The Dynamics of Self-employment in a Developing Country: Evidence from India

Tamvada, Jagannadha Pawan
Max Planck Institute of Economics

January 2010

The Dynamics of Self-employment in a Developing Country: Evidence from India

Jagannadha Pawan Tamvada*

Abstract

We examine the spatio-temporal dynamics of self-employment in India using geoadditive models and pseudo panel techniques. We test the claim of Iyigun and Owen (1999) that individuals invest in professional human capital and not in entrepreneurial human capital as an economy develops. The results suggest that in non-agriculture, higher education decreases the likelihood of individuals choosing self-employment over time; however, it has an opposite effect in agriculture. While increases in land possessed increase the likelihood of self-employment choice in agriculture, individuals with small land holdings are more likely to transition into self-employment in non-agriculture. Belonging to a backward class has a negative effect on self-employment choice in both sectors; however, the effect has increased in non-agriculture and remained stable in agriculture. The geoadditive models suggest that the propensity to be self-employed has decreased across most spatial units, although there are few pockets where self-employment is rising again.

JEL Codes: J24, J43, J44, L26.

Keywords: Entrepreneurship, Self-employment, Developing Countries, Dynamics, Pseudo Panels

*Max Planck Institute of Economics. `tamvada@econ.mpg.de`

1 Introduction

This paper examines the evolution of occupational choice over space and time in India. The years after liberalization in early nineties have unleashed many economic opportunities in the Indian economy. There are compelling reasons to assume that the dynamic economic environment is influencing the occupational behavior of people.

A compelling body of theoretical literature suggests that self-employment and economic development are related inversely. Lucas (1978) predicts that entrepreneurship decreases with economic development.[1] Iyigun and Owen (1999) show that as an economy develops, individuals invest time in accumulating professional skills through education than accumulating entrepreneurial human capital. However, these theoretical results have yet to be empirically validated using micro-databases. Further, micro analyses of self-employment dynamics in developing countries are rarely found in the literature.[2]

We make several important contributions. First, we provide empirical support to a vast theoretical literature on the evolution of self-employment. In particular, we test the claim of Iyigun and Owen (1999) that in the early stages of economic development individuals invest in professional human capital and not in entrepreneurial abilities. Second, we contribute to an emerging body of literature on self-employment in developing countries (Mohapatra et al., 2007; Audretsch, Bönte and Tamvada, 2007; Tamvada, 2008) by examining the dynamics of self-employment activity over space and time. Third, we examine the dynamics of self-employment activity in India, a geographic region that has received widespread attention for its rapid economic progress, but has surprisingly little literature on self-employment and entrepreneurship. Finally, we examine self-employment dynamics in both agriculture and non-agriculture sectors. Agricultural self-employment

[1] Recent empirical studies support the view that the per-captia Gross National Product (GNP) is negatively related to the self-employment rates (Acs, Audretsch and Evans, 1994; Fölster, 2002).

[2] There are a few exceptions. See, for instance, Mohapatra, Rozelle and Goodhue (2007) for an analysis of the changing nature of self-employment in China.

is ignored by most earlier studies on self-employment.

The spatio-temporal dynamics of self-employment are analyzed using two methodological approaches. The spatial dynamics of self-employment are examined by estimating Bayesian semi-parametric geoadditive models for large scale databases collected by the NSSO in the years 2000 and 2004. In the absence of genuine panel data, the dynamics of self-employment over time are studied by constructing pseudo panels and tracking the self-employment dynamics in cohorts of individuals. The pseudo panel analyses use three cross-sectional databases collected by the NSSO during 1994-1995, 1999-2000, and 2004.[3]

The main results of both the empirical approaches are broadly consistent with each other. In agriculture, higher education has a positive impact and increases the likelihood of individuals choosing self-employment over time. However, higher education decreases the likelihood of individuals choosing self-employment in non-agriculture. While increase in land possessed increases the likelihood of self-employment choice in agriculture, individuals with small land holdings are more likely to transition into self-employment in non-agriculture. Belonging to a backward class had a negative effect on self-employment choice in both sectors; however, while the effect remained stable in agriculture, it increased in non-agriculture. Individuals in urban locations have become more likely to be self-employed over time in non-agriculture. The geoadditive models suggest that the propensity to be self-employed has reduced across most spatial units, although there are few pockets where self-employment is rising again.

The paper is structured as follows. In the next section, the datasets used in the analysis are described. The third section presents the empirical findings relating to self-employment choice over space and time. The final section summarizes the main results and discusses the limitations of the analysis.

[3]Both methodologies are described in the appendix.

2 Theoretical Background

Lucas (1978) predicts that entrepreneurship decreases with economic development. Recent studies, however, conjecture a U-shaped relationship between economic development and entrepreneurship (Wennekers and Thurik, 1999; Wennekers, Stel, Thurik and Reynolds, 2005).[4] In order to test these predictions, we examine the spatial dynamics of self-employment choice.

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 begin to invest more time in accumulating professional skills by way of education than in 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 . . . 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." We test the claim of Iyigun and Owen (1999) by tracking the occupational behavior of individuals with greater levels of human capital over time.

In addition to the role of space, we focus on the role of three important characteristics that are found to influence self-employment choice - education, wealth and ethnic or racial background. A vast literature examines the role of these factors in shaping self-employment choice. With respect to the role of education on self-employment, there is

[4] Empirical studies support this hypothesis. For instance, Acs et al. (1994) argue that self-employment increases at later stages of development when the importance of service sector increases.

little consensus in the literature.[5] Although education expands an individual's knowledge base and increases exposure to new opportunities, education also increases the opportunity cost of being self-employed. This suggests that returns to salaried employment increase faster than returns to entrepreneurship as per-capita income grows with the result being that individuals have "more to lose" by engaging in entrepreneurship (Lucas, 1978). Thus, there are compelling reasons to posit that individuals who are more educated will opt for salaried employment instead of self-employment over time (see Sluis, van Praag and Vijverberg, 2005, for a survey).

Another determinant of self-employment discussed in the literature is wealth. Wealthier individuals have more of a "safety net" when embarking on a new venture than their less wealthy counterparts. Wealth in and of itself can make financing self-employment possible, but it also makes it easier to obtain credit. Households with very high levels of wealth have a higher propensity to take risk (Carroll, 2000). Blanchflower and Oswald (1998) find that inheritance increases the probability of self-employment. Banerjee and Neuman (1993) argue that wealth distribution determines occupational structure. For these reasons, we hypothesize that individuals whose wealth increases are more likely to enter self-employment over time.

Audretsch et al. (2007) show that individuals in backward classes are less likely to enter self-employment in the Indian context. We hypothesize that the tacit restrictions on occupational choice of individuals in the Indian society may have become less binding over time. We control for a number of other variables that are likely to influence occupational choice. Most empirical evidence suggests a positive relationship between age and entrepreneurship (Evans and Leighton, 1989a; Blanchflower and Meyer, 1994; Blanchflower, 2000). Married individuals are more likely to be self-employed than are

---

[5] For example, 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, Oswald and Stutzer (2001) find negative effects of education on the probability of selecting self-employment.

their nonmarried counterparts (Borjas, 1986; Blanchflower and Oswald, 1998). In general, male, married, and older individuals are more likely to be self-employed.[6]

3 Data

For the spatial analysis, the 55th round (collected during 1999-2000) and 60th round (collected in 2004) of the employment-unemployment surveys of the National Sample Survey Organization (NSSO) of India are used.

Individuals who have reported their principal economic activity to be self-employment (including own account workers and employers), salaried employment, casual labor, or unemployment are included in the sample.[7] We restrict the sample to those who are older than 15 years but younger than 70 years. We exclude from our analysis family members who assist household enterprises, children and the elderly, and people classified into other miscellaneous occupational categories. For the year 2000, the sample consists of 169,147 individuals, and for the year 2004, the sample consists of 88,623 individuals.

For the pseudo panel analysis, the 50th round (collected during 1994-1995), the 55th round and 60th round employment-unemployment surveys of the NSSO are used. Pseudo panels are constructed using cohorts of men based on 5 year age bands interacted with the state of their residence. Women are not considered as only 10%-14% of women in the databases are economically active and many female cohorts based on the 5 year bands and states have very few observations. Furthermore, only those cohorts that have at least 500 observations in each of the surveys are considered for the analysis, for the asymptotic

[6] See Le (1999); Parker (2004) for surveys of empirical studies on self-employment.

[7] The principal economic activity alone is considered in the analysis for two reasons. First, not all individuals are engaged in subsidiary activities. Second, as less than one-sixth of the entire sample is engaged in subsidiary activities, considering such activities would further complicate the analysis when individuals report being both self-employed and paid employees. Furthermore, the principal economic activity is the activity to which the individuals devote most of their time.

reasons described in Appendix-II.[8]

4 Empirical Results

4.1 Self-employment over Space

Geoadditive models are estimated to examine the spatial dynamics of self-employment. The geoadditive methodology allows for simultaneous estimation of non-linear effects of the continuous variables such as age and spatial location as well as discrete variables such as gender and marital status on the probability of self-employment. The methodology is described in Appendix-I.

The following geoadditive models are estimated for agriculture and non-agriculture sectors separately for the years 2000 and 2004:

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

The dependent variable is a binary variable taking value 1 if an individual is self-employed. Estimation of the geoadditive regression models gives results for the parametric part of the above equation consisting of gender, marital status, education, urban location, wealth, and class structure variables as well as the non-parametric part of the equation consisting of the age effects and the spatial effects. The estimation results of the geoadditive models for the non-agriculture sector are given in Figures 1, 2 and Table 3. The estimation results of the geoadditive model for the agriculture sector are given in Figures 3, 4 and Table 4. The estimation of the empirical models reveals consistency in the results of the models for the years 2000 and 2004.

The maps in Figure 2 and Figure 4 can be interpreted in the following manner. A

[8]The number of states and districts has increased in India in the decade following the collection of the 1994-95 survey. Hence, the databases are harmonized based on the geographic boundaries in the year 1995 for the pseudo panel analysis.

white colored area suggests that the region has a positive and significant impact on self-employment, black color suggests that the spatial location has a significantly negative impact on self-employment, and grey color suggests that the impact of spatial location on self-employment is insignificant.

Figure 2 suggests that in nonagriculture, spatial effects have remained stable, although in contrast to the year 2000, the effects have become positive and significant some districts of Andhra Pradesh and Rajasthan, and negative and significant in some districts of Uttar Pradesh and Bihar in 2004. The spatial patterns in Figure 4 show that the propensity to become self-employed has decreased in the agriculture. As the plots in Figure 4 suggest, some districts in the central part of India have turned black over the four year time period, and many districts of Uttarpradesh, Haryana and Rajasthan in the north have turned grey in the 2004 maps, indicating that the propensity to be self-employed has decreased in these regions.

In non-agriculture, age has a non-linear effect as seen in Figure 1, with probability of self-employment increasing at decreasing rates until the age of 55 and increasing remarkably thereafter.[9] This may be attributable to the important role that retirement effects play. As Figure 3 shows the age effect in agriculture is close to being linear.

Table 3 shows the estimates for non-agriculture. Although education reduces the probability of self-employment in the year 2004, it has a significantly positive effect in the year 2000. This finding supports the arguments of Lucas (1978) and Iyigun and Owen (1999) that people move from self-employment to paid employment as economy develops. However, Table 4 suggests that in the agricultural sector, education increases

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

the probability of self-employment in both years. Thus, a startling result of the analysis is that while educated people were more likely to be self-employed in non-agriculture in the year 2000, they were less likely to be self-employed in 2004. Hindus and members of backward classes have a lower likelihood to be self-employed in year 2000 as well as 2004. The estimates of other variables such as gender and marital status are remarkably similar.

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

In contrast to non-agriculture, higher education has a positive impact on self-employment in both the years in agriculture. Similarly, urban location has a positive effect in both the years. In contrast to non-agriculture, religion has no significant effect in agriculture, although belonging to a backward class has a much stronger negative effect in both the years.

4.2 Pseudo Panel Analysis

The econometrics of pesudo panels are summarized in Appendix-II. The estimation results of pseudo panel regression are presented in Table 5.[10] The cohorts are constructed on five year bands from 15 years to 70 years. These five year bands are interacted with the state regions to define cohorts. In the first column, estimation is based on cohorts of all

[10] The standard OLS model is likely to be biased as the F test in all the fixed effects regressions is significant, suggesting that cohort fixed effects are correlated with the exogenous variables.

men in the sample.[11] For the estimation in the second column, we construct a psuedo panel of men working in non-agriculture, to analyze more homogenous cohorts. In the third estimation, we similarly construct a psuedo panel of men working in agriculture.

The average age of the cohort, proportion of married individuals, proportion of people in urban regions, along with a series of variables that indicate the proportion of individuals in each education category are introduced in the estimation. The religion and caste variables are included, as these have been found to play an important role in determining self-employment choice in the Indian context (Audretsch et al., 2007).

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

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

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

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

In the third column, the analysis is done on a pseudo panel of cohorts in agriculture. In contrast to results of the non-agriculture estimation, it is seen that education has a significant positive effect on the transition into self-employment in agriculture. This is consistent with the results of the repeated cross-sectional analysis in Table 4, where coefficients of the education variables are higher in the year 2004. The effect of education on self-employment is highest for those with informal education. This effect keeps decreasing as education rises but remains positive and significant. However, the coefficient of "University" is almost half the coefficient of "Informal" education, suggesting that education has a positive effect that is non-linear and decreasing. Thus, in the agricultural sector, while a 1% increase in the proportion of people with informal education increases the self-employment transition rate by 0.42%, a similar increase in the proportion of people with university education increases the self-employment rate by only 0.23%. This suggests that educated individuals who stay in agriculture choose self-employment over paid employment.

In Table 6, the land variables are also introduced in the estimation. While the coefficients of the non-agriculture estimation in the second column are similar to the coefficients of the estimation without land variables, the coefficients of education variables in the agriculture equation in the third column have shifted downward. This suggests that the education variables captured the positive effect of the land variables in the earlier estimation. While in non-agriculture equation, land is positive and significant in the lowest category, in the agriculture equation, the highest land variables are positive and significant. Thus, while small amounts of land enable individuals to enter self-employment in non-agriculture, individuals with large amounts of land choose self-employment in agriculture. Moreover, the urban variable remains insignificant in both estimations. This result is unexpected as one would predict an increase in the share of urban population in the cohort to have a positive influence on the self-employment transition rate. It is possible that rural-urban migration increases the share of people working as self-employed and share of people working as paid employees proportionately, therefore leading to an insignificant effect on movement into self-employment.

Finally, we construct a pseudo panel of cohorts of individuals working in non-agriculture and living in rural areas as well as a pseudo panel of cohorts of individuals working in non-agriculture and living in urban areas. Table 7 presents the estimation results. In the first column, the results of rural cohorts is presented. The results suggest that education has a significant negative effect, which is more pronounced at the university level. The land variables are insignificant, suggesting that in rural areas, possession of land does not lead people into self-employment. One reason could be that in rural areas, the level of financial development is comparatively low and price of the land is low. This can reinforce credit constraints for such individuals.

The results of the estimation on cohorts of urban males is presented in the second column. The results suggest that the negative effect of education on self-employment choice is also present in the urban cohorts. However, the coefficients are much smaller than coefficients of the rural estimation. In particular, the results suggest that while a 1% increase in the proportion of university educated individuals in rural areas reduces the transition into self-employment by 0.61%, an increase in the share of university educated individuals in urban areas decreases the transition rate into self-employment by 0.24%.[12] Furthermore, the informal education variable is positive and significant in the urban equation, suggesting that, ceteris paribus, an increase in the share of people with informal education by 1% in the cohort, increases the self-employment rate by 0.65%. The lowest level land variable is positive and significant, suggesting that the overall effect of land (seen in the earlier estimation with all the workers in non-agriculture) is primarily due to such an effect in urban areas. The coefficient of the scheduled castes/scheduled tribes (SC/ST) variable is also significantly negative, and the effect is higher in urban areas. Thus, it is seen that a 1% increase in the proportion of the SC/ST people in a cohort reduces the transition rate into self-employment by 0.51% in rural areas and by 0.7% in urban areas. This could be a result of government's reservation policy that sets aside public sector jobs for individuals from these castes. The concentration of such jobs in urban areas could explain this result to some extent.

4.3 Reconciling the Results

Table 1 summarizes the results of the repeated cross section analysis and the pseudo panel analysis for the agricultural sector. The first column summarizes the results for the year 2000, the second column for the year 2004 and the third column, results of the pseudo panel analysis.

The coefficient of informal education is positive in all three columns. The pseudo panel estimation supports the finding of the repeated cross sectional analysis in Table 4, that the

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

Table 1: Results for Agriculture

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

Table 2: Results for Non-agriculture

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

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

Table 2 summarizes the main results for nonagriculture. The first two columns summarize the results of the repeated cross-sectional analysis and the next three columns summarize the results of the pseudo panel analysis.

The coefficient of informal education is positive in the first three columns, suggesting that informal education is associated with an increase in the probability of self-employment in nonagriculture as well. This is supported by the results in Table 3 which show that the coefficient of informal education is higher in the year 2004. The coefficient for rural areas is insignificant in the fourth column and is positive for urban areas in the fifth column. This suggests that the positive effect is attributable to the role of informal education in urban areas, as a determinant of self-employment.

Though the coefficient of primary schooling is positive in the first two columns, the estimated effect is negative in the pseudo panel analysis. This suggests that individuals with primary education have become less likely to be self-employed over time. This result is corroborated by the lower coefficient of the primary education variable in the year 2004, relative to the coefficient for the year 2000, in the repeated cross-sectional analysis (in Table 3). The pseudo panel analysis further suggests that the negative effect over time is seen in rural as well as urban areas.

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

The effect of technical education is negative in the year 2000 and as well as in the year 2004. The pseudo panel suggests that the nature of the relationship has not changed over time, at least in the case of technical degree. For the technical diploma variable, the effect is negative in all columns except in the rural pseudo panel. This suggests that in urban areas, individuals with technical diplomas have become less likely over time to be self-employed. The results of the repeated cross-sectional analysis in Table 3 show that the absolute value of the coefficient of technical diploma variable is smaller in 2004. This is captured by the negative sign of the coefficient estimated by the pseudo panel analysis.

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

The signs of the cultural context variables, Hindu and SC/ST suggest that the relationship has remained stable over time. While Hindus were less likely to be self-employed in the years 2000 and 2004, the insignificant coefficient in the pseudo panel analysis suggests that this relationship has remained stable over time. This is corroborated by the repeated cross-section analysis in Table 3, where the coefficient of the Hindu variable is almost equal in both years. However, the SC/ST variable is negative throughout, indicating that individuals in these castes have become less likely over time to be self-employed. This result is supported by an increase in the absolute value of the coefficients of this variable in the repeated cross sectional analysis in Table 3.

5 Conclusion

This paper uses two different empirical methods to analyze the spatio-temporal dynamics of entrepreneurship in India. The repeated cross section analysis for the years 2000 and

2004 using geoadditive models suggests that self-employment propensity has reduced across most spatial units, although there are some exceptions.

Using three different cross-sectional databases collected over 1994-2004, we also constructed pseudo panels of individuals based on 5 year bands of birth cohorts, regions and sectors. The pseudo panel analysis tracks the dynamics of self-employment over a longer period of time. The results suggest that, in agriculture, higher levels of education have a positive impact and increase the likelihood of individuals choosing self-employment over time. However, higher education decreases the likelihood of individuals choosing self-employment in non-agriculture over time. This result supports the claim of Iyigun and Owen (1999) that as an economy develops individuals prefer to invest in professional human capital instead of entrepreneurial human capital. Further, while increase in land possessed increases the likelihood of self-employment choice in agriculture, individuals with small land holdings are more likely to transition into self-employment in non-agriculture. Belonging to a backward class has a negative effect on self-employment choice in both sectors; however, while the effect remained stable in agriculture, it increased in non-agriculture. The main results of both empirical approaches are broadly consistent with each other.

In summary, the nature of self-employment is experiencing significant changes in India. In non-agriculture, individuals who acquire higher levels of human capital are shifting to wage employment. This may be attributable to the increasing number of economic opportunities in this sector. In contrast, in agriculture, individuals with higher levels of human capital are shifting to self-employment; however, the relative proportion of such individuals is low in agriculture. The spatial analysis suggests that the propensity to become self-employed has decreased across the country, although there are some pockets where it is re-emerging.

Appendix-I

Semiparametric regression techniques based on Bayesian P-Splines and geoadditive models are used for comparing two cross-sections of NSSO data collected in the years 2000 and 2004. Th methodology is extensively discussed in the appendix. In addition, we also employ the within estimator on a pseudo panel constructed from three sample surveys. The pseudo-panel method is discussed in section 5.

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 nonlinear effects of the continuous variables, such as age and spatial location, as well as individual characteristics, such as educational attainment, gender and marital status, on the probability of self-employment. As mentioned earlier, one of the main advantages of this method is that it allows simultaneous estimation of the effects of individual characteristics and spatial location on 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 and allow *ceteris paribus* interpretation. Furthermore, this method enables visual inspection of the estimated self-employment neighborhoods. A brief outline of the method is presented below.[13]

Geoadditive Models

Let (y_i, x_i, v_i) for i in {1,2,...N} describe a dataset of N observations. Let y_i be the response variable and x_i be a m-dimensional vector of continuous covariates and v_i be a vector of categorical variables.[14] Assume y_i are independent and Gaussian with mean $\eta_i = f_1(x_{i1}) + ... + f_p(x_{ip}) + v_i\gamma$, and a common variance σ^2. If f_i are unknown smooth functions of the continuous variables and $v_i\gamma$ corresponds to the parametric part of the regression, the regression model is called an 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.[15] Furthermore, Lang

[13]This section draws on Lang and Brezger (2004) and Brezger and Lang (2005).

[14]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.

[15]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.[16] These methods are used for the empirical analysis.

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

$$L = l(y, \beta_1, ..., \beta_p, \gamma) - \lambda_1\Sigma(\triangle^k\beta_1)^2 - ....\lambda_p\Sigma(\triangle^k\beta_p)^2 \tag{1}$$

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

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

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

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

Inference is based on the posterior and employs recent Monte Carlo Markov Chain

[16] 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).

(MCMC) techniques. If α is a vector of the unknown parameters, assuming conditional independence of the parameters, the posterior is given by:

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

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

Model Diagnostics

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

Appendix-II

Pseudo Panels

In the absence of genuine panel data, repeated cross-sectional data can be used to construct synthetic or pseudo panels, as suggested by the seminal paper of Deaton (1985). While cross-sectional data are collected over many years in developing countries, genuine panel data are very rare. A pseudo panel based on, for instance, age cohorts, gender, or education levels can be used to control for at least cohort fixed level effects. Such methods are similar to instrumental variable methods where group dummies are used as instruments.[17]

Consider the following linear model with individual effects,

$$y_{it} = x_{it}\beta + \alpha_i + e_{it}, \tag{4}$$

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

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

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

$$\overline{y}_{ct} = \overline{x}_{ct}\beta + \overline{\alpha}_c + \overline{e}_{ct}, \tag{5}$$

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

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

$$\hat{\beta}_W = (\sum_{c=1}^{C}\sum_{t=1}^{T}(\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)')^{-1}\sum_{c=1}^{C}\sum_{t=1}^{T}(\overline{x}_{ct} - \overline{x}_c)(\overline{y}_{ct} - \overline{y}_c) \tag{6}$$

As described in Verbeek (2006), the asymptotic behavior of pseudo panel estimators can be derived for the following alternative asymptotic sequences. First, when $N \to \infty$, with C fixed, so that $n_c \to \infty$. Second, when $N \to \infty$ and $C \to \infty$, with n_c fixed. Third,

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

$T \rightarrow \infty$, with N, C and n_c fixed. While Moffitt (1993) and Verbeek and Vella (2005) employ the asymptotics of the first type, Deaton (1985), Verbeek and Nijman (1993) employ the second type.

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

$$plim \frac{1}{CT} \Sigma\Sigma(\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)' \tag{7}$$

is finite and invertible and

$$plim \frac{1}{CT} \Sigma\Sigma(\overline{x}_{ct} - \overline{x}_c)\alpha_{ct} = 0 \tag{8}$$

As $n_c \rightarrow \infty$ the above conditions are automatically satisfied as the cohort fixed effects converge to a constant over time, that is, $\alpha_{ct} \rightarrow \alpha_c$ (Moffitt, 1993).[18] Deaton (1985) relies on asymptotics of the first type and does away with the necessity to have large numbers of observations in each cohort. This is achieved by considering the cohort averages as error-ridden measurements of the population averages of the cohorts. By assuming that measurement errors are distributed with zero mean, the moment matrices of the within estimator are adjusted to correct for the measurement error. McKenzie (2004) shows that when cohorts are based on age groups, the asymptotics of the second type seldom get satisfied, as the number of cohorts is fixed. For this reason, as the cohort sizes in the sample are very large, we assume asymptotics of the first type, to consistently estimate β.

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

References

Acs, Zoltan J., David B. Audretsch, and David S. Evans, "Why does Self Employment Rate vary Across Countries and Over Time," CEPR Discussion Paper No. 871 1994.

Audretsch, David B., Werner Bönte, and Jagannadha Pawan Tamvada, "Religion and Entrepreneurship," CEPR Discussion Paper no. 6378. Center for Economic Policy Research, London 2007.

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

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.

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.

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

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

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

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

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 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.

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

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

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

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

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

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.

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.

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

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

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

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

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

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

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.

Tamvada, Jagannadha Pawan, "The Geography and Determinants of Entrepreneurship," Discussion Paper No. 2008-026, Jena Economic Research Paper Series, Max Planck Institute of Economics and University of Jena 2008.

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

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

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

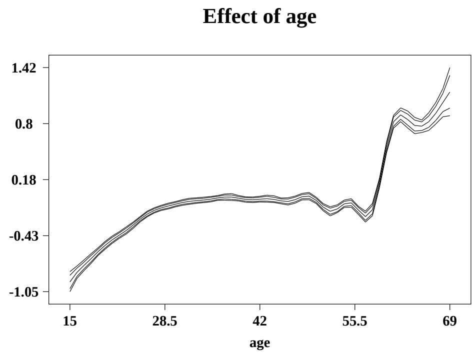
_ **and Francis Vella**, "Estimating Dynamic Models from Repeated Cross-Sections," *Journal of Econometrics*, 2005, *127*, 83–102.

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

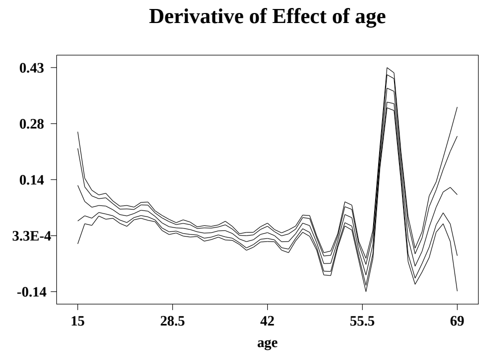
_ **and** _ , "Minimum MSE Estimation of a Regression Model with Fixed Effects and a Series of Cross Sections," *Journal of Econometrics*, 1993, *59*, 125–136.

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

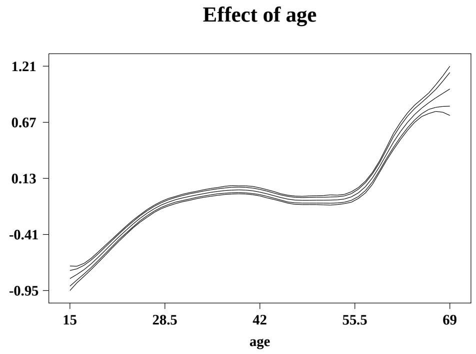
_ **, Andre van Stel, Roy Thurik, and Paul Reynolds**, "Nascent Entrepreneurship and Level of Economic Development," *Small Business Economics*, 2005, *24*, 293–309.



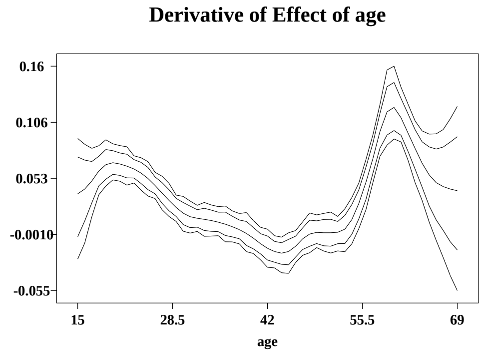
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture in 2000.



(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Nonagriculture in 2000.



(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Nonagriculture in 2004.

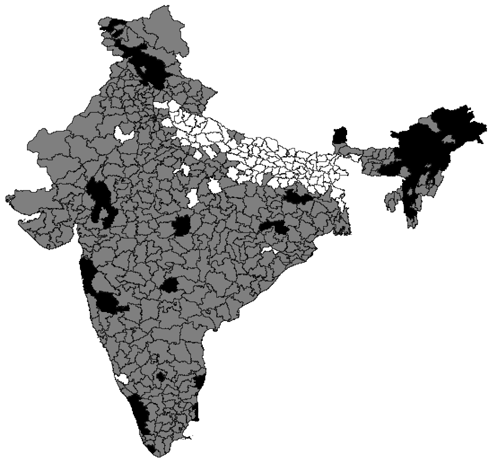


(d) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Nonagriculture in 2004.

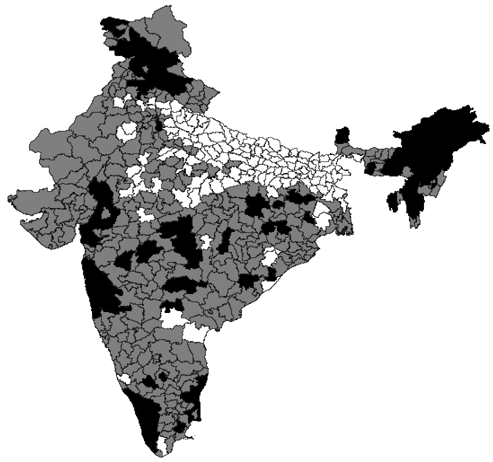
Figure 1: Non-linear Effect of Age on Self-employment in Non-agriculture

Table 3: Determinants of Self-employment (Nonagriculture)

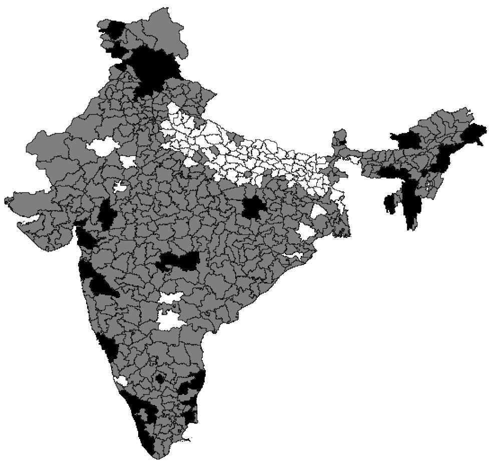
	2000		2004	
Variable	Coefficient	Std. Dev.	Coefficient	Std. Dev.
Personal Characteristics				
Female	-0.196	0.014	-0.256	0.018
Married	0.257	0.015	0.203	0.019
Divorced	0.256	0.027	0.218	0.042
General Education				
Informal	0.209	0.078	0.141	0.028
Primary School	0.259	0.016	0.130	0.021
High School	0.244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



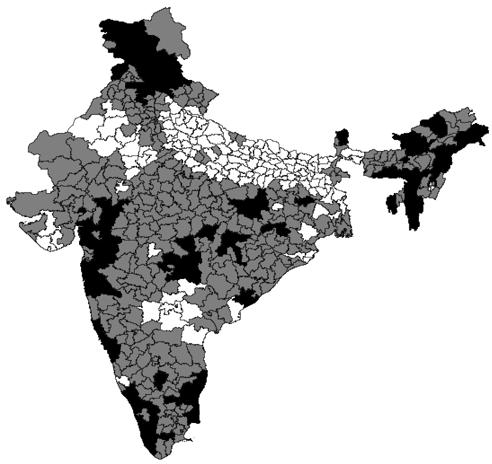
(a) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



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



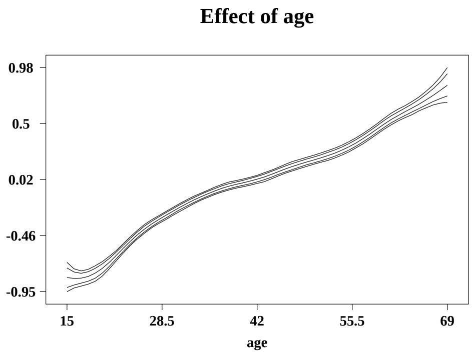
(c) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



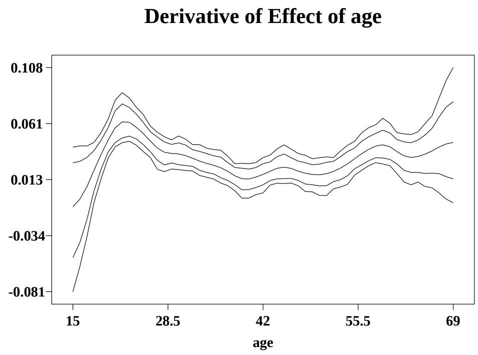
(d) Non–linear Effect of 'District' in 2004. Posterior probabilities for a nominal level of 80%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.

Notes: The maps are plotted for areas that are statistically estimable and do not show political boundaries of India.

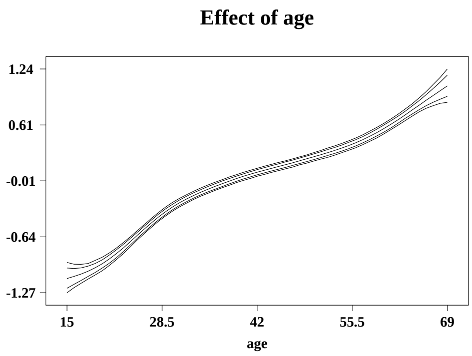
Figure 2: Spatial Effects in 'Non-agriculture'



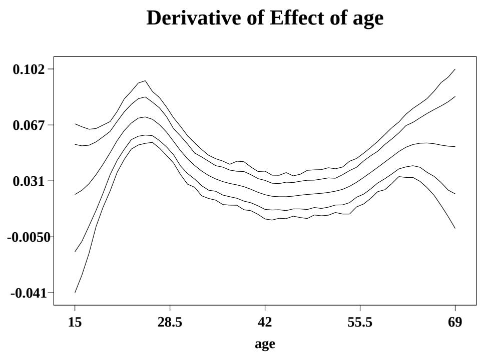
(a) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Agriculture in 2000.



(b) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Agriculture in 2000.



(c) Posterior mean of the non-linear effect of 'age' together with 95% and 80% pointwise credible intervals in Agriculture in 2004.

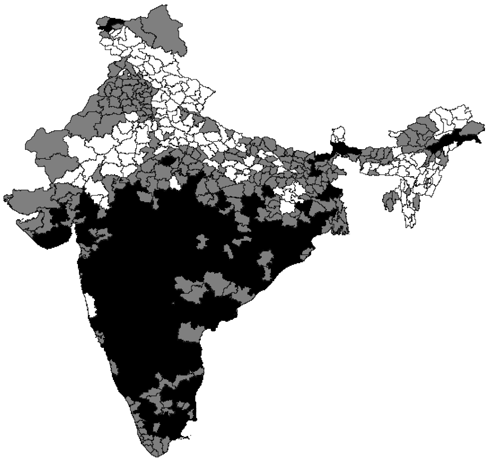


(d) Derivative of the posterior mean of the 'age' function with 95% and 80% pointwise credible intervals in Agriculture in 2004.

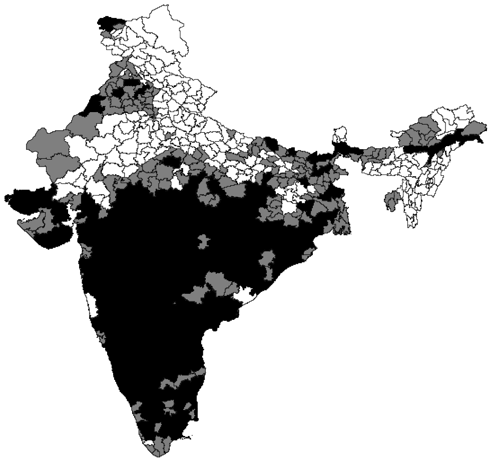
Figure 3: Non-linear Effect of Age on Self-employment in Agriculture

Table 4: Determinants of Selfemployment (Agriculture)

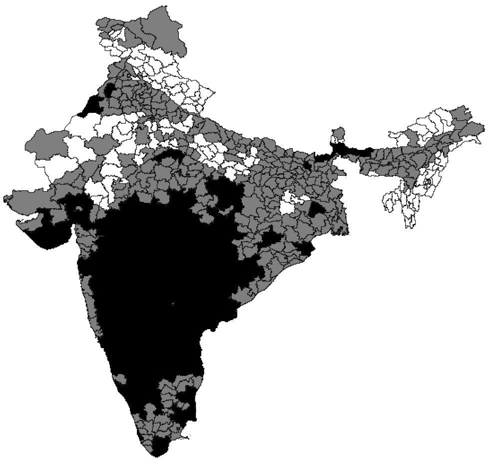
	2000		2004	
Variable	Coefficient	Std. Dev.	Coefficient	Std. Dev.
Personal Characteristics				
Female	-0.495	0.018	-0.540	0.027
Married	0.175	0.029	0.206	0.042
Divorced	0.282	0.040	0.447	0.058
General Education				
Informal	0.156	0.088	0.233	0.032
Primary School	0.244	0.021	0.435	0.025
High School	0.496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



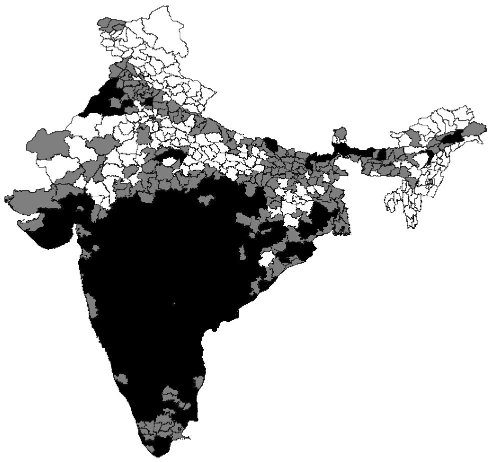
(a) Non–linear Effect of 'District' in 2000. Posterior probabilities for a nominal level of 95%. Black denotes regions with strictly negative credible intervals, white denotes regions with strictly positive credible intervals.



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



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



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

Notes: The maps are plotted for areas that are statistically estimable and do not show political boundaries of India.

Figure 4: Spatial Effects in 'Agriculture'

Table 5: Pseudo Panel Estimation

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

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

Table 6: Pseudo Panel Estimation

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

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

Table 7: Pseudo Panel Estimation: Rural-Urban Areas (Nonagriculture)

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

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