



**Research Network on Innovation  
Réseau de Recherche sur l'Innovation**

**WORKING PAPERS  
DOCUMENTS DE TRAVAIL**

**N°40 / 2014**

**DOES FIRM CREATION DEPEND ON LOCAL CONTEXT?  
A FOCUS ON THE NEIGHBOURING EFFECTS**

**Nadine LEVRATTO**

# DOES FIRM CREATION DEPEND ON LOCAL CONTEXT? A FOCUS ON THE NEIGHBOURING EFFECTS

Nadine LEVRATTO<sup>1</sup>

**Summary:** This paper seeks to shed some light on the influence that the characteristics of the local context in a given area and in adjacent ones exert on the entrepreneurial process. In order to make a distinction between the purely local factors and the role played by the neighbourhood, we mobilize the so-called Exploratory spatial data analysis which determines the degree of spatial dependence and its consequence on entry rate. We empirically address this question by considering the case of French employment areas from 2006 to 2010 using spatial econometric models adapted to panel data. Our results show that financial, material, human and organisational resources locally available are far from being the unique geographical determinants of firm creation. Instead, the entry rate in a given area also strongly depends on the propensity to create firms in the adjacent places. Spillover effects taking their origin in adjacent areas should thus be considered in explaining the local determinants of firm creation.

**Keywords:** Firm creation, spatial dependence, spatial matrix

**JEL Codes:** L26, R11, C21

© Réseau de Recherche sur l'Innovation, 2014

---

<sup>1</sup> EconomiX, CNRS, University of Paris Ouest Nanterre, La Défense, Centre d'Etudes de l'Emploi, Kedge Business School, nadine.levratto@u-paris10.fr

**DOES FIRM CREATION DEPEND ON LOCAL CONTEXT?  
A FOCUS ON THE NEIGHBOURING EFFECTS**

**Nadine LEVRATTO**

**TABLE**

<b>INTRODUCTION</b>	<b>4</b>
<b>1. THEORETICAL AND EMPIRICAL REVIEW</b>	<b>6</b>
<b>2. METHODOLOGY</b>	<b>7</b>
<b>3. MEASURING THE LOCAL DETERMINANTS OF START-UPS RATE</b>	
3.1. The empirical model	12
3.2. Operationalization of variables	12
<b>4. RESULTS AND DISCUSSION</b>	<b>14</b>
4.1. Spatial distribution of entry rates	14
4.2. Regression results	18
<b>CONCLUSION</b>	<b>23</b>
<b>REFERENCES</b>	<b>24</b>

## INTRODUCTION

From Carree, and Thurik (2010), it is broadly admitted that facilitating entrepreneurship is a vital determinant of economic growth. After having been mainly conceived as a substitute to job creation, policies which facilitate firm creation are definitely seen as a booster of innovative and competitive power. This idea of a positive and statistically robust link between entrepreneurship and economic growth has been supported by a huge variety of empirical studies (Acs 2010, Audretsch et al. 2008, Zacharakis et al. 2000), strengthening the will of policy makers to facilitate entrepreneurial activity. This tendency has been encouraged by recommendations of the OECD (OECD 2010) which lay emphasis on the way governments should encourage business creation and the right practices to adopt in order to shift to a so-called “entrepreneurial society” (Audretsch 2007).

Most of the policies implemented in the OECD countries have aimed at creating the most favourable institutional framework and at giving the most appropriate incentives in order to encourage people to create their own business. Despite the national scope of the policies and measures adopted, many reports have pointed out that territories remain unequal faced with firm creation. Moreover, regional administrations have also placed business creation at the forefront of their economic action with various degrees of effectiveness (Huggins Williams 2011). This local consideration has opened a new research avenue for entrepreneurship studies.

The role played by local elements in determining the entrepreneurial process has been documented by Keeble and Walker (1994). Audretsch and Fritsch (1994) proposed a key synthesis of this question in their research about the local determinants of business creation in Germany. This paper inaugurated a long series of empirical studies aiming at explaining why regions differ from one another as far as entrepreneurship is concerned. Fritsch (1997) shows that the number of regional start-ups launched clearly depends on the industrial structure in the region. Schutjens and Wever (2000) point out the responsibility of the local business climate. Armington and Acs (2002) also note that traditionally most studies on determinants of regional entry use variables such as the unemployment rate and population density as explanatory variables. Some scholars also consider that observed regional differences in levels of entrepreneurial activity can, to a large extent, be explained by individual cultural differences (Boschma et al. 2008). This point of view is, however, contested. Instead, building on Baumol (1990), according to whom, there is little evidence that entrepreneurial spirit differs across regions, Hall and Sobel (2008) affirm that regional differences in entrepreneurship are the result of different institutional arrangements across regions.

Theories of new economic geography and endogenous growth theories (e.g. Krugman 1991; Aghion Howitt 1998) also had a significant influence on subsequent contributions to the study of why entrepreneurs choose a particular location in which to set up their new business activity. These theories imply that spatial agglomerations may create location advantages in terms of spillovers and co-operation between firms (Sorenson Audia 2000). Kibler (2013) recently proposed an integrated framework enabling him to show that individual and local characteristics are mutually reinforcing.

All these papers have in common the fact that they compare employment areas *ceteris paribus*, without taking into account the neighbouring areas. Though, Plummer (2010) recently pinpointed the importance of spatial dependence in entrepreneurship research. This is

especially relevant since the data used are spatial in the sense that the location of the observations is observed (Shane Venkataraman 2000).

In a first attempt to shed some light on the clustering schemes that may prevail in the entrepreneurial process, Levratto (2014) showed evidence on the spatial dependence of firm creation in the French departments using data from 2011. Our purpose is to go one step further considering not only one year, but a whole period to provide more accurate results, eliminating the bias of a narrow, time-specific economic climate.

This paper contributes to the literature by empirically investigating three issues. Does the region matter in the decision to start a new business in France? If region matters, what is inside “the black box” of the regional effect? And, last but not least, to what extent is the entry rate in a given area influenced by the adjacent areas?

Answering these questions when considering a period of several years simultaneously is not straightforward. Indeed, several problems arise when panel data have a locational component. They have been listed by Elhorst (2003). The first one is that spatial dependence may exist between the observations at each point of time so that the model may either incorporate a spatial autoregressive process in the error term or may contain a spatially autoregressive dependent variable. The second one is that parameters may vary over space. It has been recognized that traditional panel techniques are not appropriate as, among many drawbacks, they do not show the differences in behaviour among individual spatial units (Quah, 1996). Spatial panel data models and corresponding estimation techniques have thus been developed to treat spatial dependence and spatial heterogeneity simultaneously.

Our econometric study is based on a unique dataset combining different sources computed at the fine geographical level (the French employment area<sup>2</sup>). The French National Institute of Statistics (INSEE) provided all the databases. We apply spatial econometric techniques to take into account the spatial dependence in the business creation process. We show that the entrepreneurial process is characterized by a positive spatial auto-correlation so that entrepreneurial areas tend to be surrounded by similar ones. In addition, we confirm that the entrepreneurial profile of an area significantly depends on its own characteristics but that the strength of some links between firm creation and some local variables do not resist the introduction of contiguity phenomena.

The remainder of the paper is structured in the following way. Section 2 provides an overview of the literature on local determinants of business creation. In Section 3, we describe the French situation and the difference observed among the 304 mainland employment areas. The specification of the econometric model and the estimation techniques are presented in Section 4. Section 5 presents and discusses the results. Section 6 concludes.

---

<sup>2</sup> An employment area covers a territory in which people both live and work. Consequently, this zone takes into account the daily commute between the home and the place of work. It is defined by the French National Institute of Statistics and Economic Studies and regularly adapted to the changes detected. No change in the definition happened during the period under review. It is roughly equivalent to the NUT 4 level in the European Nomenclature of Territorial Units for Statistics.

## 1. THEORETICAL AND EMPIRICAL REVIEW

The analysis of local differences in the entrepreneurial process began with Reynolds and Storey (1993) and Reynolds *et al.* (1994, 1995). They strongly contributed to popularizing the idea that local characteristics matter in shaping the entrepreneurial profile of a given area.

In this field of research, the most documented analysis has been proposed by Keeble and Walker (1994) who insist upon the diversity of local determinants of firm creation. They identify 31 factors possibly influencing entrepreneurship in a given area and show that the magnitude of the entry rate is a function of the development of the banking system, the specialization of the local labour market, and the size of the city. If entrepreneurial spirit also depends on the average size of already operating companies, the role of this factor differs according to the industry: in the consumer goods industry, the entry rate positively depends on the share of small companies (seedbed effect), whereas in the services industries the higher the share of large companies, the greater the number of new firms created. From these extensive empirical analyses of the spatial differences in business creation, it is possible to identify three major categories of factors able to explain why some areas are more or less entrepreneurial than others. Following Keeble *et al.* (1993) and Johnson and Parker (1996), one may distinguish local demand factors, the local supply of founders and the policy environment.

It can be expected that an entrepreneur's decision about where to locate a business will be influenced by the market size and potential demand expected in the various areas. The solvent demand in a given region depends on the number of companies and inhabitants in the area and their incomes. The literature broadly agrees that higher levels of income increase demand, but they also provide access to capital that a potential entrant needs in order to start a business (Reynolds, 1994). A large population may also be favourable to entrepreneurship. Areas experiencing a high entry rate tend to become more and more attractive for new entrepreneurs (Krugman 1998). On the opposite the same self-reinforcing mechanism exists in areas where an already small population is also decreasing, deterring potential entrepreneurs from launching their business there.

In addition to these market side considerations, firm creation also depends on the economic climate in a given area. Among the different reasons which may explain why an individual decides to act as an entrepreneur, difficulties in finding a job due to a high level of unemployment in a particular employment area is a major one. This idea is developed by Storey (1991), according to whom high unemployment rates can cause higher entry rates, since they force unemployed workers to start their own companies as an alternative to unemployment. This assumption has been tested often, as the survey of literature proposed by Carree recalls (2002). However, the results are quite unclear as they are highly sensitive to the nature of data. Empirical studies using time series often confirm a positive relation between unemployment and firm creation, whereas cross-section studies find a negative one. Santarelli *et al.* (2009) investigate the relationship between regional unemployment and firm entry and exit in Italian regions. They do not find evidence for the 'unemployment push' hypothesis.

The size structure of the productive system in a given area is also expected to play a role in the propensity to create new companies. A high rate of large companies in a given area may then encourage or deter the creation of new business depending on the nature of the industry. In some activities characterized by economies of scales a higher rate of large companies may deter new entrepreneurs from entering the market. On the opposite, in industries where entry

barriers are low, starting a new firm is easier as incumbents do not have the possibility to erect barriers to entry thanks to a growth strategy (Shapiro and Khemani 1987).

The levels of education and skills have been identified as an important determinant of the probability for people to start a new firm (Armington and Acs, 2002; Fritsch and Mueller, 2005). Lee et al. (2004) reconsider this question taking into account skills, the social context and the diversity of the population. They conclude that firm creation is strongly associated with cultural creativity when controlled for the variables suggested in the literature.

Based on Marshall's (1920) suggestion about co-location, many papers agree that agglomeration and the external effects of urbanization may boost entrepreneurial spirit (Puga 2010). These relationships may have different aspects: a specialized local labour market creating potential with a large, skilled labour force, the functioning of local entrepreneurs' networks making it possible to reduce risk, and lack of information that may deter business creation (Acs et al. 2009). As these external and network effects are stronger in high-density areas, these agglomeration effects justify a higher entry rate in urban areas with a high economic and demographic density.

These localisation economies give economic advantages to dense areas. The prevailing rule is that collocation of enterprises generates over-performances (Krugman 1991; Combes et al. 2009; Martin et al. 2010) because of the positive external effects. The different factors generating these effects may also play in favour of an increased entry rate (Joffre-Monseny et al. 2011). Dense areas should thus be characterized by a higher rate of entry because potential entrepreneurs foresee that they have better chances of surviving in such places.

In addition to the influence of regional characteristics on entrepreneurship, neighbourhood effects are more and more considered as key factors in explaining the entrepreneurial process. Plummer (2010) and Plummer and Pe'er (2010) emphasise this spatial dimension. Building on Cooper and Folta (2000), according to whom the performance of start-ups is sensitive to local resources, Plummer (2010) asserts that "there is good theoretical motivation for expecting that many firms-level variables are more spatially dependent in the case of new firms rather than older firms." (p. 150) We are thus required to use not only spatially oriented variables but also econometric techniques able to capture this spatial effect.

Two challenges result from the spatial feature of entrepreneurship. The first one consists in detecting the spatial dependence affecting the "georeferenced" data used to estimate the propensity to create new business. The second one lies in the method to adopt to mitigate this spatial dependence in the context of linear regression. They are successively addressed in the next Section.

## **2. METHODOLOGY**

In this paper we mobilize the so-called Exploratory spatial data analysis (ESDA) to handle the problem of detecting patterns in spatial data (Haining et al. 1998: 457).

Our approach involves two steps: first, to determine the intensity of the spatial distribution of propensity to launch businesses in France, we assess both global and local spatial clustering in the entry rate; then we estimate standard panel data and spatial panel data regression models to compare their respective performance in explaining the entry rate in the 304 French

employment areas. In this paper, we use a first order spatial contiguity matrix that assumes a spatial interaction between an employment area and its immediate neighbours. We adopted a Queen contiguity criterion, which stipulates that two areas are neighbours when they share a common side or vertex. The spatial weight matrix is standardized so that the sum of its rows is equal to one. Such a matrix facilitates the interpretation of neighbouring phenomena underlying the administrative breakdown and improves the efficiency of algorithms during the estimation process (Torres-Preciado et al. 2013).

In order to examine spatial patterns of business creation rates, we assess their global pattern using Moran's I test (Moran 1950, Cliff and Ord, 1981):

$$I = \left( \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \right) * \left( \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \right) \quad (1)$$

where  $n$  is the number of spatial units indexed by  $i$  and  $j$ ;  $y$  is the outcome variable of interest;  $\bar{y}$  is the mean of  $y$ ; and  $w_{ij}$  is an element of spatial weighting matrix,  $W$ , which corresponds to the spatial weights assigned to pairs of units  $i$  and  $j$  (Anselin 1994).

Moran's I statistic is a weighted correlation coefficient used to explore a specific type of spatial clustering. It helps determine whether high values are located in proximity to other high values or whether low values are located in proximity to other low values. Values range from -1, corresponding to a perfect negative correlation, to +1, corresponding to a perfect positive correlation, whereas 0 implies no spatial correlation.

We also calculate Geary's  $C$  (Geary, 1954) defined as:

$$C = \frac{(N - 1) \sum_i \sum_j w_{ij} (X_i - X_j)}{2W \sum_i (X_i - \bar{X})^2} \quad (2)$$

where  $N$  is the number of spatial units indexed by  $i$  and  $j$ ;  $X$  is the variable of interest;  $\bar{X}$  is the mean of  $X$ ;  $w_{ij}$  is a matrix of spatial weights; and  $W$  is the sum of all  $w_{ij}$ .

Geary's contiguity ratio is inversely related to Moran's  $I$ , but whereas Moran's  $I$  is a measure of global spatial autocorrelation, Geary's  $C$  is more sensitive to local spatial autocorrelation. The value of Geary's  $C$  lies between 0 and 2: 1 means no spatial autocorrelation; values lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values higher than 1 illustrate increasing negative spatial autocorrelation.

Finally, we calculate Getis and Ord's  $G$  (Getis and Ord, 1992). The basic statistic is defined as:

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad (3)$$

In this equation, the  $x_j$  are the weighted values of the points in the study area,  $w_{ij}$  is a binary, symmetric weights matrix with ones for all points  $j$  within distance  $d$  of point  $i$  and zeros otherwise.

Getis and Ord's  $G$  measures the concentration of a parameter and indicates the type of cluster that exists (for example in one part of an area there are clusters of higher values than in other



parts). In Ord and Getis (1995), the authors reformulated the statistic so that the results are given in standard normal variants. The statistic is normally distributed and can be used for normal as well as skewed frequency distributions of the underlying variable. However, when the number of neighbours is small, the statistic is less reliable.

In this paper we choose a contiguity spatial weight matrix.<sup>3</sup> In such a matrix (named  $W$ ), the same weight is attributed to all neighbours of an observation ( $w_{ij} = 1$  if  $i$  and  $j$  are neighbours). This square matrix has  $N$  rows  $\times$   $N$  columns corresponding to the 304 employment areas.<sup>4</sup> Its diagonal elements are set to zero by assumption, since no unit can be viewed as its own neighbour. We also scale the individual rows (or columns) of a spatial weight matrix by the row totals to avoid a singularity problem. In such a row-standardized contiguity matrix the spatial weights of each observation  $i$  depend only on the number of its neighbours  $n_i$ :

$$\forall i, \quad w_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } d_{ij} \leq r, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$\text{and } \sum_{j=1}^N w_{ij} = 1$$

Moran's I statistic is a weighted correlation coefficient used to explore a specific type of spatial clustering. It helps determine whether high values are located in proximity to other high values or whether low values are located in proximity to other low values. Values range from -1 corresponding to a perfect negative correlation to +1 corresponding to a perfect positive correlation, whereas 0 implies no spatial correlation.

Moran's I values can then be transformed to Z-scores for statistical hypothesis testing. If the Moran's I rejects the null hypothesis of residuals spatial autocorrelation for most time periods, then this can be interpreted as a good signal that, once controlled for spatial fixed effects and spatial lag dependence in dependent and independent variables, the residuals are free of spatial autocorrelation.

After having controlled for the presence of spatial correlation, the second step consists in estimating a spatial panel data model in order to identify variables responsible for firm creation at the local level. A spatial panel data modelling approach is appropriate as pointed out by some recent studies insisting upon the geographical dimension of the entrepreneurial process (Plummer 2010; Backman and Karlsson, 2013). Thus, the spatial dependent nature of entrepreneurship has to be introduced in our analysis. We also compare the results from spatial panel data models with the standard panel data regression model, such as fixed-effects (FE) estimators.

---

<sup>3</sup> The first-order queen contiguity spatial weight matrix defines all observations that share common boundaries or vertices as neighbours. The first-order rook contiguity spatial weight matrix defines the observations that share common boundaries as neighbours. In the spatial economics literature, a variety of weighting metrics has been proposed. The specification of these matrices, in general, is arbitrary and based on some measure of distance between the geographic units (Moscone et al. 2007). Some of the suggested matrices are based on geographical distance (Anselin 1988) whereas others are more general and are based on social proximity (Conley and Topa 2002) and economic proximity (Conley 1999; Pesaran et al. 2004).

<sup>4</sup> In order to create the weighting, we first translated employment area shapefiles into Stata format by using a user-written "shp2dta" program (Crow 2006) in STATA. Then, the "spmat" procedure developed in STATA was used to generate the  $W$ s.

Spatial econometrics initially adapted to cross-section data has been adapted to estimate econometric relationships based on spatial panels such as:

$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u \quad (5)$$

$$u = \lambda Wu + \varepsilon \quad (6)$$

where the variable  $WY$  denotes the endogenous interaction effects among the dependent variables,  $WX$  the exogenous interaction effects among the independent variables, and  $Wu$  the interaction effects among the disturbance terms of the different units.  $\rho$  is called the spatial

autoregressive coefficient,  $\lambda$  the spatial autocorrelation coefficient, while  $\theta$ , just as  $\beta$ ,

represents a  $K \times 1$  vector of fixed but unknown parameters.  $W$  is a nonnegative  $N \times N$  matrix of known constants describing the arrangement of the units in the sample.

As mentioned by Elhorst (2011), this model has been adapted to account for spatial and temporal heterogeneity. The space-time model extended with spatial specific and time-period specific effects reads as:

$$Y_t = \rho WY_t + \alpha \iota_N + X_t \beta + WX_t \theta + \mu + \xi_t \iota_N + u_t, \quad (7)$$

$$u_t = \lambda W u_t + \varepsilon_t, \quad (8)$$

where  $\mu = (\mu_1, \dots, \mu_N)^T$ . The spatial and time-period specific effects may be treated as fixed effects or as random effects

A large number of model specifications for spatial processes have been proposed<sup>5</sup> to integrate the weight matrix into the regression model. In this paper, we consider a spatially-weighted dependent variable (the so-called "spatial lag model") and a spatially auto-correlated error ("spatial error model") model for incorporating spatial heterogeneity.

The spatial lag or spatial autoregressive model (SAR hereafter) refers to a situation where a phenomenon in one region is affected by a similar one in nearby regions. Such a model is appropriate when there are spillover effects from neighbouring regions.

The SAR model for a panel of  $N$  observations over  $T$  time periods is specified as (Elhorst 2010a):

$$y_{it} = \rho \sum_{j=1}^N w_{ij} y_{jt} + b X_{it} + m_i + e_{it}, \quad (9)$$

where  $\rho$  is the spatial autoregressive coefficient,  $X_{it}$  is a  $1 \times k$  row vector of explanatory variables for an employment area  $i$  at time  $t$  ( $i = 1, \dots, N$  and  $t = 1, \dots, T$ ),  $w_{ij}$  is the element of spatial weighting matrix  $W$ .  $\mu_i$  is an employment area unobserved heterogeneity term and  $\varepsilon_{it}$  is the classical disturbance term with assumptions that  $E(\varepsilon_{it}) = 0$  and  $\text{Var}(e_{it}) = S_e^2$  and  $E(\varepsilon_{it} \varepsilon_{js}) = 0$  for  $i \neq j$  and/or  $t \neq s$ . This FE estimator allows for correlation between time-invariant unobserved heterogeneity at the employment area level and explanatory variables. The spatial autoregressive coefficient,  $\rho$ , measures the extent to which entrepreneurship in

---

<sup>5</sup> LeSage, Pace (2009) and Elhorst (2010) provide a comprehensive presentation of major available spatial panel data econometrics models.

one area is related to the entry rate in neighbouring areas conditional on  $X_{it}$ . A statistically significant value of  $\rho$  indicates that there are spillover effects across geographic entities sometimes difficult to justify.

The spatial error model (abbreviated SEM hereafter) is an alternative to the spatial-lag one. It assumes that one or more explanatory variables have been omitted from the model whereas they influence the dependent variable and are spatially correlated. Such a model is more relevant than the spatial-lag approach when a random shock in a given area spreads to neighbouring regions, i.e. when the distribution of residuals in different employment areas displays spatial correlation. Following Elhorst (2010), we specify the SEM model as:

$$y_{it} = \beta X_{it} + \mu_i + v_{it}, \quad (10)$$

$$v_{it} = \lambda \sum_{j=1}^N w_{ij} v_{jt} + \varepsilon_{it}, \quad (11)$$

where  $\lambda$  represents the spatial error parameter, and  $\varepsilon_{it}$  and  $\mu_i$  are as discussed earlier.

All the previously mentioned spatial approaches have to deal with estimation bias. The multidirectional nature of spatial dependence in the spatial-error model implies that generalized least-square estimators are inconsistent. The spatial-lag model exhibits endogeneity that can be solved using an appropriate maximum likelihood estimator (see Anselin 1988, for details).

A third class of models adds lag effect of the independent variables (LeSage 1999), so that the model is:

$$y = \rho W_1 y + \beta_0 + X \beta_1 + W_1 X \beta_2 + \varepsilon \quad (12)$$

where  $\beta_2$  is parameter of lag on  $W_1 X$ . It is referred as Spatial Durbin Model (SDM). This model was developed because the dependencies in the spatial relationships not only occur in the dependent variable, but also in the independent variables (Anselin 1988, Brasington and Hite 2005, Kissling and Carl 2007). According to LeSage and Pace (2009), the SDM model has several advantages with respect to SAR and SEM. It produces unbiased coefficients in case of problems with the data generating process and, in addition, is not affected by the problem of bias caused by omitted variables.

In the context of panel data, fixed effects can be included in the estimation of equation (9), which leads to a fixed-effects spatial-lag model (SLFE). A maximum likelihood (ML) procedure can be used to estimate the model (Lee and Yu, 2010).<sup>6</sup>

To further test if a spatial effects model outperforms a model without any spatial interaction effects, one may use Lagrange Multiplier (LM) tests for a spatially lagged dependent variable and for spatial error autocorrelation (Debarsy and Ertur, 2010). If a spatial lag model and a spatial error model are estimated separately then likelihood ratio (LR) tests can be conducted to determine which model provides the best fit for the data.

---

<sup>6</sup> The procedure developed by Pisati (2001) to investigate spatially correlated cross-sectional data using maximum likelihood has been adapted to panel data. We thus used the `xsmle` command proposed by Belotti et al. (2013). Similarly, one can estimate a fixed-effects spatial error-model (SEFE) (Elhorst, 2003) using equation (10) and (11).

### 3. MEASURING THE LOCAL DETERMINANTS OF START-UPS RATE

In this paper we use a dataset constructed from different sources made available by the French National Institute of Statistics and Economic Studies (INSEE). They consist in data coming from three main databases:

- (1) *Ad hoc* ones directly provided at the local level describing the number of firm creations, the number of operating companies and the rate of unemployment
- (2) Locally aggregated data such as the rate of skilled employees come from individual data provided at the plant level with an indication of location issued from CLAP (*Connaissance Locale de l'Appareil Productif*), which provides information about the number of employees according to their qualification and the number of employed workers
- (3) Information about the rate of stand-alone companies also result from individual data made available at the company level to determine the degree of independence or of integration within a business group (LIFI, or Financial Linkages dataset)

These different sources enable us to estimate different proxy variables to illustrate the explanatory factors presented above and to determine the entry rate in every employment area. The data are made available yearly from 2006 to 2010, so that we have a panel composed of 304 areas over 5 years. Appendix 1 presents the construction and the summary statistics of the dependent and independent variables. Appendix 2 reports the correlation matrix.

#### 3.1. The empirical model

Noted *CREA*, the entry rate is computed as:

$$CREA_{i,t} = NewEnt_{i,t} / ExistEnt_{i,t} \quad (13)$$

where *NewEnt<sub>i,t</sub>* is the number of new firms created in an employment area *i*, year *t* and *ExistEnt<sub>i,t</sub>* the number of already operating companies in an employment area *i*, year *t*.

The different kinds of local variables identified by the literature as possibly influencing the rate of new business created in an area are inserted in an empirical model aiming at explaining the entry rate at the employment area level. The model reads as follows:

$$CREA_{it} = f(Comp_{it}, Dens_{it}, Manuf_{it}, C5_{it}, Unemp_{it}, DumYear) \quad (14)$$

where the subscript *i* denotes the employment area and *t* the time period. *CREA<sub>it</sub>* is the entry rate; *Comp<sub>it</sub>* is the share of white collar and "grey matter" positions workers in the total number of employees, *Dens<sub>it</sub>* indicates the number of workers par squared kilometre, *Manuf<sub>it</sub>* is the share of people employed in manufacturing industry compared to the total number of workers, *C5<sub>it</sub>* represents the share of employees working in the five biggest plants and *Unemp<sub>it,t</sub>* is the unemployment rate. Finally, the model includes dummy variables for years.

For the estimation of equation (14), we apply a log transformation to density (*DENS*) since the other covariates are defined as percentages ratios. Equation (14) can thus be written as:

$$CREA_{it} = \beta_0 + \beta_1 Comp_{it} + \beta_2 \ln\_Dens_{it} + \beta_3 Manuf_{it} + \beta_4 C5_{it} + \beta_5 Unemp_{it} + \beta_j Year + \mu_i + \lambda_t + u_{i,t} \quad (15)$$

Coefficients  $\beta_j$  with  $j = (6, \dots, 9)$  are associated to the temporal dummies with 2006 as a reference. Finally, we include fixed-region effects ( $\mu_i$ ) as well as fixed (annual) time effects ( $\lambda_t$ ), while  $u_{i,t}$  is the i.i.d. error term of the specification.

### 3.2. Operationalization of variables

The high degree of precision in the breakdown of geographic data used in this paper is challenging because of the lack of easily available data at this spatial level. We circumvent the problems raised by this data shortage, by combining the data made available in the datasets mentioned above and calculating the ratios which enable us to properly approximate the determinants usually identified in the literature.

The definition of the dependent variable, new firm entry, is known for being a real dilemma. The literature shows how different measures of new firm entry may produce different results. Two alternative approaches are possible. The ecological approach standardizes the number of entrants relative to the number of firms in existence to investigate the amount of start-up activity relative to the size of the existing population of businesses. It differs from the ecological approach which extends the concept of entrepreneurial choice proposed by Evans and Jovanovic (1989), according to whom all firms are the result of individual actions. This approach usually introduces the ratio of new companies to working population. According to Garofoli (1994), this approach makes it possible to appreciate the local entrepreneurial spirit more accurately.

As our purpose has mainly to do with the renewal of the productive system, in this paper, we define the entrepreneurial intensity of an area as the ratio of the number of new companies registered there to the total number of companies operating in the same area over the same year.

The independent variables introduced cover the field of potential local determinants of the entry rate. They can be categorized into three different sets. The first set refers to the local economic climate, the second one to the characteristics of the local production system, and the third one considers the agglomeration effects.

The incentives to create a business depend firstly on the capacity to find a job. The different aspects of the relation between unemployment and firm creation are discussed in Audresch and Dhose (2010). Two opposite influences may be identified, the refugee effect, on one hand, and the Shumpeterian opportunity effect, on the other. The refugee effect is inversely related to the regional economic business climate. It is especially strong when people are lead, ideologically, to consider firm creation as a means to improve their skills, to accumulate experience and to be occupied (Aubry et al. 2013). The refugee effect tends to take precedence over the opportunity effect during an economic slowdown. Due to the economic crisis France was still facing in 2011, the coefficient associated with the variable *Unempl* is expected to be positive.

As the education and skills of a given population have been identified factors which influence firm creation, we introduce the share of people in the labour force having at least a Bachelor's degree as an explanatory variable in the model. This variable is noted *Qual*. If the sign associated is positive, it means that skilled people are more able than others to detect business opportunities. If it is negative, it means that low skilled workers may face more difficulties finding a job and have a stronger incentive to create their own.

Average size and competition in a given area may either encourage entrepreneurs to carry out their projects or deter them from doing so. We consider that scale economies are the most powerful driving force in such a process. To capture them, we compute the share of

employees working at the five largest companies in any employment area as a percentage of the total number of employees in the same area. Denominated *C5*, this variable is expected to exert a negative effect.

It is complemented by the share of employees working in stand-alone companies as a function of the total number of employees in a given area (*Indep*). This proxy reflects the capacity of the local productive system to accept the emergence of new competitors and the ease of market entry. The influence of this variable should thus be positive.

When the “net benefits to being in a location together with other firms increase with the number of firms in the location” (Arthur 1990: 237), the entry rate is supposed to be higher in areas where a large number of firms are already located. We adopt this perspective and, as in Ciccone and Hall (1996), we approximate the agglomeration effects using the ratio, total number of employees in a given employment area divided by the area measured in square kilometres of this employment area (*Dens*). As an alternative measurement ratio, we also consider the number of employees in industrial manufacturing compared to the total number of employees (*Manuf*), which may also be introduced as a proxy for agglomeration effects. Whenever the sign is positive, one may conclude that new industrial business tends to locate in industrial areas; whenever it is negative it is possible to consider that entry and firm creation are a complement to industrial activities.

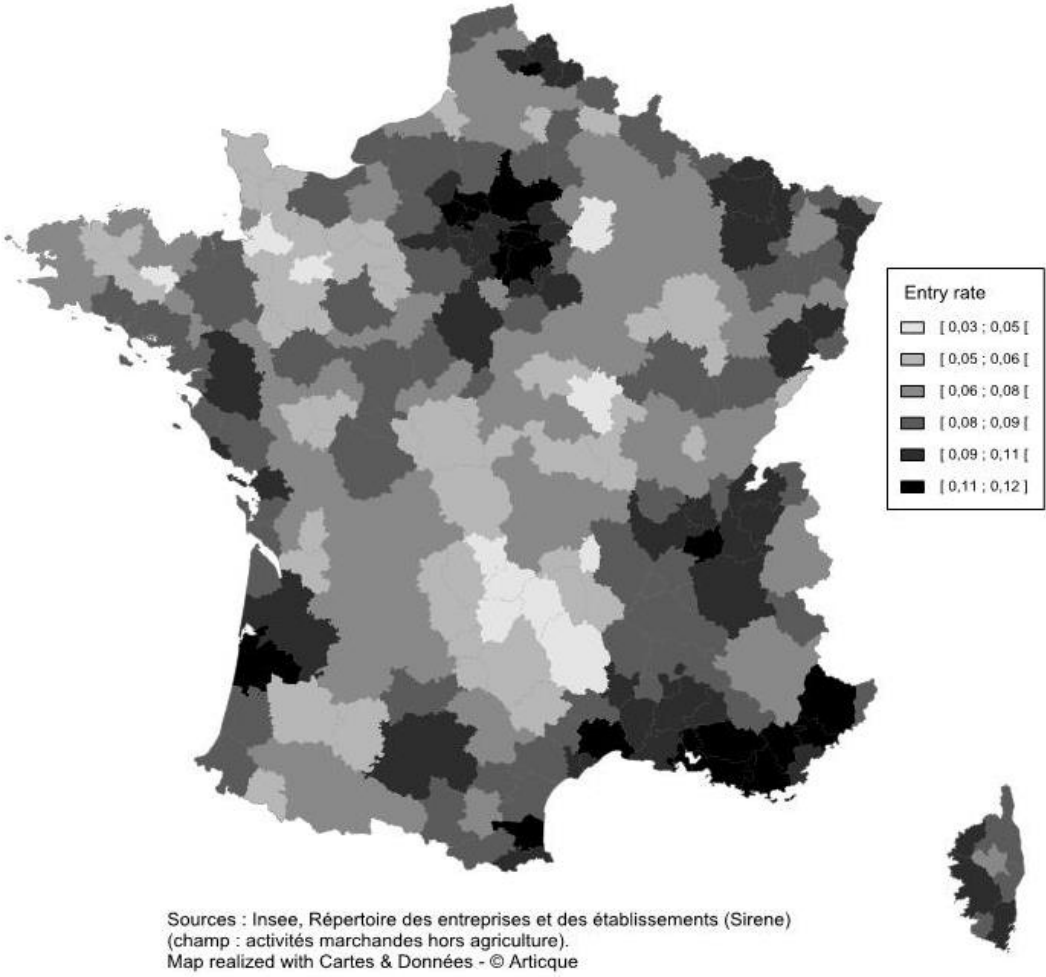
## **4. RESULTS AND DISCUSSION**

This Section presents the results of the estimations. The results of the “classical” panel regression approaches might be biased, because apart from adopting the standard errors according to the Driscoll and Kraay (1998) approach, we neglect any sort of spatial correlation. To take into account these local spillover effects and the fact that those employment areas, which are located next to each other, might disclose a stronger spatial dependence than employment areas at a greater distance, we follow the synthesized methodology proposed by Linderhof et al. (2011). It mainly consists of testing the presence of a spatially dependent scheme and, if so, to run different tests to determine the most appropriate model, mainly fixed or random effects.

### **4.1. Spatial distribution of entry rates**

The average value over the period 2004-2010 per employment area is reported on the following map (Graph 1). At first glance, it appears that entry rates are not random but, instead, that there is a phenomenon of clustering. Mediterranean employment areas, and those of the Atlantic coast and Paris region appear to be more entrepreneurial than those of the centre of France where many weakly entrepreneurial areas are agglomerated.

Graph 1– Average entry rate by employment area (2006-2010)



In order to measure a possible clustering process we perform firstly two tests for spatial dependence: Moran’s *I* and Geary’s *C* (Table 1). Both reveal a strong positive spatial autocorrelation. It is clearly attested to by the value of Moran’s *I* and confirmed by Geary’s *C* which, when positive and close to zero, reveals a positive spatial dependence. This second test is more sensitive than Moran’s to local rather than global spatial autocorrelation. They are complemented by Getis and Ord’s *G*, which identifies the type of cluster that exists.<sup>7</sup>

These tests indicate that there is a positive significant spatial autocorrelation across employment areas for firm creation rates in France over the period 2006-2010.

---

<sup>7</sup> In a panel setting, the analysis can be repeated for all available years. However, if spatial patterns exist for one year, that is already enough to merit the inclusion of spatial econometrics in a model. Performing an analysis for the first and last available year can, however, be relevant in order to visually estimate whether spatial concentration increases or decreases over time (Linderhoff et al. 2011).

Table 1 – Results of the tests for spatial dependency for the variable *CREA* (2006 and 2010)<sup>8</sup>

	Value	E(I)	Sd(I)	Z	p-value*
2006					
Moran's <i>I</i>	0.516	-0.003	0.035	14.872	0.000
Geary's <i>C</i>	0.463	1.000	0.043	-12.577	0.000
Getis & Ord's <i>G</i>	-2.727	0.017	0.185	-14.872	0.000
2010					
Moran's <i>I</i>	0.577	-0.003	0.035	16.627	0.000
Geary's <i>C</i>	0.451	1.000	0.044	-12.599	0.000
Getis & Ord's <i>G</i>	-3.050	0.017	0.184	-16.627	0.000

\*1-tail test

Moran's diagram (Graph 2) compares normalized values of *CREA* in the 304 employment areas with the normalized neighbour's average, generating a two-dimensional plot of *CREA* versus *W1*, so we can visualize the spatial dependence of such a variable. The first quadrant, Q1, the high-high quadrant, shows high values for the variable for both the region and its neighbours. Q2, the low-low quadrant, shows low values for regions as well as their neighbours. If the region has low values, but is surrounded by neighbours with high values, it will be plotted in the Q3 quadrant (low-high) and regions with high levels surrounded by low levels regions will be in Q4 (high-low). It confirms the positive spatial dependence in the data and leads to consider OLS estimators as inappropriate since they are biased.

---

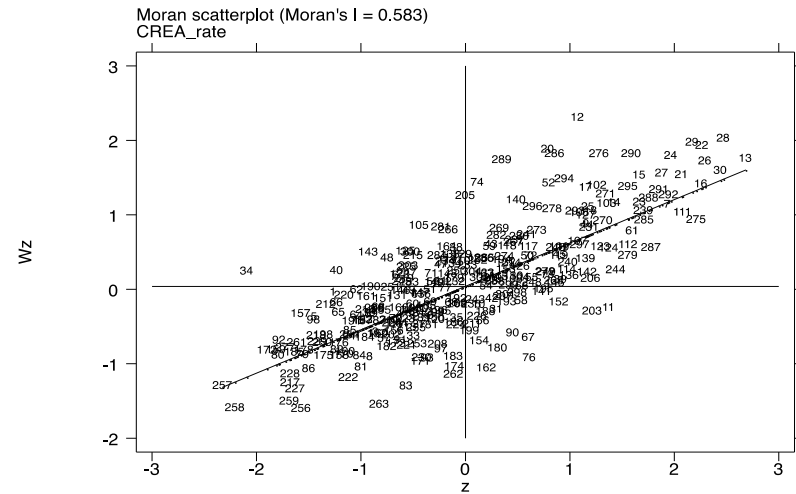
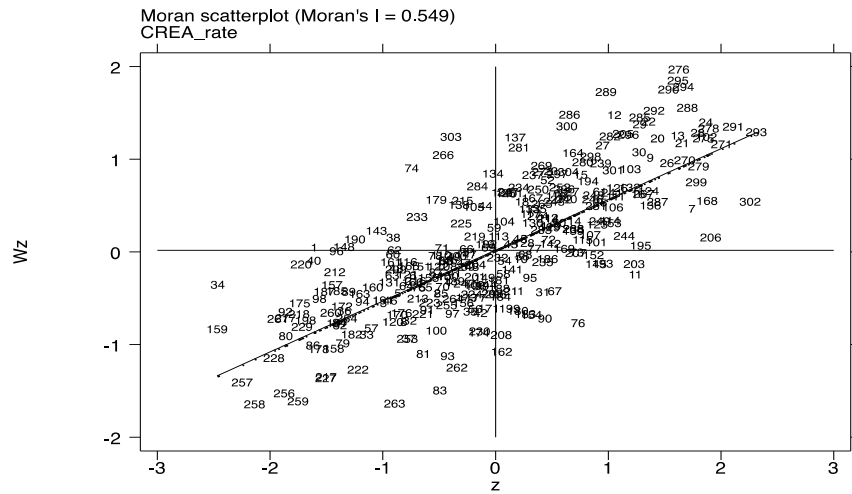
<sup>8</sup> The results do not differ for the years in between. They are available on request from the authors.



Graph 2 – Moran's diagram for the normalized value of entry rate (*CREA*) (2006-2010)

2006

2010



## 4.2. Regression results

Considering the spatial dependence exhibited, we turn to spatial econometric on panel data models to estimate the role played by the diverse local determinants of business creation and the influence of the neighbouring areas. Our initial tests indicate that the OLS model can be rejected in favour of the fixed effects (FE) and the random effects (RE) models. The Breusch and Pagan Lagrangian multiplier test indicates that there are effects other than those captured by the exogenous variables in OLS regressions. The F test, which check for the homogeneity of constant terms across regions and time periods, is also rejected. Moreover, the Hausman test suggests that the FE approach should be preferred to the RE approach. The Hausman test has a value of 249.50 and a p-value of 0.00, so that we opt for an FE model.

Table 2 carries out three tests for spatial error dependence (Moran's  $I_\lambda$ ,  $LM_\lambda$ ,  $LM_\lambda^*$ ) and two tests for spatial dependence ( $LM_\rho$  and  $LM_\rho^*$ ).  $I_\lambda$ ,  $LM_\lambda$  test for the null hypothesis that  $\lambda = 0$ ,

while  $LM_\rho$  tests that  $\rho = 0$ . As mentioned by Pisati (2001),  $I_\lambda$ , and  $LM_\lambda$  also respond to nonzero  $\rho$ ; likewise, when testing for  $\rho = 0$ ,  $LM_\rho$  also respond to nonzero  $\lambda$ . The robust tests  $LM_\lambda^*$  and  $LM_\rho^*$  have been designed to solve this problem (Anselin *et al.* 1996). The results of

the tests lead us to reject  $H_0$  (no spatial dependence) in both tests. They also imply that a model specification with a spatially dependent variable may be favoured over a non-spatial model since we find consistent rejection of the hypothesis of no spatially lagged dependence without being able to determine which kind of model suits better. We will thus use both SAR and SEM with fixed effects models as well as a Durbin model, and we will compare the AIC and BIC statistics to determine the most efficient one.

Table 3 presents the estimation results for different techniques used to estimate the model. Three spatial models for panel data are estimated (1-3) using different techniques whereas the fifth column presents the results obtained using a simple fixed effect model. All the spatial models are run using a row normalized contiguity matrix.

Table 2 – Identification of the source of spatial dependency for the variable *CREA*

Test	Statistic	df	p-value
Spatial error SEM:			
Moran's $I_\lambda$	12.789	1	0.000
$LM_\lambda$	151.384	1	0.000
$LM_\lambda^*$	21.636	1	0.000
Spatial lag SAR:			
$LM_\rho$	154.653	1	0.000
$LM_\rho^*$	24.905	1	0.000

Note:  $LM_\rho$  simple Lagrange multiplier,  $LM_\rho^*$ = robust Lagrange multiplier

Overall, a fixed effect model is the most preferable estimation procedure as suggested by the Hausman test which measures the difference between FE and RE estimators of  $\beta$ . It yields a  $\chi^2(6)$  value of 39.54 and a p-value of 0.00. This leads us to reject the null hypothesis,

according to which differences in coefficients are not systematic, and to conclude that the FE estimator is consistent.

The non-spatial model may suffer from misspecification, as spatial dependence exists within the data as shown in the previous Section. Considering the properties of the models pointed out by LeSage and Pace (2009), SDM should be used when one believes that there are omitted variables in the model that are spatially correlated and that, in addition, these spatially correlated omitted variables are correlated with an excluded variable in the model. Whenever these two conditions hold, the SDM is the appropriate model. In our model, a potentially omitted variable may be a measure of the local gross domestic product, which is strictly impossible to know and to obtain from the National Institute of Statistics.

Table 3 - Determinants of new business formation

	SAR with spatial Fixed effects (1)	SEM with spatial Fixed effects (2)	SDM with spatial Fixed effects (3)	FE without spatial effect (4)				
	Coef.	Std. Err.		Coef.	Std. Err.			
Direct								
Qual	0.037462	0.0278673		0.004336	0.0326927			
In_Dens	0.018574	0.0121604		0.030699***	0.0110419			
Manuf	-0.038266	0.033999		-0.031526	0.0300248			
C5	-0.023847	0.0294286		-0.062821	0.0324278			
Indep	-0.025652***	0.0054903		-0.026061***	0.005859			
Unempl	-0.017817	0.0619768		-0.012550	0.058892			
y2007	0.000726	0.0007159		-0.000619	0.0009868			
y2008	0.001490	0.0010161		0.000758***	0.0002346			
y2009	0.020670***	0.0015191		0.02134***	0.0003175			
y2010	0.019916***	0.0016073		-0.004352***	0.0003286			
Constant								
Indirect								
Qual	0.058862	0.0441202		-0.64845***	0.1569432			
In_Dens	0.02923	0.0197064		-0.007454	0.0369022			
Manuf	-0.060437	0.05421		-0.29513***	0.146983			
C5	-0.037082	0.046162		-0.42953***	0.1626548			
Indep	-0.040007***	0.0092102		-0.04202	0.0276046			
Unempl	-0.027016	0.0948477		-0.168125	0.2006509			
y2007	0.001114	0.0011075		-0.001058	0.0014388			
y2008	0.002301	0.0015209		0.000020	0.0031256			
y2009	0.032108***	0.0018104		0.013847***	0.0042461			
y2010	0.030917***	0.0016876		0.037077***	0.0044313			
Total								
Qual	0.096325	0.0718313	0.086759***	0.0297278	-0.644118***	0.1759954	-0.023724	0.0409408
In_Dens	0.047809	0.0317715	0.037766***	0.0111302	0.023244	0.0407634	0.019245	0.0136264
Manuf	-0.098704	0.0880505	0.007923	0.0298204	-0.326655**	0.1621566	-0.064207	0.0395673
C5	-0.0609	0.0754372	-0.022608	0.0263249	-0.492350***	0.1843437	-0.075188**	0.0377108
Indep	-0.065659***	0.0143892	-0.02422***	0.0050542	-0.068090**	0.0311541	-0.028724***	0.0071458

Unempl	-0.044833	0.156681	0.0636	0.0584483	-0.180675	0.2193015	-0.060589	0.0741802
y2007	0.0018	0.0018204	0.003189***	0.0005828	-0.001678	0.0024213	0.001719*	0.0009372
y2008	0.003791	0.0025297	0.006297***	0.0008482	0.000778	0.0033599	0.00412***	0.0012176
y2009	0.052779***	0.0024229	0.055305***	0.0008815	0.035193***	0.0045624	0.052055***	0.001286
y2010	0.050834***	0.0025003	0.05442***	0.0009391	0.032724***	0.0047594	0.050599***	0.0013505
							0.038621	0.0506772
Lambda			20.60411***	0.0338232				
Rho	0.651408***	0.0225847			0.63418***	0.023123	0.7958084	
$\sigma_e^2$	0.00004***	1.38e-06	0.000038***	1.38e-06	0.00004***	1.36e-06		
lgt_theta								
$\sigma_e^2$							9.68***	
Ident	304		304		304		304	
Observations	1520		1520		1520		1520	
R <sup>2</sup> within	0.910		0.9093		0.917		0.910	
Log likelihood	5517.30		5561.04		5537.08			
AIC	-11010.61		-11098.08		-11038.16		-10470.11	
BIC	-10946.69		-11034.16		-10942.29		-10411.52	

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The spatial Durbin Model with spatial fixed effects provides the best estimation as shown by the respective values of AIC and BIC criteria which reach their lowest value for this model. For brevity, we do not report estimates of local dummy coefficients. When we discuss the significance of SDM parameters, we primarily refer to the total effects defined as the sum of the direct and indirect effects.<sup>9</sup>

The importance of the spatial dependence phenomenon is confirmed by the significant value of Rho which is equal to 0.643 and significant at 1% in the SDM with spatial fixed-effects (column 3, Table 3 in the Appendix). Our first focus is the direct, indirect and total effects of the variables depicting the local business climate. The direct effect of a change in an employment area's rate of skilled workers affects business creation in that same area. According to the result in Table 3, the direct effect is not significant. This means that the rate of skilled workers in a given area does not influence the entry rate in the same area. The spatial econometric technique we use in this paper allows, however, for the quantifying of spatial spillovers in the form of indirect effects. The indirect effect estimated is -0.648 and it is significant at 1% level. So, when the rate of skilled workers in a given employment area increases by 10%, entry rate decreases by 6.48% in the adjacent areas. One possible explanation is that when the rate of qualified workers increases in an employment area, it also reduces the creation of unskilled jobs there. Consequently, adjacent areas become more attractive for low-skilled workers. The total effect defined as the sum of direct and indirect effects remains significantly negative, a result in line with previous studies about the correlation between education and entrepreneurship that do not utilize econometric spatial techniques (Barruel et al. 2012).

Our second concern refers to the local characteristics of the production system approximated by the index of concentration (*C5*) and the share of stand-alone companies (*Indep*). Both exhibit a significant and positive total effect, which strongly differ from the direct one. This legitimates the use of spatial econometric techniques to estimate our model. Indeed, whereas the direct effect of the variable *C5* is not significant, the indirect effect estimated is -0.429 and is significant at 1%. The higher the contribution to total employment of the five largest companies in an employment area, the lower the entry rate in adjacent areas. The total effect requires, thus, discussion as, taking into account both strictly local and spillover effects, it reaches -0.492 and is significant at 1%. This effect is greater than the consensus level.<sup>10</sup> The difference confirms that controlling for spatial dependence matters as not doing so leads to an underestimate of the effect of concentration on firm creation.

That same analysis is echoed by the discussion about the variable *Indep*. The negative direct effect means that an increase of 10% in the share of stand-alone companies leads to a decrease of 2.6% in the entry rate in a given employment area. Changes in the total effect are partially offset by the indirect effect so that, in the end, the coefficient is equal to -0.068 and significant at 1%. Looking at the variable *Manuf*, which captures the agglomeration effects resulting from an industrial profile in the area, one may conclude that if the share of manufacturing industry in a given area does not play any role in determining the entry rate in the same area, spillover effects matter. Indeed, when the share of industrial employment in total employment

---

<sup>9</sup> Statistical inference regarding these effects estimates and how they are calculated are contained in LeSage and Pace (2009).

<sup>10</sup> The regression coefficient estimated in the non-spatial panel cannot be directly compared with the one estimated thanks a spatial technique because the latter contains feedback effects (LeSage and Pace, 2009).

increases by 10% in an employment area, the rate of business creation decreases by 29.5% in the adjacent areas. People detecting employment opportunities in the manufacturing industry are then deterred from becoming entrepreneurs close to their home. In the end, the total effect of the variable *Manuf* is -0.326 and significant at 5% a radically different result from the one obtained with a non-spatial panel model.

## CONCLUSION

Most of the empirical evidence on determinants of entrepreneurship at a local level is based on a-spatial models as the papers generally take into account the local characteristics without considering the influence of adjacent areas. This approach has a few drawbacks such as the inability to explain the clustering structure of the propensity to create new business. Looking at the spatial structure of entry rates immediately convinces us that entrepreneurship can no longer be regarded as independently generated within regions and that possible spillover effects have to be taken into consideration. Firm creation in a given employment area may exert a strong influence on the entrepreneurial spirit in the surrounding areas, either because it generates an imitation effect (positive externality) or because it serves a market which becomes thus unavailable for other potential entrepreneurs in the neighbourhood (negative externalities). As a consequence, standard estimation techniques employed in the majority of empirical studies are no longer relevant as they can lead to biased results. Therefore, in this paper, we introduce a spatial panel approach to account for these spatial effects across areas. Our empirical evidence suggests that spatial spillovers indeed affect the entry rates in contiguous French employment areas.

Some policy implications may result from the negative effect of the agglomeration effects we highlight. Indeed, if concentration plays a role in the firm creation effect, this effect is, nevertheless, negative. Whereas the regional level becomes a key dimension for understanding entrepreneurship, policy makers should thus exert caution when promoting the concentration of production activity or of manufacturing as a means of strengthening the entrepreneurial process. As pointed out by Huggins and Williams (2011), this may be important in remote or peripheral areas where the impact of entrepreneurial policies should not be limited to the start-up rates. Instead, entrepreneurship should be understood and appreciated as a complex phenomenon encompassing measures aiming at educating potential entrepreneurs to raise their chances of success as well as promoting a change in culture. These criteria are barely introduced not only in the evaluation process but also in the academic studies on measuring entrepreneurship. This lack of plurality is also a limit of this paper.

This study may also have future implications for research aiming to explain the spatial pattern and causes of entrepreneurship. Indeed, the data we use do not capture the whole local climate. Some key variables are missing from the analysis. Thus, the challenge consists either in making income variables available or in adapting the geographical breakdown to the official statistics grid. It also pleads for an increased diversity in the criteria used in the literature as well as in public policy evaluation to analyse the different dimensions of the entrepreneurial process.

## REFERENCES

- Acs Z.J. 2010. Entrepreneurship and economic development: the valley of backwardness. *Annals of Innovation & Entrepreneurship* 1: 5641 - DOI: 10.3402/aie.v1i1.5641 (accessed February 2014)
- Acs Z.J., P. Braunerhjelm, D.B. Audretsch and B. Carlsson. 2009. The knowledge spillover theory of entrepreneurship. *Small Business Economics* 32(1): 15-30
- Aghion P. and P. Howitt. 1998. *Endogenous Growth Theory*. Cambridge, MA. MIT Press.
- Anselin L. 1994. Exploratory spatial data analysis and geographic information systems. In: *GeoComputation: A primer*. Eds Longley P, Brooks S, Macmillan B, McDonnell R New York: Wiley, pp 77-94.
- Anselin L., A.K. Bera, R. Florax and M.J. Yoon. 1996. Simple diagnostic test for spatial dependence. *Regional Science and Urban Economics* 26: 77-104
- Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht (the Netherlands): Kluwer.
- Armington C. and Z.J. Acs. 2002. The determinants of regional variation in new firm formation. *Regional Studies* 36(1), 33–45.
- Arthur W., 1990. Silicon Valley locational clusters: When do increasing returns imply monopoly? *Mathematical Social Sciences* 19 (3): 235-273.
- Aubry M., J. Bonnet and P. Renou-Maissant. 2013. *Les cycles macroéconomiques et le comportement entrepreneurial sur données régionales françaises*. 6ème Atelier Stratégies, Espaces et Territoires, AIMS, ASLRF, AEI, Rouen, France, 24th January.
- Audretsch D.B. 2007. *The Entrepreneurial Society*. Oxford: Oxford University Press.
- Audretsch D.B. and D. Dhose. 2010. Cultural diversity and entrepreneurship: a regional analysis for Germany. *The Annals of Regional Science* 45: 55-85.
- Audretsch D.B. and M. Fritsch. 1994. The geography of firm birth in Germany, *Regional Studies* 28(4): 359-365.
- Audretsch, D.B., M.A. Carree, and A.R. Thurik. 2008. Does Entrepreneurship Reduce Unemployment? *Journal of Business Venturing* 23(6): 673–686
- Backman M. and C. Karlsson. 2013. *Determinants of entrepreneurship. Is it all about the individual or the region*. CESIS Electronic Working Paper Series, Paper No. 338, December.
- Barruel F., N. Penaud and S. Thomas. 2012. Créations et créateurs d'entreprises - Première interrogation 2010, profil du créateur. *Insee Résultats* N° 58, February
- Baumol, W.J. 1990. Entrepreneurship: Productive, Unproductive, and Destructive. *Journal of Political Economy* 98(5), 893–921.
- Belotti F., G. Hughes, and A. Piano Mortari. 2013. XSMLE - A Command to Estimate Spatial Panel Models in Stata. *2013 German Stata Users Group Meeting*, Potsdam, June 7<sup>th</sup>.
- Boschma N., V. Schutjens, and K. Suddle. 2008. *Whither a flat landscape? Regional differences in Entrepreneurship in the Netherlands*, SCALES-initiative Paper H 200805 Retrieved from <http://www.entrepreneurship-sme.eu/pdf-ez/H200805.pdf> (Accessed February 2014).
- Brasington, D.M. and D. Hite. 2005. Demand for environmental quality: A spatial hedonic analysis. *Regional Science Urban Economics* 35(1): 57-82.
- Carree M.A. 2002. Does unemployment affect the number of establishments? A regional analysis for US states. *Regional Studies* 36(4): 389–398
- Carree M.A. and A.R. Thurik. 2010. The Impact of Entrepreneurship on Economic Growth. In *Handbook of Entrepreneurship Research*. Eds Z.J. Acs, and D.B. Audretsch, Springer Science+Business Media: 557-594.
- Ciccone A. and R.E. Hall. 1996. Productivity and the Density of Economic activity. *American Economic Review* 86(1), 54-70.



- Cliff A. and J.K. Ord. 1981. *Spatial processes: Models and applications*. London: Pion
- Combes P., G. Duranton, L. Gobillon, D. Puga and S. Roux. 2009. *The productivity advantages of large cities: distinguishing agglomeration from firm selection*. Working Paper, Crest, February.
- Conley T.G. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* 92:1-45.
- Conley T.G. and G. Topa. 2002. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics* 17:303-327
- Cooper A.C. and T.B. Folta. 2000. Entrepreneurship and high technology clusters. In *Handbook of Entrepreneurship*. Eds Sexton D.L. and H. Landstrom, Blackwell: Oxford, 348-367.
- Crow K. 2006. SHP2DTA: Stata module to convert shape boundary files to Stata datasets, *Statistical Software Components*, Boston College Department of Economics. Retrieved from <http://ideas.repec.org/c/boc/bocode/s456718.html>. Accessed February 2014.
- Debarys, N. and C. Ertur. 2010. Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science Urban Economics* 40(6): 453–470.
- Driscoll, J., and A. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80(4), 549-560.
- Elhorst, J.P. 2003. Specification and estimation of spatial panel data Models. *International Regional Science Review* 26(3): 244-68.
- Elhorst, J.P. 2010. Spatial panel data models. In *Handbook of applied spatial analysis*. Eds Fischer M.M. and A. Getis. Berlin, Heidelberg and New York: Springer, pp 377-407
- Elhorst, J.P. 2011. *Spatial panel models*. Mimeo, September.
- Evans D. and B. Jovanovic. 1989. An Estimated Model of Entrepreneurial Choice under Liquidity Constraints. *Journal of Political Economy* 97: 808-827.
- Fritsch, M. 1997. New Firms and Regional Employment Change. *Small Business Economics* 9(5): 437-448.
- Fritsch, M., and P. Mueller. 2004. The effects of new business formation on regional development over time. *Regional Studies* 38(7): 961–76
- Garofoli G. 1994. New Firm Formation and Regional Development: The Italian Case. *Regional Studies* 28(4): 381-393
- Geary R.C. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5 (3), 115–145.
- Getis. A. and J. K. Ord.,1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, 189-206
- Haining, R., S. Wise and J. Ma. 1998. Exploratory spatial data analysis in a geographic information system environment. *The Statistician* 47: 457-469.
- Hall J.C. and R.S. Sobel. 2008. Institutions, Entrepreneurship, and Regional Differences in Economic Growth. *American Journal of Entrepreneurship* 1(1), 69-96
- Huggins R. and N. Williams. 2011. Entrepreneurship and regional competitiveness: The role and progression of policy. *Entrepreneurship & Regional Development* 23(9-10), 907-932.
- Jofre-Monseny J., R. Marin-Lopez, and E. Viladecans-Marsal. 2011. The mechanisms of agglomeration : evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics* 70: 61-74.
- Johnson P. and S. Parker. 1996. Spatial variations in the determinants and effects of firm births and deaths. *Regional Studies* 30(7), 679–688
- Keeble D. and S. Walker. 1994. New Firms, Small Firms and Dead Firms: Spatial Patterns and Determinants in the United Kingdom. *Regional Studies* 28(4):411-427.

- Keeble D., S. Walker, and M. Robson. 1993. *New firm formation and small business growth in the United Kingdom: spatial and temporal variations and determinants*. Research Series Department of Employment, 15, London
- Kibler E. 2013. Formation of entrepreneurial intentions in a regional context. *Entrepreneurship & Regional Development* 25(3-4): 293-323.
- Kissling W.D. and G. Carl. 2007. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology Biogeography* 17: 59-71.
- Krugman P. 1991. *Geography and Trade*. Cambridge, MA: MIT Press
- Krugman P. 1998. Space: The final frontier. *Journal of Economic Perspectives* 12(2), 161-174
- Lee S.Y., R. Florida, and Z.J. Acs. 2004. Creativity and Entrepreneurship: A Regional Analysis of New Firm Formation. *Regional Studies* 38(8), 879-891
- Lee, L. F. and Yu, J. 2010. Estimation of spatial autoregressive panel data models with fixed effects, *Journal of Econometrics* 154(2): 165-185.
- LeSage J.P. 1999. *The Theory and Practice of Spatial Econometrics*. Department of Economics, University of Toledo
- LeSage J.P. and R.K. Pace, 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC
- Levratto N. (2014) *Regional dynamics and start-ups: Evidence from French departments in 2011*. Mimeo EconomiX, February.
- Linderhof V., Nowicki P., van Leeuwen E., Reinhard S. and Smit M. 2011. *Manual for the tests of spatial econometric model*. Research report, The Hague: LEI, part of Wageningen UR, 2011 (SPARD Deliverable D4.1)
- Marshall A. 1920. *Principles of Economics; An Introductory Volume*. London, U.K.: Macmillan and Co.
- Martin P., T. Mayer, and F. Maynerie. 2010. Spatial concentration and plant-level productivity in France. *Journal of Urban Economics* 69(2): 182-195.
- Moran P.A.P. 1950. Notes on Continuous Stochastic Phenomena, *Biometrika* 37(1): 17-23.
- Moscone F., M. Knapp, and E. Tosetti. 2007. Mental health expenditure in England: a spatial panel approach. *Journal of Health Economics* 26:842-864.
- OECD. 2010. *SMEs, Entrepreneurship and Innovation*. Paris : OECD Publications.
- Ord J.K. and A. Getis. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27, 286-305.
- Pesaran M.H., T. Schuermann and S.M. Weiner. 2004. Modelling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model. *Journal of Business and Economic Statistics* 22:129-162
- Pisati, M. (2001) Tools for spatial data analysis. *Stata Technical Bulletin* 60: 21-37. *Stata Technical Bulletin Reprints*. College Station, TX: Stata Press. 10: 277-298.
- Plummer, L.A. 2010. Spatial dependence in entrepreneurship research: Challenges and methods. *Organizational Research Methods*. 13(1): 146-175.
- Plummer L.A. and A. Pe'er. 2010. The Geography of Entrepreneurship. in *Handbook of Entrepreneurship Research*. Eds. Acs Z. and D.B. Audretsch. Volume 5, Berlin: Springer Verlag, 519-556
- Puga D. 2010. The magnitude and causes of agglomeration economies, *Journal of Regional Science*. 50(1): 203–219
- Quah, D. T. 1996. Empirics for economic growth and convergence. *European Economic Review*. 40(6): 1353-1375.
- Reynolds P. 1994. Autonomous firm dynamics and economic growth in the United States, 1986–1990. *Regional Studies*. 28(4): 429–442.

- Reynolds P.D. and D.J. Storey 1993. *Local and regional characteristics affecting small business formation: a cross national comparison*, Paris: OECD Publications.
- Reynolds P.D., D.J. Storey and P. Westhead, 1994. Cross-national comparisons of the variation in new firm formation rates. *Regional Studies*. 28(4): 443-456.
- Reynolds P.D., B. Miller and W.R. Maki.1995. Explaining regional variation in business births and deaths: U.S. 1976-1988. *Small Business Economics*. 7(5): 389-407.
- Santarelli E., M. Caree and I .Verheul. 2009. Unemployment and Firm Entry and Exit: An Update on a Controversial Relationship. *Regional Studies* 43(8): 1061-1073.
- Schutjens V.A.J.M. and E. Wever. 2000. Determinants of new firm success. *Papers in Regional Science*. 79(2): 135–159.
- Shane S. and S. Venkataraman. 2000. The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25: 217-226.
- Shapiro D. and R.S. Khemani. 1987. The determinants of exit and entry reconsidered. *International Journal of Industrial Organisation*. 15(1): 15–26.
- Sorenson O. and P.G. Audia. 2000. The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940-1989. *American Journal of Sociology*. 106(2): 424-462.
- Storey D.J. 1991. The birth of new firms—Does unemployment matter? A review of the evidence. *Small Business Economics*. 3(3): 167–178.
- Torres-Preciado V.H., M. Polanco-Gaytán and M.Á. Tinoco-Zermeño, 2013. Technological innovation and regional economic growth in Mexico: a spatial perspective. *Annals of Regional Science*. DOI 10.1007/s00168-013-0581-1 (published on line 6 November 2013)
- Zacharakis, A., W. Bygrave, and D. Shepherd. 2000. *Global entrepreneurship monitor: National entrepreneurship assessment, United States of America*. Kansas City: Kauffman Center for Entrepreneurial Leadership.

### Appendix 1 - Descriptive statistics

Variables	Definition	Mean	Std. Error	Min	Max
$CREA_i$	Entry rate = Number of entering firms in an employment area $i$ divided by number of firms in the same employment area	0.076	0.032	0.021	0.180
$Qual_i$	Share of college-educated employees compared to the total number of employees in employment area $i$	0.120	0.033	0.061	0.320
$ln\_Dens_i$	Number of employees in employment area $i$ divided by the area in square kilometres of employment area $i$	3.453	0.977	1.283	8.643
$Manuf_i$	Share of employees in Manufacturing Industry compared to the total number of employees in employment area $i$	0.165	0.076	0.026	0.454
$C5_i$	Share of the five largest operating companies in the total number of employees in employment area $i$	0.112	0.046	0.013	0.284
$Indep_i$	Share of employees in stand-alone companies in the total number of employees in operating companies in an employment area $i$	0.824	0.115	0.319	0.999
$Unempl_i$	Share of unemployed aged 16–64 in employment area $i$ divided by population aged 16–64 in employment area $i$	0.085	0.022	0.036	0.164

### Appendix 2 - Correlation Matrix

	crea_rate	qual	ln_dens	manuf	C5	indep	unempl
crea_rate	1.000						
qual	0.087***	1.000					
ln_dens	0.380***	0.638***	1.000				
manuf	-0.236***	-0.198***	-0.061*	1.000			
c5	-0.149***	-0.196***	-0.273***	0.339***	1.000		
indep	-0.246***	-0.520***	-0.629***	-0.200***	0.101***	1.000	
unempl	0.424***	-0.127***	0.157***	-0.167***	-0.102***	0.107***	1.000

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### Appendix 3 - Summary of the normalized spatial-weighting object

Matrix	Description
Dimensions	304 x 304
Stored as	304
Values	
min	0
min>0	0.0006306
mean	0.0032895
max	0.0873724