Spatial Patterns

Consider estimating the geoaddivitive 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 geoaddivitive 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 geoaddivitive 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 geoaddivitive 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 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.<sup>17</sup> 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

---

<sup>17</sup>We drop 17 individuals who adhere Zoharastrianism.

and paid employment. Table 1 lists the number of 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.<sup>18</sup> 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.<sup>19</sup>

The descriptive statistics in Table 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 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-possessed 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

---

<sup>18</sup>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.

<sup>19</sup>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.

enterprize.<sup>20</sup>

These descriptive tables also show that more than 50% of individuals in agriculture 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).<sup>21</sup> 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.

## 5 Empirical Results

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.

---

<sup>20</sup>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.

<sup>21</sup>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.

## 5.1 Aggregate Model

In the first model, participation in the agricultural sector is controlled using a dummy variable. The following semiparametric geoaddivitive 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.<sup>22</sup> Figure 1(a) shows that the probability of being self-employed increases with age, confirming the age-effect. The derivative of the ‘age’ function in Figure 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.<sup>23</sup> 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).

---

<sup>22</sup>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.

<sup>23</sup>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.

The results of the parametric part of this regression model, also referred to as fixed effects, in Table 4, suggest that both married and divorced people are more likely to be self-employed compared to unmarried individuals.<sup>24</sup> 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.<sup>25</sup> 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 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.<sup>26</sup> 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

---

<sup>24</sup>This is consistent with Taylor (1996), Fairlie and Meyer (1996) and other studies that find positive effects of marital status on self-employment.

<sup>25</sup> Though dowry is legally prohibited in India, it is prevalent in numerous forms.

<sup>26</sup>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.

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.<sup>27</sup> We keep the land variables as there are compelling reasons to assume that wealth determines the entrepreneurial choice, in the Indian context.<sup>28</sup>

The map of structured spatial effects in Figure 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(c) and 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(b) are very small compared to the structured effects.<sup>29</sup> 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.

## 5.2 Sector Specific Models

### 5.2.1 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 indi-

---

<sup>27</sup>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.

<sup>28</sup>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.

<sup>29</sup>The structured spatial effects range from -0.8 to 0.8, the random unstructured local spatial effects range from -0.2 to 0.18.

viduals 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.<sup>30</sup> The relationship of age with self-employment is very close to being linear in the agricultural sector, as seen in Figure 1(e), while in the nonagricultural sector, as Figure 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 5 and Table 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 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. This is consistent with Glaeser (2007), who suggests that half of the spatial variation in self-employment rates can be explained by the underlying education, demographic and industry structure. People of Kerala and some districts of Tamil Nadu in the south, Maharashtra 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 4 show that the result of north-

---

<sup>30</sup>The autocorrelation files and plots of the sampling paths show that sufficient convergence is achieved in these models also.



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 self-employed in agriculture in southern and central states. As Figures 3(b) and 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.<sup>31</sup>

### 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 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.

---

<sup>31</sup>These plots are available from the author.

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 technological progress. Empirical studies of Acs et al. (1994) and Fölster (2002) find that per-capita 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 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 8](#) and [Table 9](#), we estimate multinomial logit models with the dependent variable as the estimated 95% spatial effects in the maps in [Figure 2](#), [Figure 3](#) and [Figure 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 8](#), we use per-capita credit as a proxy for financial development and in [Table 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 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.<sup>32</sup> 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.

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

---

<sup>32</sup>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.

measure trends that are more relevant to the nonagricultural sector.<sup>33</sup> 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 8 and Table 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 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 3(c). Similarly, the positive effect of unemployment vanishes in the nonagricultural sector in the multinomial estimations. This shows that the results of Table 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 the influence of factors such as population density and agglomeration index are controlled.

## 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 geoaddivitive 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 the determinants in urban and rural areas. We control for regional effects using a set of state level regional dummies. We estimate

---

<sup>33</sup>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.

this for the sub-sample of individuals in the nonagricultural sector alone.<sup>34</sup> 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 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 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. 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

---

<sup>34</sup>As the agricultural sector is mostly found in the rural areas only, we restrict the urban-rural analysis to the nonagricultural sector.

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 individuals 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.

## 6 Conclusion

The field of entrepreneurship in economics provides insights into the individual determinants of the self-employment choice in developed countries. While the determinants of self-employment choice in less developed countries received little attention, the role of spatial location as a *micro-determinant* of self-employment choice has rarely been analyzed. We contribute to these aspects of literature that remained neglected for a long time.

We use recent advances in Bayesian semiparametric methodologies and geoaddivitive models to examine spatial location as well as individual characteristics as 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 nonagriculture, self-employed people are more or

less uniformly distributed across space. Thus, with the exception of some districts, self-employment choice is largely independent of spatial context, in nonagriculture, once the individual characteristics are controlled. However, self-employed in people in agriculture are concentrated in 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.



## References

- 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.
- Armington, Catherine and Zoltan J. Acs**, “The Determinants of Regional Variation in New Firm Formation,” *Regional Studies*, 2002, 36, 33–45.
- Audretsch, David B., 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.
- Bates, Timothy**, “Entrepreneur Human Capital Inputs and Small Business Longevity,” *Review of Economics and Statistics*, 1990, 72 (4), 551–559.
- Baumol, William J.**, *Free Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*, Princeton: Princeton University Press, 2002.
- 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.
- Brezger, Andreas and Stefan Lang**, “Generalised Structured Additive Regression Based on Bayesian P-Splines,” *Computational Statistics and Data Analysis*, 2005, 50 (4), 967–991.
- 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 Statistics*, 1999, 61, 315–341.
- Carroll, Christopher D.**, “Portfolios of the Rich,” NBER Working Paper 2000.
- 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.
- Cunningham, Wendy V. and William F. Maloney**, “Heterogeneity in the Mexican Micro-enterprise Sector,” *Economic Development and Cultural Change*, 2001, 50, 131–156.
- Eilers, Paul H. C. and Brian D. Marx**, “Flexible Smoothing Using VB-Splines and Penalized Likelihood,” *Statistical Science*, 1996, 11 (2), 89–121.
- 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.
- 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.
- Georgellis, Yannis and Howard J. Wall**, “What Makes a Region More Entrepreneurial? Evidence from Britain,” *Annals of Regional Science*, 2000, *34*, 385–403.
- 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.
- Glaeser, Edward L.**, “Entrepreneurship and the City,” Working Paper No.13551, National Bureau of Economic Research, Cambridge, MA. 2007.
- 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.

- 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.
- 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, 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.
- Iyigun, Murat F. and Ann L. Owen**, "Entrepreneurs, Professionals, and Growth," *Journal of Economic Growth*, 1999, 4, 213–232.
- 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.
- , "Firm Formation with Heterogenous Management and Labor Skills," *Small Business Economics*, 1994, 6 (3), 185–191.
- Kammann, E. E. and Matthew P. Wand**, "Geoaddivitive Models," *Journal of the Royal Statistical Society C*, 2003, 52, 1–18.
- Kandala, Ngianga Bakwin, Ludwig Fahrmeier, 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 Fahrmeier**, "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.
- 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.
- Le, Anh T.**, “Empirical Studies of Self-Employment,” *Journal of Economic Surveys*, 1999, *13* (4), 381–416.
- Lewis, Arthur W.**, *Economic Development with Unlimited 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.
- McClland, David**, *The Achieving Society*, Princeton, NJ: D. Van Nostrand, 1964.
- Miller, Robert**, “Job Matching and Occupational Choice,” *Journal of Political Economy*, 1984, *December 1984* (92), 1086–120.
- 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.
- 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.
- Parker, Simon C.**, *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.
- Pratap, Sangeeta and Erwaan Quintin**, “Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach,” *European Economic Review*, 2006, *50*, 1817–1841.

- 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.
- 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.
- 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.
- Taylor, Mark P.**, “Earnings, Independence or Unemployment: Why Become Self-Employed?,” *Oxford Bulletin of Economics and Statistics*, 1996, 58, 253–266.
- Torrini, R.**, “Cross-country differences in self-employment rates: the role of institutions,” *Labour Economics*, 2005, 12 (5), 661–683.
- van der Sluis, Justin, Mirjam van Praag, and Wim Vijverberg**, “Entrepreneurship Selection and Performance: A Meta-Analysis of the Impact of Education 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.
- White, Lawrence J.**, “The Determinants of the Relative Importance of Small Business,” *Review of Economics and Statistics*, 1982, 64, 42–49.

**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 Marilyn B. Brewer**, “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises,” *American Economic Review*, 1998, 88, 290–306.



Table 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	

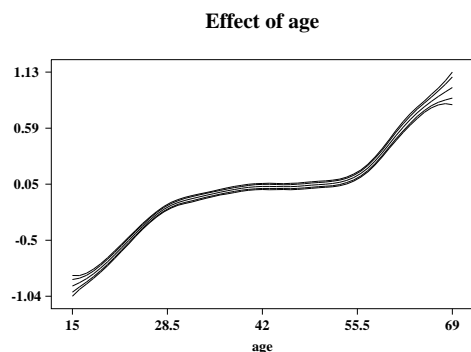


Table 2: Agricultural and Nonagricultural Sectors (Descriptives)

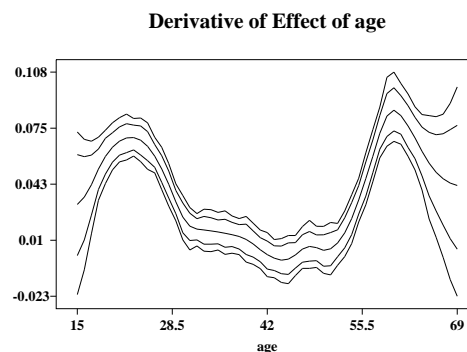
Variable	All	Non-Agri	Agri
Selfemployed	0.430	0.341	0.564
Age	37.130	34.910	40.464
Age (Std. Dev)	12.88	12.08	13.31
Male	0.809	0.837	0.767
Female	0.191	0.163	0.233
Unmarried	0.210	0.278	0.108
Married	0.745	0.690	0.828
Divorced	0.045	0.032	0.064
No Education	0.256	0.135	0.437
Informal Education	0.085	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 3: Sector Specific Self-employment (Descriptives)

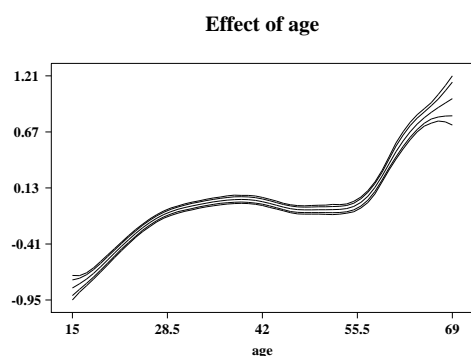
	Nonagri Employee	Nonagri Self-employed	Agri Employee	Agri Self-employed
Age	33.304	38.015	35.542	44.266
Age (Std. Dev)	11.712	12.189	12.639	12.552
Male	0.812	0.885	0.634	0.870
Female	0.188	0.115	0.366	0.130
Unmarried	0.335	0.168	0.168	0.062
Married	0.636	0.795	0.772	0.871
Divorced	0.029	0.037	0.060	0.068
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



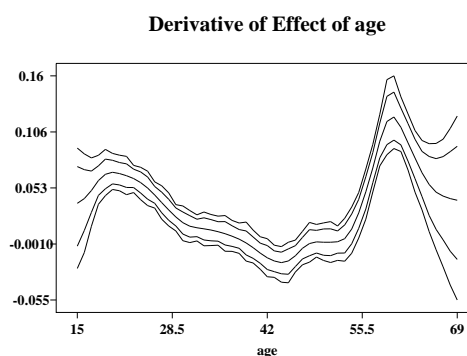
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in the Aggregate Model.



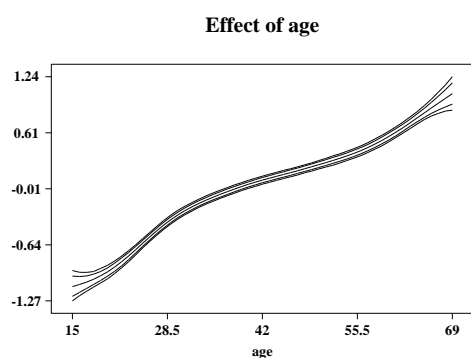
(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in the Aggregate Model.



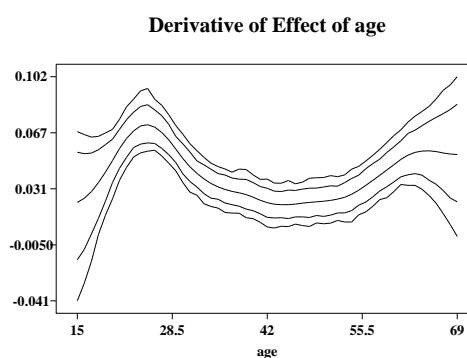
(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture.



(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 1: Non-linear Effects of Age on Self-employment

Table 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 5: Determinants of Self-employment in Nonagriculture

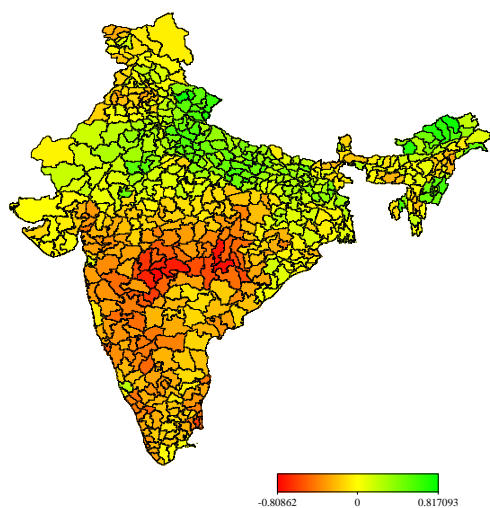
Variable	Mean	Std. Dev.	2.5%-Qt.	97.5%-Qt.
Personal Characteristics				
Female	-0.256	0.018	-0.290	-0.221
Married	0.203	0.019	0.165	0.240
Divorced	0.218	0.042	0.137	0.298
General Education				
Informal	0.141	0.028	0.085	0.195
Primary School	0.130	0.021	0.086	0.169
High School	-0.039	0.022	-0.078	0.004
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.

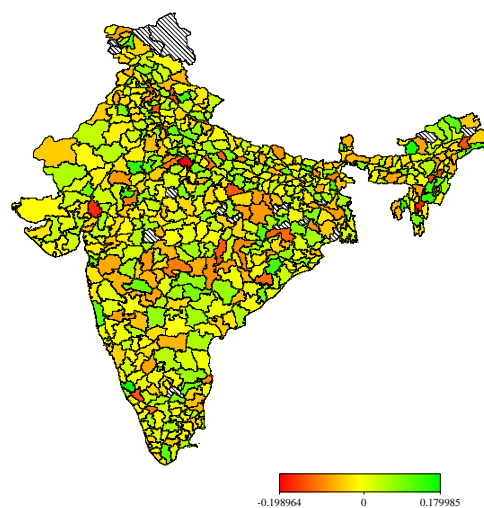
Table 6: Determinants of Self-employment in Agriculture

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			

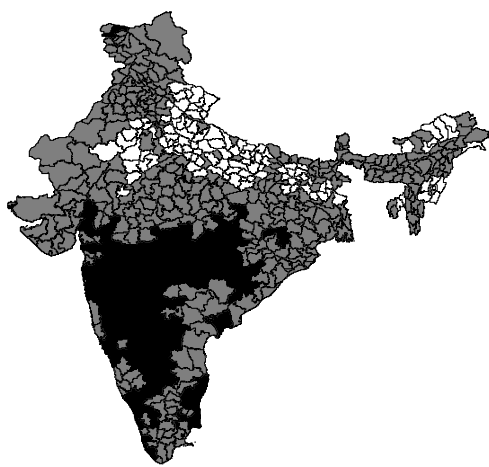
Notes: Dependent variable is binary self-employment status of the individual. Base categories for marital status, general education, technical education, land dummies are unmarried, no general education, no technical education and less than 0.2 hectares of land respectively.



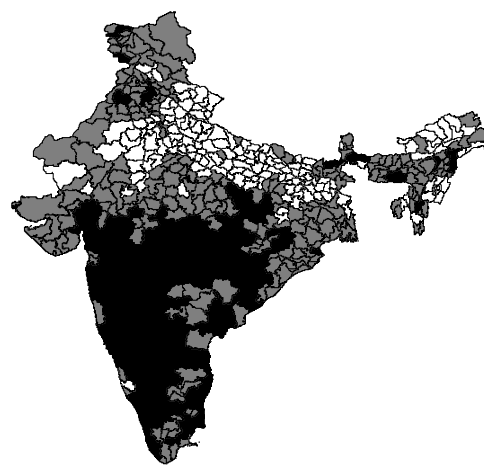
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



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

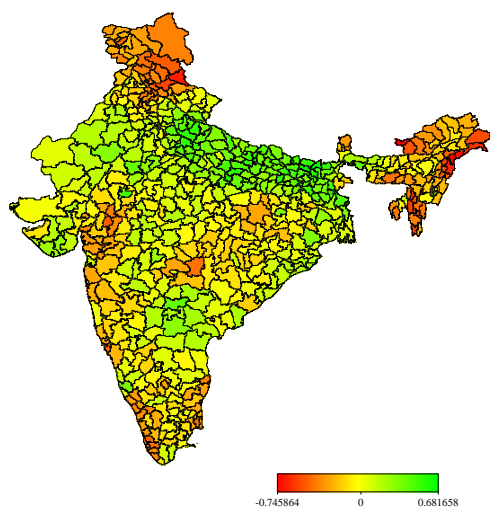


(c) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

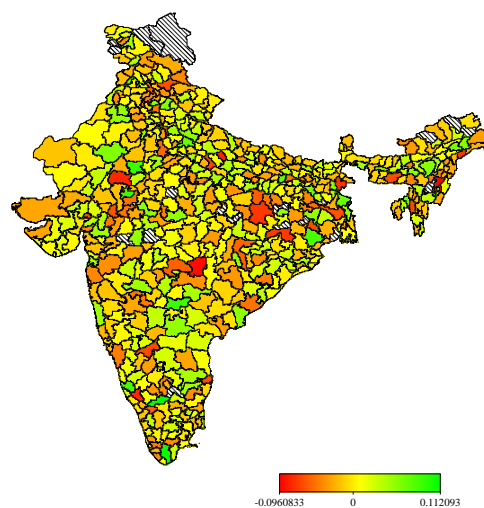


(d) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

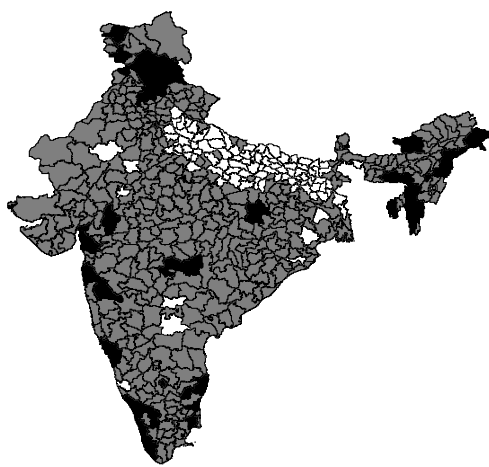
Figure 2: Spatial Effects on Self-employment Choice



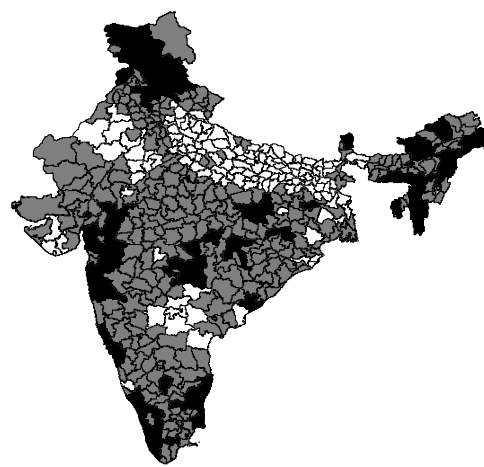
(a) Structured Non-linear Effect of 'District'. Shown are the posterior means.



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



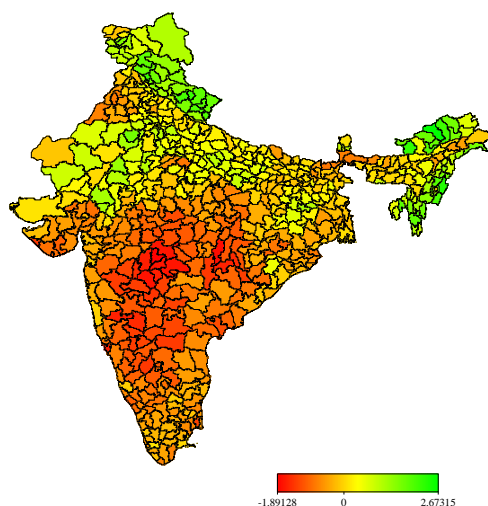
(c) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



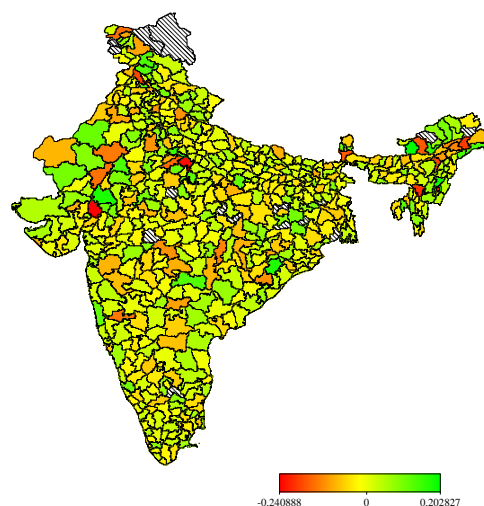
(d) Non-linear Effect of 'District'. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

Figure 3: Spatial Effects in 'Nonagriculture'

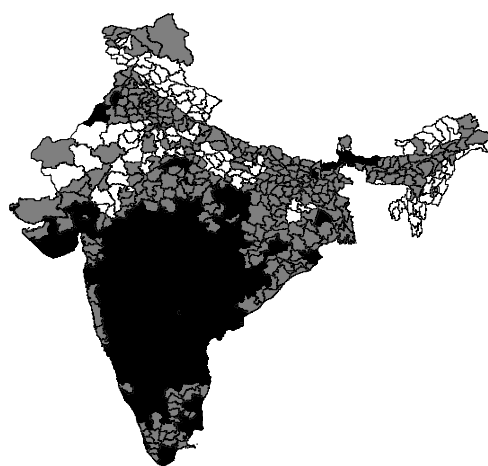




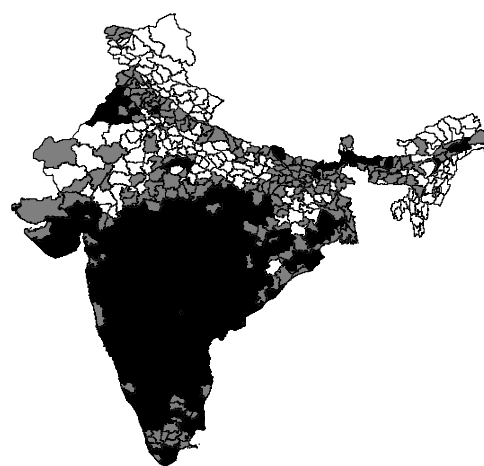
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(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 4: Spatial Effects in 'Agriculture'

Table 7: Determinants of Spatial Patterns in Figure 2, Figure 3 and Figure 4

	All		Nonagriculture	Agriculture
Financial Development				
Per-capita Credit	0.00622 (0.016)	-0.0183 (0.012)	0.0275 (0.044)	-0.402*** (0.061)
Credit-Deposit Ratio	-0.102*** (0.023)	0.0436** (0.018)		
Economic Development				
Per-Capita NSDP	-0.310*** (0.034)	-0.291*** (0.026)	-0.325*** (0.023)	-0.418*** (0.091)
Unemployment	-0.0603*** (0.016)	0.0406*** (0.012)	0.0369*** (0.012)	-0.239*** (0.040)
Demographics				
Mid Size District	0.00325 (0.030)	0.0869*** (0.022)	0.0819*** (0.022)	-0.147* (0.076)
Large District	0.0280 (0.091)	0.0750 (0.068)	0.0719 (0.068)	-0.161 (0.23)
Population Density	-0.0189 (0.015)	0.0594*** (0.011)	0.0554*** (0.011)	-0.129*** (0.037)
Agglomeration Index				
Firm Density	-0.00767 (0.013)	-0.00290 (0.0094)	-0.00874 (0.0090)	0.00453 (0.031)
Constant	2.926*** (0.35)	2.534*** (0.36)	2.749*** (0.27)	2.778*** (0.93)
Observations	534	531	531	532
$R^2$	0.20	0.23	0.40	0.22
F	19.08	22.46	49.43	21.42
$R^2$ Adjusted	0.192	0.220	0.390	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 geoaddivitive models.

Table 8: Determinants of 95% Spatial Patterns in Figure 2, Figure 3 and Figure 4

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

Table 9: Determinants of 95% Spatial Patterns in Figure 2, Figure 3 and Figure 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*** (0.037)	0.00945 (0.043)	-0.152*** (0.036)	-0.0790*** (0.030)	0.100*** (0.039)	-0.0212 (0.030)	0.233*** (0.046)	-0.0656 (0.047)	-0.168*** (0.038)
Economic Development									
Per-Capita NSDP	0.368*** (0.050)	-0.0897 (0.055)	-0.278*** (0.047)	0.316*** (0.039)	-0.0333 (0.053)	-0.283*** (0.041)	0.225*** (0.057)	-0.115** (0.057)	-0.110** (0.047)
Unemployment	0.0232 (0.024)	0.0159 (0.028)	-0.0390* (0.023)	-0.0500*** (0.017)	0.0154 (0.025)	0.0345 (0.022)	0.124*** (0.030)	-0.0183 (0.031)	-0.106*** (0.023)
Demographics									
Mid Size District	0.0707 (0.049)	-0.177*** (0.052)	0.106** (0.049)	-0.0316 (0.033)	-0.153*** (0.052)	0.185*** (0.049)	0.143** (0.056)	-0.157*** (0.053)	0.0138 (0.046)
Large District	0.00133 (0.14)	-0.211 (0.15)	0.210 (0.19)	-0.0152 (0.079)	-0.217 (0.20)	0.232 (0.21)	0.441*** (0.13)	-0.321*** (0.089)	-0.119 (0.092)
Population Density	-0.0160 (0.022)	0.0364 (0.025)	-0.0204 (0.022)	0.00542 (0.017)	-0.0591** (0.023)	0.0537*** (0.018)	-0.0731*** (0.028)	0.157*** (0.031)	-0.0841*** (0.023)
Agglomeration Index									
Firm Density	0.0198 (0.018)	-0.0350* (0.021)	0.0152 (0.018)	0.0223* (0.014)	-0.0250 (0.018)	0.00264 (0.013)	0.0113 (0.021)	-0.0313 (0.022)	0.0200 (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 geoaddivitive models.

Table 10: Self-employment in Nonagriculture

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