The Distribution of Firm Start-Up Size Across Geographic Space*

David B. Audretsch† Jagannadha Pawan Tamvada‡

Abstract

A growing body of literature shows that geographic location plays an important role in influencing economic phenomena. While the impact of firm-specific and industry-specific characteristics on size of new firms has been analyzed, the role of geographic location has been largely neglected. Using geoadditive models, we estimate geographic location as a micro-determinant of firm start-up size. The estimations based on a comprehensive database of firm start-ups in India suggest that the size of new firm start-ups exhibits remarkable spatial patterns that are not explainable by firm and industry characteristics. The results reveal a clear north-south divide in the size of new firm start-ups in India and provide important first insights identifying those factors shaping startup size in the context of a developing economy. Further, an investigation into the factors that shape these residual spatial patterns shows that that regions that are more developed give birth to superior start-ups.

*We thank Zoltan Acs, Thomas Astebro, Werner Bönte, Alex Coad, Michael Dahl, Stephan Klasen, Mili Shrivastava and participants at the European Economic Association Conference in Milan, DRUID Conference in Copenhagen, and the Schumpeter Society Conference in Rio De Janeiro, for their comments and suggestions on an earlier version of the paper.

†Max Planck Institute of Economics, Kahlaische Strasse 10, Jena 07745, Germany. Email: audretsch@econ.mpg.de

‡Max Planck Institute of Economics, Kahlaische Strasse 10, Jena 07745, Germany. Email: tamvada@econ.mpg.de

1 Introduction

A growing body of literature examines the determinants of firm size distribution and its evolution over time (Cabral and Mata, 2003; Angelini and Generale, 2008) and a component of this literature examines the determinants of the start-up size of firms. The literature on firm start-up size finds that industry characteristics, such as the extent of scale economies (Mata and Machado, 1996; Mata, 1996), and firm specific characteristics, such as the degree of human capital of the founding entrepreneurs (Astebro and Bernhardt, 2005; Colombo, Delmastro and Grilli, 2004; Colombo and Grilli, 2005), influence the start-up size of new firms. However, despite a vibrant literature on economic geography emphasizing geographic location as an important determinant in shaping economic activity (Krugman, 1991b; Fujita and Krugman, 2003; Gilbert, 2004; Gilbert, McDougall and Audretsch, 2006), to the best of our knowledge, the role of geographic location on the size of new firm start-ups has rarely been examined.

The purpose of this paper is to examine the impact of geographic location on the distribution of firm start-up size. Using recent methodological advances in spatial econometrics and a large, comprehensive firm database consisting of small firms that registered in India from 1998-2000, we find that the size distribution of start-ups is remarkably spatially skewed and exhibits distinct spatial patterns. The results presented in this paper contribute to the growing literature on firm start-up size by highlighting that the firm start-up size is not independent of the spatial context.

The findings resulting from the analysis make several important contributions to the literature. The empirical evidence suggests that, despite being overlooked in the previous studies on startup size, the spatial location is an important micro-determinant of the size of firm start-ups. The size of new firm start-ups exhibits remarkably distinct spatial patterns after filtering firm level and industry level effects. This also suggests that the right skewed nature of firm size distribution may be attributable to the emergence of many small firms in some regions.

For India, the results reveal a clear north-south divide in the size of new firm start-ups. The results provide important first insights identifying those factors shaping startup size in the context of a developing economy. In particular, ownership structure and initial knowledge endowments are found to influence the size of new firms. Further, an investigation into the factors that shape these residual spatial patterns suggests that superior regions give birth to superior new firm start-ups.

The paper is structured as follows. The next section presents the literature on firm start-up size and the hypotheses linking geographic location with the firm start-up size. The third section presents the database. In the forth section, the geoadditive modeling technique with Bayesian inference based on Monte Carlo Markov Chain (MCMC) methods is discussed. The fifth section presents the empirical model and the results linking the region with the firm start-up size. The final section provides conclusions and possible avenues for future research.

2 Start-Up Size and Location

2.1 The role of industry and firm characteristics for Firm Start-Up Size

As Geroski (1995) and Caves (1998) summarize an extensive empirical literature, the likelihood of survival for new entrants is low, and those that do survive grow at a higher rate than the incumbents. Firms with a larger start-up size have a higher likelihood of survival (Dunne, Roberts and Samuelson, 1989; Mata, Portugal and Guimaraes, 1995). Thus, the size of a new firm at start-up has a significant impact on its post-entry performance and survival.[1]

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

The handful of empirical studies on the determinants of firm start-up size primarily examine the role played by industry characteristics such as the minimum efficient scale (MES) of the industry, industry growth, effects of operation at suboptimal scale (defined as the proportion of those employed in firms that are operating at sub-optimal scale), impact of market size (Mata and Machado, 1996; Mata, 1996; Görg, Strobl and Ruane, 2000; Görg and Strobl, 2002) and firm specific characteristics such as human capital of founders and credit constraints (Astebro and Bernhardt, 2005; Colombo et al., 2004; Colombo and Grilli, 2005).

The size of new firms tends to be greater in industries with a higher minimum efficient scale (MES), greater turbulence, and in industries where few suboptimal firms operate (Mata, 1996). As Mata and Machado (1996, p. 1321) note, "entry on a relatively large scale in each industry is much more sensitive to the minimum efficient scale and to the extent of firm turnover in the industry than entry in small scale. Put differently, it seems that small new firms appear everywhere, while relatively large ones only appear where economies of scale make it crucial, or where sunk costs are low, therefore leading to low

[1] Many empirical studies categorically reject the assumption underlying the Gibrat's Law which, in essence, states that firm growth is independent of size (Evans, 1987; Hall, 1987; Sutton, 1997; Lotti and Santarelli, 2004). An emerging body of literature that explains the evolution of FSD (Cabral and Mata, 2003) also supports the view that the relationship between firm growth and size is not constant.

losses in case of failure."[2]

The start-up size has also been found to be influenced by characteristics specific to the firm. In particular, as the age and education of the founder increase, firm start-up size has been found to also increase. Professional knowledge about the particular industry along with managerial and entrepreneurial experience have been found to have a greater positive impact than education and working experience on the size of new firms (Colombo et al., 2004).[3] Similarly, Astebro and Bernhardt (2005) find that the entrepreneurial human capital of founders co-determines their household wealth as well as their firm's start-up capital.

2.2 The role of spatial location for Firm Start-Up Size

Why should spatial location matter for the size of new firm start-ups? There are compelling reasons already established in the economics literature to suggest that start-up size may not be independent of the geographic location.

A significant body of literature has examined the factors influencing the differences in new firm formation rates across regions in developed countries (Acs and Armington, 2006). This literature has found that regional specific variables such as human capital, knowledge stock, unemployment, sector specialization, and income and population growth explain the variation in new firm formation rates across regions. However, by solely focusing on firm entry rates, this literature ignores the size distribution of new firm start-ups or assumes that the size of new firm start-ups is independent of the spatial context. By

[2] A study on Irish firms shows comparable results, but finds a negative effect of industry size and positive effect of industry growth on start-up size (Görg et al., 2000). Nurmi (2006) studies sectoral differences in size of firm start-ups in Finland and finds that results for manufacturing and service sectors are very similar. Görg and Strobl (2002) find that the presence of multinationals in an industry negatively effects the size of domestic Irish entrants. Mata and Machado (1996) analyze a sample of 1079 new firms from Portugal. In their sample, not more than 25% have greater than the average size of 17 employees, and 50% of the firms employ less than 10 people.

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

contrast, this paper tests the validity of this assumption by explicitly analyzing if the size of new firm start-ups systematically varies across space.

A new literature has emerged suggesting that knowledge, competence, and know-how spills over from the firm producing it to a different, third-party firm (Griliches, 1992). This view is supported by theoretical models which have focused on the role that spillovers of knowledge play in generating increasing returns and ultimately economic growth (Romer, 1986, n.d., 1994; Krugman, 1991a,b). An important theoretical development is that geography may provide a relevant unit of observation within which knowledge spillovers occur. The theory of localization suggests that because geographic proximity is needed to transmit knowledge and especially tacit knowledge, knowledge spillovers tend to be localized within a geographic region. The importance of geographic proximity for spillovers has been supported in a wave of recent empirical studies by (Jaffe, 1989; Jaffe, Trajtenberg and Henderson, 1993; Audretsch and Feldman, 1996; Audretsch and Stephan, 1996). Such external knowledge resources should impact the firm start-up size by enabling the new firm to attract more early stage resources in order to start up at a greater scale, i.e., size. Thus, such spatially constrained externalities will have an impact on the size of new firm start-ups, reflecting the availability of knowledge external to the start-up.

A recent literature finds that financial development influences firm size distribution (Beck, Demirguc-Kunt, Laeven and Levine, forthcoming). Guiso, Sapienza and Zingales (2004) find a positive impact of financial development on firm entry and entrepreneurship. A compelling body of literature suggests that financial constraints have an impact on the size of new firm start-ups. For example, Colombo and Grilli (2005) find that firms receiving external private equity financing have a larger size at start-up. Cabral and Mata (2003) suggest that the firm size distribution evolves from a right skewed distribution to a less skewed distribution over time as financial constraints become less binding over time.[4] As the level of financial development varies across regions and as financial constraints have an impact on the size of new firm start-ups, geographic location is likely to influence the start-up size.

3 Data

The main source of data for linking the geographic location of the firm with the start-up size is from the Ministry of Small Scale Industries in India. We use firm level data from the third census of registered small scale firms. This census was conducted in 2001. We consider manufacturing firms that have started producing in 1998, 1999 or 2000 as new start-ups for the analysis following Audretsch, Keilbach and Lehman (2006), who consider the three year period, as firm start-ups are subject to a very high degree of stochastic disturbance if only a very short period is considered. This rich dataset of entrants consists of 149,708 firms. Each new firm was asked a series of questions concerning the set of initial conditions under which it was founded (such as the original value of its plant and machinery, its year of initial production, the sector, the source of its technical knowledge, its spatial location).

As the dataset consists solely of small firms, we do not have information about large entrants. This limitation of the dataset, however, does not pose serious problems for testing our hypothesis, since the theory of firm size distribution suggests that majority of entrants are small and numerous. Furthermore, if few large entrants are also present

[4] However, by using measures of financial constraints from firm level data, Angelini and Generale (2008) find that although the FSD of constrained firms is more right skewed, the FSD of unconstrained firms is indistinguishable from the FSD of all firms. Further, they find that financial constraints are more relevant in developing countries in explaining the FSD.

in the dataset, they would at best be outliers and introduce heterogeneity.[5]

As the descriptive statistics in section 6 suggest 88.5% of the sample consists of firms that were started by proprietors. 6.8% of the firms are owned by two or more partners and are referred to as partnerships. Firms having other ownership structures such as co-operatives account for 4.7% of the sample. 15.8% of the firms are managed by women. 73.8% are small scale industrial units. Thus, 26.2% of the firms are small scale business enterprises, primarily consisting of repairing, servicing and maintenance units. More than 14% of the firms have reported that they have technical knowledge. While only 0.94% of the firms in the sample reported having obtained knowledge from sources outside India, as many as 6.6% have their technical knowledge from other firms and 6.67% from universities. 20.9% of the firms are in the industrial sub-sector of apparels manufacture and 19.2% are firms dealing with food products. With 11.67% of all the firms, the next largest group comprises of firms in the industrial sub-sector of fabricated metals.

4 Methodology: Geoadditive Models

Geoadditive models are used for estimating the spatial distribution of firms' start-up size. This technique enables the simultaneous estimation of the effects of firm and industry characteristics as well as the spatial location on the firms' start-up size. Thus, the estimated spatial patterns allow *ceteris paribus* interpretation. A brief outline of the methodology is presented here.[6]

Let (y_i, v_i) for i in {1,2,...N} describe a dataset of N observations. Let y_i be the response variable and v_i be a q-dimensional vector of categorical variables. Let y_i be independent and gaussian with mean $\eta_i = v_i\gamma$ and variance σ^2. In the Bayesian set-

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

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

up, γ_j are considered as random variables and assigned prior distributions. Independent diffuse priors are assumed for the parameters γ_j, for $j = 1....q$. To the above predictor, if a spatial effect $f_{spatial}$ is added, a geoadditive model is obtained.

The spatial effect may be split into a spatially correlated and uncorrelated effect, $f_{spatial} = f_{str} + f_{unstr} = X_{str}\beta_{str} + X_{unstr}\beta_{untr}$, as the spatial effect may comprise of a component that has strong spatial structure and a component that is only locally present. Following Besag, 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, then the MRF prior is given as,

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

for, $u \neq s$, where, N_s is the number of adjacent regions (pixels) and ∂_s is the neighborhood of s. This prior may be seen as an extension of a first order random walk into a two dimensional space. For the second component, β_{unstr}, independent and identically distributed (i.i.d.) Gaussian random priors, $\beta_{unstr}(s) \sim N(0, \tau^2_{unstr})$, are assumed for s=1.....,S. For τ^2_{str} and τ^2_{unstr} 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 MCMC techniques. If α is a vector of the unknown parameters, assuming conditional independence of the parameters, the posterior is given by:

$$\begin{aligned} p(\alpha \backslash y) \propto\ & L(y, \beta_{str}, \beta_{unstr}, \gamma, \sigma^2) \times \\ & p(\beta_{str} \backslash \tau^2_{str}) p(\tau^2_{str}) p(\beta_{unstr} \backslash \tau^2_{unstr}) p(\tau^2_{unstr}) p(\gamma) p(\sigma^2) \end{aligned} \tag{2}$$

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 (p. 603) "classical estimate of fit, plus twice the effective number of parameters." Of the competing models, the specification with the least DIC is selected and reported.

5 Estimation and Results

5.1 Empirical Model

Using the Bayesian methodology described in the previous section, firm and industry characteristics as well as spatial location are introduced simultaneously in the following geoadditive regression framework.[7]

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

The *dependent variable* is the start-up size of new firms. Two different measures of the dependent variable are used for the empirical analysis. In the first model, the start-up size is given by the initial value of fixed assets. In the second model, it is given by the initial employment of the firm.

The estimation output for each model comprises of a table with the estimated effects of the parametric part of the above equation and four maps with the estimated spatial effects. The maps are presented as follows. In the sub-plot (a) the estimated posterior mean of the structured spatial effects are given. As described in the methodology section, the structured spatial effects model spatial dependence and estimate the impact of space as a continuous variate. As the legend shows, a green color indicates that a region has a positive impact on the size of the new firm start-ups, and a red color indicates that the

[7] The model is estimated using the Bayesian software BayesX. The structured spatial effects are estimated based on Markov random field priors and random spatial effects are estimated with gaussian priors. The variance components in all the cases are estimated based on inverse gamma priors with hyperparameters a=0.001 and b=0.001. The number of iterations is set to 120000 with burnin parameter set to 20000 and the thinning parameter set to 100. The autocorrelation files and the sampling paths show that the MCMC algorithm has converged. These plots are available from the authors.

region has a negative impact on the size of new firm start-ups. In the sub-plot (c) the map shows if the estimated spatial effects in sub-plot (a) are statistically significant at 95% confidence level, and in sub-plot (d) the map shows if the estimated spatial effects in sub-plot (a) are statistically significant at 80% confidence level. A white color in these plots indicates that the impact of the spatial unit on the size of new firm start-ups is significantly positive, and a black color indicates that the impact is significantly negative. The sub-plot (b) gives the estimated results for the localised random effects–unlike the structured spatial effects given in (a), these effects are estimated at the local district level. These effects correspond to the random error component of the above model.

After the spatial patterns are estimated using the geoadditive models, the role of regional factors in influencing these estimated spatial patterns is also examined. As firm and industry characteristics are introduced into the geoadditive model, the resulting spatial patterns allow *ceteris paribus* interpretation–they represent spatial effects that remain after filtering firm and industry level effects. A simple strategy to examine the factors that influence these spatial patterns is to regress these residual spatial effects on the regional variables.

5.2 Empirical Results

Figure 1 shows that the size of the new firm start-ups, measured by initial value of fixed assets, systematically varies across space even after controlling for the firm and industry characteristics.

As the structured spatial effects in Figure 1(a) show, while districts of Maharastra and Andhra Pradesh form a belt exhibiting a significantly larger size new firm start-ups, the districts of Uttar Pradesh, Madhya Pradesh, and Bihar, three of the poorest Indian states, form a belt with significantly smaller new firm start-up size.[8] As the 95% and 80% confidence maps in Figures 1(c), 1(d) show, the effects are mostly statistically significant.[9]

The estimated effects of the variables in the parametric part of the geoadditive model are broadly consistent with the existing literature. As the results in Table 2 show, ownership type has a significant influence on the size of new firm start-ups.[10] New-firm start-ups by proprietary owners and female entrepreneurs have a smaller size at start-up. The results suggest that entrepreneurs with technical know-how start-up with a larger firm size. Technical knowledge from abroad increases the size of new firm start-ups by 44%, technical knowledge from other firms increases it by 21% and technical knowledge from universities increases it by 29%. Thus, firms that have technical knowledge at the start-up phase tend to have a larger start-up size than firms that do not have any technical knowledge.

The estimated spatial effects in the second specification-with initial employment as the dependent variable-demonstrate a similar spatial pattern as in the first model. As the maps plotted in Figure 2 show that the new firm start-ups in northern states of Uttaranchal, Uttar Pradesh, Bihar, Madhya Pradesh, and Rajasthan as well as some districts of the southern states of Kerala, Karnataka, and Tamil Nadu are likely to be smaller. In contrast, new firm start-ups in the states of Maharastra, Andhra Pradesh, West Bengal, Punjab and, the northeastern states have a larger size.

The results for the parametric part of the estimation in this empirical specification

---

[8] As discussed in the method section, the total spatial effect is split into the structured spatial component and an unstructured local random component. The unstructured random component corresponds to the random error component of the geoadditive model. The map of the estimated unstructured random component in Figure 1(b) shows that the local spatial effects are smaller than the structured spatial effects suggesting that the effect of geographic space is continuous and not purely random.

[9] Because industry effects are controlled in the estimation, these results can not be attributed to the effect of spatial distribution of labor-intensive industries.

[10] The effect of firm ownership structure on start-up size is estimated as firms that have proprietary ownership structure are more likely to be small compared to those that have partnership or co-operative ownership structures (Spulber 2008).

are consistent with the the results of the first model.

Table 3 shows that the estimated effects of variables in the parametric part of the geoadditive model are robust to in the alternate specification with the firm start-up size measured by the initial employment. Proprietary owners and women entrepreneurs start-up with a smaller firm size. Entrepreneurs with technical knowledge start-up with a larger firm size. New firms started by entrepreneurs in urban locations have a larger size.

The estimation results for both models demonstrate that firm and industry effects do not completely explain the size of new firm start-ups–the size of new firm start-ups systematically varies over space even after accounting for firm and industry effects. Thus, spatial location has an impact on the size of new firm start-ups. In the case of India, the estimated spatial effects clearly suggest a north-south divide in the size of new firm start-ups. New firms in the northern regions have a significantly smaller size when compared to the new firm start-ups in the southern regions, *ceteris paribus*. The maps show that the spatial effects are continuous over space–contiguous districts in the north have a significantly negative effect (these districts are colored either red or black in the maps) and contiguous districts in the south have a significantly positive effect (these districts are colored either green or white in the maps).

We now investigate the factors influencing these spatial patterns. In particular, it may be conjectured that factors such as economic development, financial development, literacy, and unemployment are shaping these spatial patterns. There is overwhelming evidence that financial development has an important influence on economic growth (see Levine, 1997, for a survey). At the level of the firm, Cabral and Mata (2003) find that the firm size distribution is right skewed for start-ups but evolves over time to a less skewed distribution as financial constraints become less binding over time.[11] Lucas (1978) argues that the average firm size increases with economic development. Thus, regional variations in the level of economic development may play a significant role in shaping these spatial patterns. High level of unemployment may compel people into self-employment and make them start small micro-enterprises as an alternative. Literacy rate in the region reflects the abilities of people and is likely influence the size of new firms. Agglomeration and population density are likely to have an impact on the size of new firm start-ups.

In Table 4, the determinants of the mean spatial effects of the first model shown in Figure 1(a) are presented.[12] The R-square indicates that close to 60% of the variation across space that remains after filtering the firm and industry effects can be explained by the regional variables discussed above. Financial development, measured by per-capita credit flows, credit-deposit ratio in the region and density of bank branches has a positive effect on the spatial patterns. This suggests that financial development increases the likelihood of a region having positive impact on the size of new firm start-ups. Economic development, measured by the per-capita net state domestic product, and literacy rate in the region have significant positive effects suggesting that spatial units that are more economically developed give birth to firm start-ups that have a larger size. The coefficient on the unemployment variable is negative and significant suggesting that unemployment in a region leads to new firm start-ups with a small size. The agglomeration index, given

[11] The role of financial constraints on the evolution of FSD is supported by Desai, Gompers and Lerner (2003). Similarly, Fazzari, Hubbard and Petersen (1988) show that financial constraints influence a firm's investment decisions. Holtz-Eakin, Joulfaian and Rosen (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 on the investment in the capital of a new enterprise. Evans and Jovanovic (1989) and Holtz-Eakin et al. (1994) show that the probability of survival depends on assets. Under an assumption that banks lend under the security of collateral, this suggests presence of credit-rationing. Hurst and Lusardi (2004) argue that households with higher levels of wealth have a higher tolerance for risk and are most likely to be business owners.

[12] The data on financial indicators for the regions, the per-capital net state domestic product (NSDP) come from the Reserve Bank of India's Handbook of Statistics on Indian Economy. The data on population density and size of the districts is taken from 2001 Census. The data on small firm density for constructing the agglomeration index is from the third census of small scale industries. The data on unemployment is based on the number of people registered at the employment exchanges.

by the density of small firms in a region, is significantly negative.[13] The demographic variables are mostly insignificant.

Thus, the estimation results suggest that the spatial patterns of firms' start-up size that remain after filtering the firm level and industry level effects, are attributable to a large extent to the variations in economic and financial development across spatial units. Table 5 explains the spatial effects of the second model, shown in Figure 2(a). The results confirm the effects of financial and economic development on start-up size.

6 Conclusion

A growing body of literature examines the determinants of firm size distribution and its evolution over time (Cabral and Mata, 2003; Angelini and Generale, 2008) and a component of this literature examines the determinants of the start-up size of firms (Mata and Machado, 1996; Colombo et al., 2004). This literature focuses mainly on the industry characteristics and personality traits of the entrepreneurs in explaining why initial size varies systematically across start-ups. While a large literature has emerged emphasizing the importance of location in influencing economic phenomenon, the impact of location has generally been ignored in these studies focusing on start-up size. Thus, in this paper, a new database of entrants in India is analyzed using geoadditive models to examine the link between geographic location and the size of new firm start-ups.

The findings resulting from our analysis make several important contributions to the literature. The empirical evidence suggests that, despite being overlooked in the previous studies on startup size, the spatial location is an important micro-determinant of the size of firm start-ups. In particular, spatial neighborhood effects exert a strong influence on firm size at entry. The size of new firm start-ups exhibits remarkably distinct spatial

[13]This may reflect the presence of number of ancillary firms providing vertical integrated services to the few large firms in the region.

patterns after filtering firm level and industry level effects. This also suggests that the right skewed nature of firm size distribution may be attributable to the emergence of many small firms in some regions.

For India, the results reveal a clear north-south divide in the size of new firm start-ups. The results provide important first insights identifying those factors shaping startup size in the context of a developing economy. In particular, ownership structure and initial knowledge endowments are found to influence the size of new firms.

Further, an investigation into the factors that shape these residual spatial patterns shows that variation in financial and economic development across regions can explain, to a large extent, the spatial patterns of firms' start-up size that remain after filtering firm and industry level effects. Thus, the results suggest that regions that are more developed give birth to superior start-ups. Examining the spatial dynamics of firm size distributions is an interesting avenue for future research.

References

Acs, Zoltan J. and Catherine Armington, *Entrepreneurship, Geography, and American Economic Growth*, Cambridge University Press, 2006.

Angelini, Paolo and Andrea Generale, "On the Evolution of Firm Size Distributions," *American Economic Review*, 2008, *98* (1), 426–438.

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

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

_ **and Maryann P. Feldman**, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 1996, *86* (3), 630–640.

_ **and Paula E. Stephan**, "Company-Scientist Locational Links: The Case of Biotechnology," *American Economic Review*, 1996, *86* (4), 641–52.

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

Beck, Thorsten, Asli Demirguc-Kunt, Luc Laeven, and Ross Levine, "Finance, Firm Size, and Growth," *Journal of Money, Credit and Banking*, forthcoming.

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.

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.

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

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

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

Desai, Mihir, Paul Gompers, and Josh Lerner, "Institutions, Capital Constraints and Entrepreneurial Firm Dynamics: Evidence from Europe," Working Paper No. 10165, NBER, National Bereau of Economic Research. December 2003.

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

Evans, David S., "Tests of Alternative Theories of Firm Growth," *Journal of Political Economy*, 1987, *95* (4), 657–674.

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

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

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

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

Gilbert, B. A., "The Implications of Geographic Cluster Locations for New Venture Performance." PhD dissertation, Indiana University 2004.

_ , P.P. McDougall, and D.B. Audretsch, "New Venture Growth: A Review and Extension," *Journal of Management*, 2006, *32* (6), 926.

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

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

Griliches, Zvi, "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 1992, *94*, 29–47.

Guiso, Luigi, Paolo Sapienza, and Luigi Zingales, "Does Financial Development Matter?," *The Quarterly Journal of Economics*, 2004, pp. 929–969.

Hall, Bronwyn H., "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector," *Journal of Industrial Economics*, 1987, *35* (4), 583–606.

Holtz-Eakin, Douglas, David Joulfaian, and Harvey S. Rosen, "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints," *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.

Jaffe, A.B., "Real Effects of Academic Research," *American Economic Review*, 1989, *79* (5), 957–970.

_ , M. Trajtenberg, and R. Henderson, "Geographic localization of knowledge spillovers as evidenced in patent citations," *Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Krugman, Paul, "History and Industry Location: The Case of the Manufacturing Belt," *American Economic Review*, 1991, *81* (2), 80–83.

_ , "Increasing Returns and Economic Geography," *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.

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

Lotti, Francesca and Enrico Santarelli, "Industry Dynamics and the Distribution of Firm Sizes: A Nonparametric Approach," *Southern Economic Journal*, 2004, *70* (3), 443–466.

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

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

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

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

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

Romer, Paul M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 1986, *94* (1986), 1002–1037.

_ , "The Origins Of Endogenous Growth," *Journal Of Economic Perspectives*, 1994, *8* (1), 3–22.

_ , "Endogenous Technological Change."

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.

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

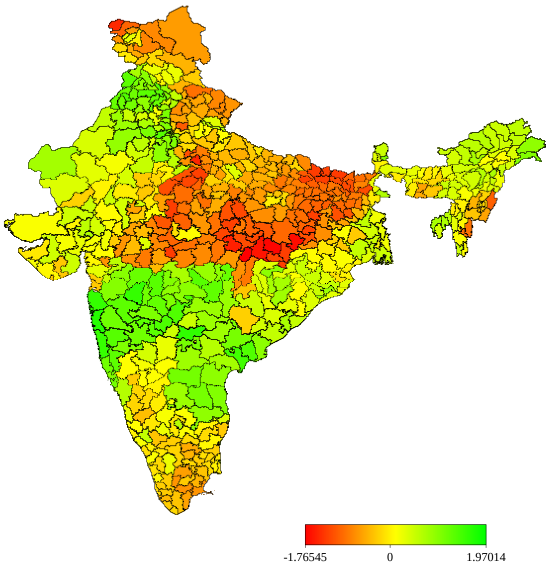
Table 1: Characteristics of Start-ups (Descriptives)

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

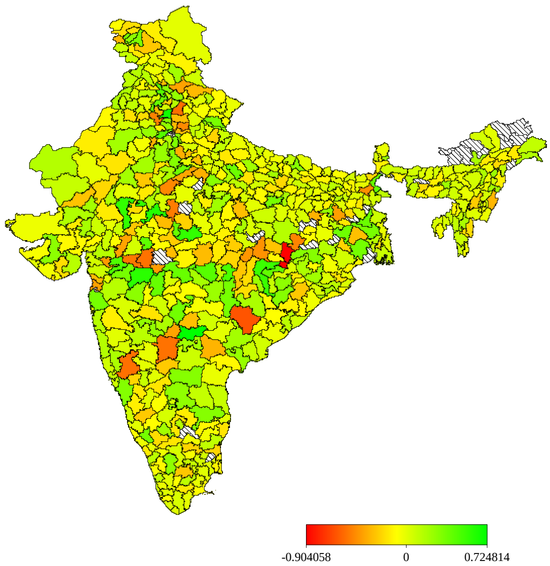
Table 2: Determinants of Start-up Size (Model I)
(Initial Value of Fixed Assets)

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

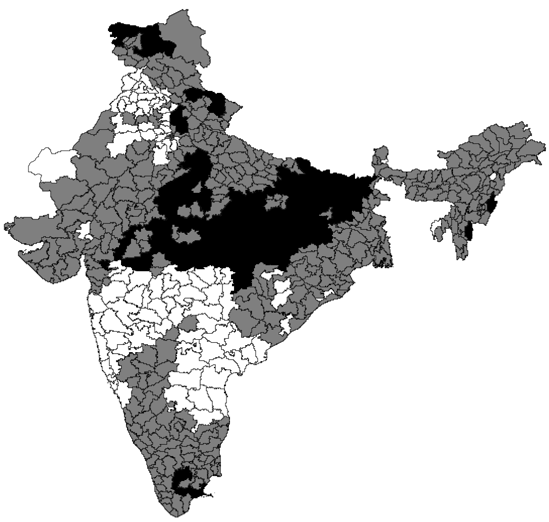
Notes: Dependent variable is log of initial value of fixed assets of the firm.



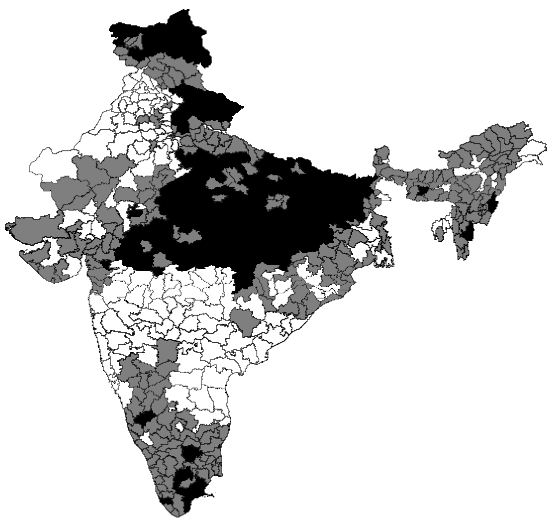
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



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



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



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

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

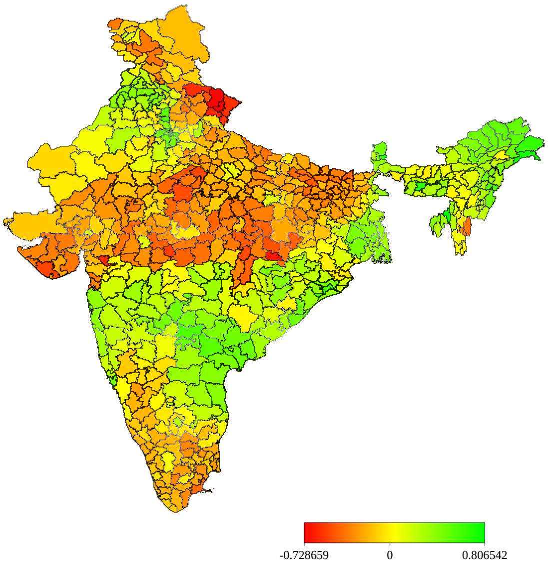
Figure 1: Spatial Effects in Model I

Table 3: Determinants of Start-up Size (Model II)

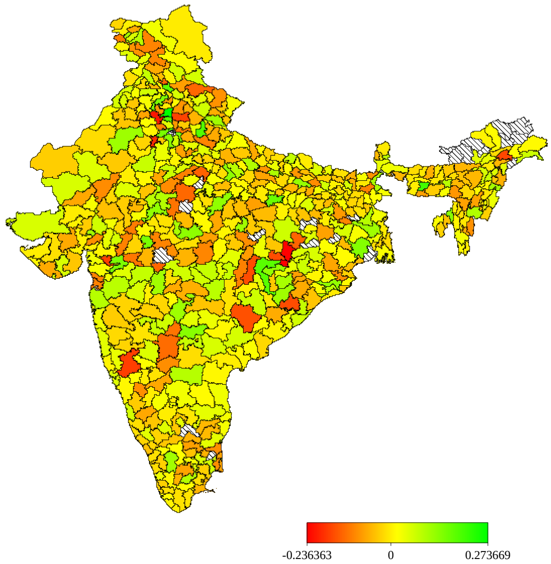
(Determinants of Initial Employment)

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

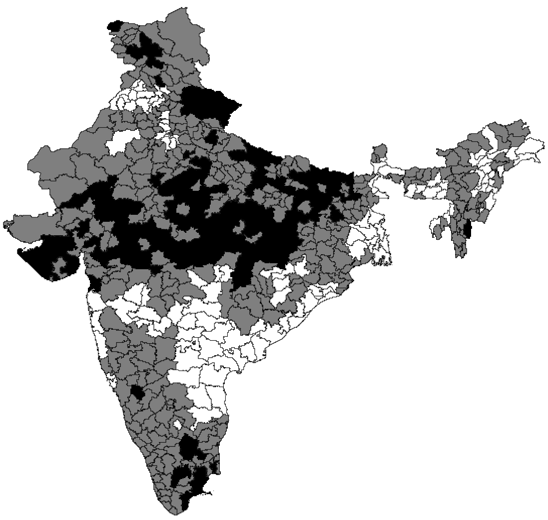
Notes: Dependent variable is log of initial employment of the firm.



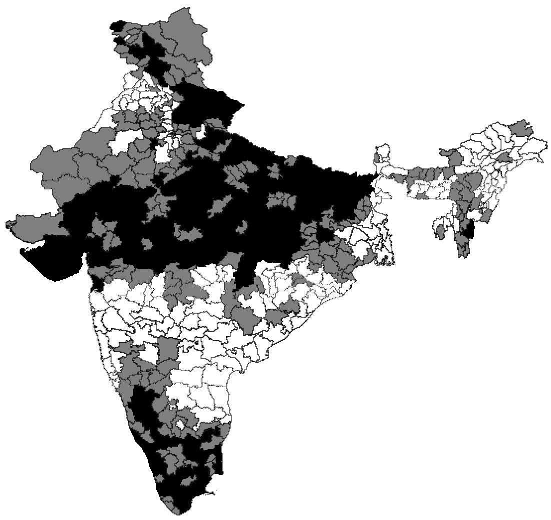
(a) Structured Non linear Effect of 'District'. Shown are the posterior means.



(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.

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

Figure 2: Spatial Effects in Model II

Table 4: Determinants of the Mean Spatial Effects in Figure 1
(Start-up Size given by initial fixed assets)

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

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

Table 5: Determinants of the Mean Spatial Effects in Figure 2

(Start-up Size given by initial employment)

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

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