Working Paper IMRI
2006/05

Spatial spillovers in France: a study on individual count data at the city level

Emilie-Pauline Gallié* and Diègo Legros[†]

28th August 2006

Abstract:

In this research, we focus on two additional dimensions of the issue of spatial autocorrelation in spillovers measures. First, we use patent data as dependent variable. These are count data, characterised by non-negative integer numbers and an important number of zeros. The model usually used in this case, is based on Poisson distribution. The methods used to take spatial auto-correlation into account are not anymore appropriate. We then propose in this article to use a new original method to estimate spatial dimension of spillovers with count data. The method is based on generalised cross entropy approach. Second, contrary to previous studies which use spatial aggregated data, we concentrate on individual data. The idea is that the local dimension can be smaller than this of a region, a French department or even an American MSA. This allows us to take a smaller geographical dimension and to test if firms benefit from spillovers of their very close neighbours or of their further neighbours.

Keywords: Geography of innovation, Knowledge spillovers, Spatial dependence, Entropy.

JEL Classification: O31, O18, R12, C21, C25.

Acknowledgements: We are thankful to P. O. Flavigny for his help to create the contiguity matrix.

*Institut pour le Management de la Recherche et de l'Innovation, Université Paris IX Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 PARIS CEDEX 16. Phone : +33 (0)1 4405 4819, Fax : +33 (0)1 4405 4849, E-mail: emilie-pauline.gallie@dauphine.fr

[†]Institut National de **RE**cherche sur les **T**ransports et leur **S**écurité, 2 avenue du Général Malleret-Joinville, 94 114 Arceuil cedex, Phone: + 33 (0)1 4740 7266, Fax: +33 (0)1 4547 5606, Email: legros@inrets.fr and ERMES-CNRS FRE 2887, Université Panthéon-Assas Paris II, 12 place du Panthéon, 75231 Paris Cedex 05.

1 Introduction

For 20 years, the economic literature has showed a considerable attention to the contribution of spillovers to local innovation and more generally to economic growth. Jaffé (1989) is the first one to propose an econometric model to measure the local dimension of spillovers, thanks to a knowledge production function. He shows that spillovers are localised. Several studies (Feldman, 1994...) confirm these results. However, these early models only test spillovers inside a delimited area. Consequently, they confirm the presence of local spillovers but this does not imply that spillovers only occur at local levels.

Anselin et al. (1997) improve the model in testing the impact of local research and this of the proximate research. In the same idea, Autant-Bernard (2001) introduce three spatial levels of research in the knowledge production function in order to test the distance spillovers can cover. Her study confirms that spillovers are geographically limited. Introducing spatially lagged exogenous variables, as these authors do, allows to treat spatial auto-correlation. This is an important topic in new economic geography as it is often considered that it can have a relation between different geographic observations, that means the statistical hypothesis of the independence of the observations would not be verified. The main advantage of the method (cross regressive model) used by Anselin et al. (1997) and Autant-Bernard (2001) is that it does not require specific methodology. The estimation can be based on ordinary least squares (Le gallo, 2000).

Recently, models propose two others methods to take into account spatial autocorrelation: the substantive dependence and the nuisance dependence. The first one is based on lagged dependent variable, the second one on error-autocorrelation. These methods are judged more relevant in applied empirical work (Anselin et al., 2000) but they require specific estimation methods. For the moment, they are available for data requiring ordinary least squares regression.

In this research, we focus then on two additional dimensions of the issue of spatial autocorrelation in spillovers measures. First, we use patent data as dependent variable. These are count data, characterised by non-negative integer numbers and an important number of zeros. The model usually used in this case, is based on Poisson distribution¹. The methods used to take spatial autocorrelation into account are not anymore appropriate. We then propose in this article to use a new original method to estimate spatial dimension of spillovers with count data. The method is based on generalised cross entropy approach (Golan et al., 1996).

Second, contrary to previous studies which use spatial aggregated data, we concentrate on individual data. The idea is that the local dimension can be smaller than this of a region, a French department or even an American MSA. This allows us to take a smaller geographical dimension and to test if firms benefit from spillovers of their very close neighbours or of their further neighbours.

¹The models of the geography of innovation also use the number of patents but the studies generally take into account the whole of the economic activity ; the data relating to this variable contain then very few zeros and big numbers. They can be regarded as quasi-continuous and be estimated by the OLS method.

The remainder of the paper is organized in five sections. The model based on a knowledge production function (of the sort proposed by Griliches-Jaffe) is developed in section 2. Section 3 explains the estimation methodology. Section 4 presents the data. The results of the analysis are detailed in section 5. Finally, section 6 concludes.

2 The model

The model is specified with the objective of picking up the effects of the spatial dimension of spillovers. The specification is based on that employed in studies belonging to the geography of innovation. These models must simultaneously integrate two dimensions: spillovers and spatial dimension (Feldman, 1994). As in many models, spillovers are tested as an external stock of knowledge using a knowledge production function (Griliches, 1979). This function expresses the relation between the inputs of R&D and the output of R&D. The main problem from an econometric viewpoint resides in modelling the spatial dimension of spillovers. The first models of the geography of innovation (Jaffé, 1989; Audretsch and Feldman, 1996) test spillovers through the level of R&D concentration. They show that there are spillovers inside a limited geographical area but do not capture the geographic extend of diffusion. On the other hand, Anselin et al. (1997), Autant-Bernard (2001), and Bottazzi et al. (2003), are interested in determining the distance that spillovers can cover. To do this, they compare the impact of R&D investments in more or less further geographical areas on the output of the studied area. More precisely, “the external stock of knowledge is divided into two, an external stock of research carried out in the vicinity and an external stock of research carried out at a longer distance”(Autant-Bernard, 2000, p.111). This method makes it possible to show that if firm activities are more affected by the local activities than by those of geographically distant agents, then spillovers are localised. Thus, it seems more adapted to measuring spillovers and their spatial transferability. This method is used to model R&D.

Moreover, the economic literature traditionally distinguishes two types of spillovers: the intra-sector-based ones according Marshall (1920) and the inter-sector-based ones following Jacobs (1969). Indeed, are spillovers more important between firms of a same sector or between firms of different sectors? This first one would facilitate exchange between firms which have the same competences via a local labour market, local atmosphere for exchange between rivals and between clients and sellers. The second type would contribute to the effects of cross fertilization and offer new opportunities. Many studies (Glaeser et al., 1992, Henderson, 2003, Autant-Bernard and Massard, 2004; Leblanc, 2004;...) deal with this question in order to know which spillovers have a higher impact on local activities. However, they do not reach the same conclusion. Some authors underline the role of inter-sector-based spillovers (Glaeser et al., 1992) whereas others show the predominance of intra-sector-based spillovers (Jaffe, 1989). As for Audretsch and Feldman (1999), they find a negative impact of the specialisation. It seems that the results differ according to the sectors, the geographical level of study. We propose then to distinguish in our model the

intra-sector-based R&D from the inter-sector-based R&D in order to know when we study spillovers in a small area which is the city (in comparison with other studies on American States or French department), which dimension prevails? The model is then:

$$Kic = f(RDic, RDcz, RDcz, RDck, RDck, Sc) \quad (1)$$

The variable $RDic$ represents R&D carried out by the firm i ,. $RDcz$ is the R&D of the firms located in the same city and in the same sector z than the firm i . $RDcz'$ measures the R&D carried out in the sector z , in the bordering cities of the city where i is located. $RDck$ is the R&D in other sectors carried out in the city c . $RDcz'$ measures the R&D carried out in the other sectors, in the bordering cities. Sc represents the whole sector-based dummy variables, characteristic of the firm.

3 Methodology

The origin of entropy dates back to 19th century. In 1948, the entropy concept as a measure of uncertainty was developped by Shannon. A decade after in 1957, Jaynes formulated Shannon's entropy as a method for estimation and inference particularly for ill-posed problems by proposing the so-called Maximum Entropy (ME) principle. More recently, Golan et al. (1996) developed the Generalized Maximum Entropy (GME) estimator and started a new discussion in econometrics.

3.1 The generalized maximum entropy (GME) approach

Suppose that we observe a T -dimensional vector $\mathbf{y}$ of noisy indirect observations on a unknown and unobservable K -dimensional parameter vector β , where $\mathbf{y}$ and β are related through the following linear model relationship:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \quad (2)$$

where $\mathbf{X}$ is the $T \times K$ know matrix of explanatory variables, β is a $(K \times 1)$ vector of parameters to be estimated and $\mathbf{u}$ is a $T \times 1$ disturbance vector.

In order to be able to use Jaynes' Maximum Entropy principle for the estimation of regression parameters, the vector of parameters β must be written in terms of probabilities because of the fact that the arguments of the Shannon's maximum entropy function are probabilities. Following Golan et al. (1996), if we define $M \geq 2$ equally distanced discrete support values, z_{km} as the possible realizations of β_k with corresponding probabilities p_{km} , we can convert each parameter β_k as follow:

$$\beta_k = \sum_{m=1}^M z_{km} p_{km} \quad \text{for } k = 1, 2, \dots, K \quad \text{where } M \geq 2 \quad (3)$$

Let us define the M dimensional vector of equally distanced discrete points (support space) as $\mathbf{z}'_k = (z_{k1}, \dots, z_{kM})$ and associated M dimensional vector of probabilities as $\mathbf{p}_k = (p_{k1}, p_{k2}, \dots, p_{kM})'$. Now, we can write β in equation (2) as:

$$\beta = \mathbf{Z}\mathbf{p} = \underbrace{\begin{pmatrix} \mathbf{z}'_1 & 0 & \cdots & 0 \\ 0 & \mathbf{z}'_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{z}'_K \end{pmatrix}}_{(K, KM)} \underbrace{\begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{pmatrix}}_{(KM, 1)} \quad (4)$$

where $\mathbf{Z}$ is a block diagonal matrix of support points with:

$$\mathbf{z}'_k \mathbf{p}_k = \sum_{m=1}^M z_{km} p_{km} = \beta_k \quad \text{for } k = 1, 2, \dots, K \quad , \quad m = 1, 2, \dots, M \quad (5)$$

where $\mathbf{p}_k$ is a M dimensional proper probability vector² corresponding to a M dimensional vector of weights $\mathbf{z}_k$. Recall that the vector $\mathbf{z}_k$ define the support space of β_k . By this way, each parameter is converted from the real line into a well-behaved set of proper probabilities defined over the supports.

As we can see, the implementation of the maximum entropy formalism allowing for unconstrained parameters starts by choosing a set of discrete points by researcher based on his *a priori* information about the value of parameters to be estimated, where these set of discrete points are called the support space for all parameters. In most cases, where researchers are uninformed as to the sign and magnitude of the unknown β_k , they should specify a support space that is uniformly symmetric around zero with end points of large magnitude, say $\mathbf{z}'_k = (-C, -C/2, 0, C/2, C)$ for $M = 5$ and for some scalar C (Golan et al., 1996:77).

Similarly, we can also transform the noises $\mathbf{u}$ as follows (Golan et al., 1996:87):

$$u_t = \sum_{j=1}^J \nu_{tj} w_{tj} \quad \text{for } t = 1, 2, \dots, T \quad \text{where } J \geq 2 \quad (6)$$

Notice that by this conversation, Golan et al. (1996:121) propose a transformation of the possible outcomes for u_t to the interval $[0, 1]$ by defining a set of discrete support points $\mathbf{v}'_t = (v_{t1}, v_{t2}, \dots, v_{tJ})$ which is distributed uniformly and evenly around zero (such that $v_{t1} = -v_{tJ}$ for each t if we assume that the error distribution is symmetric and centered about 0) and a vector of corresponding unknown probabilities $\mathbf{w}_t = (w_{t1}, w_{t2}, \dots, w_{tJ})$ where $J \geq 2$. Now we

²A proper probability vector is characterized by two properties: $p_{km} \geq 0 \quad \forall m = 1, \dots, M$ and $\sum_{m=1}^M p_{km} = 1$.

can rewrite $\mathbf{u}$ in (2) as:

$$\mathbf{u} = \mathbf{V}\mathbf{w} = \underbrace{\begin{pmatrix} \mathbf{v}'_1 & 0 & \cdots & 0 \\ 0 & \mathbf{v}'_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{v}'_T \end{pmatrix}}_{(T,KJ)} \underbrace{\begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_K \end{pmatrix}}_{(KJ,1)} \quad (7)$$

with:

$$\mathbf{v}'_t \mathbf{w}_t = \sum_{j=1}^J \nu_{tj} w_{tj} = u_{ttj} \quad \text{for } t = 1, 2, \dots, T \text{ and } j = 1, 2, \dots, J \quad (8)$$

In equations (5) and (8) the support spaces $\mathbf{z}_k$ and $\mathbf{v}_t$ are chosen to span the relevant parameter spaces for each β_k and u_k , respectively. As for the determination of support bounds for disturbances, Golan et al. (1996) recommend using “the three-sigma rule” of Pukelsheim (1994) to establish bounds on the error components: the lower bound is $v_L = -3\sigma_y$ and the upper bound is $v_U = 3\sigma_y$, where σ_y is the (empirical) standard deviation of the sample $\mathbf{y}$.

Under this reparametrization, the inverse problem with noise given in equation (2) may be rewritten as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} = \mathbf{XZp} + \mathbf{Vw} \quad (9)$$

Jaynes (1957) demonstrates that entropy is additive for independent sources of uncertainty. The details of this property can be found in Kapur and Kesavan (1992:31-32). Therefore, assuming the unknown weights on the parameter and the noise supports for the linear regression model are independent, we can jointly recover the unknown parameters and disturbances (noises or errors) by solving the constrained optimization problem of

$$\max_{\mathbf{p}, \mathbf{w}} H(\mathbf{p}, \mathbf{w}) = -\mathbf{p}' \ln \mathbf{p} - \mathbf{w}' \ln \mathbf{w} \quad (10)$$

$$\text{subject to } \mathbf{y} = \mathbf{XZp} + \mathbf{Vw} \quad (11)$$

Hence, given the reparameterization in equation (9) where β_k and u_t are transformed to have the properties of probabilities, in scalar notation the GME formulation for a noisy inverse problem may be stated as:

$$\max_{\mathbf{p}, \mathbf{w}} H(\mathbf{p}, \mathbf{w}) = \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln p_{km} - \sum_{t=1}^T \sum_{j=1}^J w_{tj} \ln w_{tj} \quad (12)$$

subject to the constraints:

$$\sum_{k=1}^K \sum_{m=1}^M x_{tk} z_{km} p_{km} + \sum_{j=1}^J w_{tj} v_{tj} = y_t \quad \text{for } t = 1, 2, \dots, T \quad (13)$$

$$\sum_{m=1}^M p_{km} = 1 \quad \text{for } k = 1, 2, \dots, K \quad (14)$$

$$\sum_{j=1}^J w_{tj} = 1 \quad \text{for } t = 1, 2, \dots, T \quad (15)$$

where equation (13) is the data (or consistency) constraint whereas equations (14) and (15) provide the required adding-up constraints for probability distributions of p_{km} and w_{tj} , respectively. The solution for $\hat{p}_{km}$ is:

$$\hat{p}_{km}^{GME} = \frac{e^{-\sum_{t=1}^T \hat{\lambda}_t z_{km} x_{tk}}}{\Omega_k^p(\hat{\lambda}_t)} \quad \text{where } \Omega_k^p(\hat{\lambda}_t) = \sum_{m=1}^M e^{-\sum_{t=1}^T \hat{\lambda}_t z_{km} x_{tk}} \quad (16)$$

The solution for $\hat{w}_{tj}$

$$\hat{w}_{tj}^{GME} = \frac{e^{-\hat{\lambda}_t v_{tj}}}{\Omega_k^w(\hat{\lambda}_t)} \quad \text{where } \Omega_k^w(\hat{\lambda}_t) = \sum_{m=1}^J e^{-\hat{\lambda}_t v_{tj}} \quad (17)$$

Notice that, in the expressions above, $\hat{\lambda}_t$ represent the dual value of data constraint. Substituting the solutions of $\hat{p}_{km}$ and $\hat{w}_{tj}$ into (3) and (6) produces the GME estimates of β_k and u_t , as:

$$\hat{\beta}_k^{GME} = \sum_{m=1}^M \hat{p}_{km} z_{km} \quad \text{for } k = 1, 2, \dots, K \quad (18)$$

and

$$\hat{u}_k^{GME} = \sum_{j=1}^J \hat{w}_{tj} v_{tj} \quad \text{for } t = 1, 2, \dots, T \quad (19)$$

As can seen, the GME estimates depend on the optimal Lagrange multipliers $\hat{\lambda}_t$ for the model constraints. There is no closed-form solution for $\hat{\lambda}_t$, and hence no closed form solution for $\mathbf{p}$, $\mathbf{w}$, β and $\mathbf{u}$. Therefore numerical optimization techniques should be used to obtain the solutions and solutions must be found numerically.

3.2 The generalized cross entropy (GCE) approach

If, in addition to the moment constraints, we have some non-sample information about the signals in the form of prior probabilities p_{mn}^0 , then an equivalent problem is to minimize the informational distance or Cross Entropy between the prior and the posterior probabilities (Kullback, 1959; Good, 1963). The Cross Entropy is defined as:

$$CE = \sum_j p_j \log \left(\frac{p_j}{p_j^0} \right) \quad (20)$$

if p_j^0 are the priors. Therefore, given prior probabilities $\mathbf{p}^0$, the resulting constrained optimization problem is to:

$$\min_{\mathbf{p}} CE(\mathbf{p}, \mathbf{p}^0) = \sum_{n=1}^N \mathbf{p}'_n \log \left(\frac{\mathbf{p}_n}{\mathbf{p}_n^0} \right) = \sum_{mn} p_{mn} \log \left(\frac{p_{mn}}{p_{mn}^0} \right) \quad (21)$$

subject to:

$$\sum_m p_{mn} = 1 \quad \forall n \text{ and } p_{mn} > 0 \quad \forall m, n \quad (22)$$

$$\sum_n x_{kn} y_n = \sum_n x_{kn} \mathbf{z}'_n \mathbf{p}_n \quad \forall k \in K \quad (23)$$

If the prior probabilities are assumed to be uniform, then the ME formulation results. The optimization problem can be solved numerically and each set of probabilities can be computed as non-linear functions of $\mathbf{x}_n$, $\mathbf{z}_n$, $\mathbf{p}_n^0$ and $\hat{\lambda}$. Here $\hat{\lambda}$ are a set of K Lagrange multipliers corresponding to the moment constraints resulting from solving the optimization problem.

3.2.1 Nonspherical errors

GCE with heteroscedastic errors In the sections above, a fixed weight of $\frac{1}{N}$ were applied to each of the observations in the sample. To allow for heteroscedasticity of an unspecified form, we replace this fixed weight by an unknown weight π_n that is allowed to vary across individuals. The π_n are an additional set of probabilities that must be estimated. π_n have to meet the following conditions, $\pi_n > 0 \quad \forall n$ and $\sum_{n=1}^N \pi_n = 1$. The resulting heteroscedastic consistent flexible moment constraints can be written as:

$$\sum_{n=1}^N x_{kn} y_n = \sum_{n=1}^N x_{kn} \mathbf{z}'_n \mathbf{p}_n + \sum_{n=1}^N x_{kn} \pi_n \mathbf{v}'_n \mathbf{w}_n \quad \forall k = 1, 2, \dots, K \quad (24)$$

with the added requirement that $\sum_{n=1}^N \pi_n = 1$. Assuming independence between π_n and w_{mn} , we can define an auxiliary joint probability $q_{mn} = w_{mn} \cdot \pi_n$ so that $\pi_n = \sum_m q_{mn} \quad \forall n$ and $w_{mn} = \frac{q_{mn}}{\sum_n q_{mn}} \quad \forall m, n$, are marginal and conditional probabilities (respectively) derivable from q_{mn} . In addition, since $\sum_n \pi_n = 1$ and $\sum_m w_{mn} = 1 \quad \forall n$, then $\sum_m q_{mn} = 1$ over all m and n . Finally, given the above definition of q_{mn} , we may specify its associated priors as $q_{mn}^0 = \pi_n^0 \cdot w_{mn}^0$. The constrained optimization problem that incorporates all available knowledge is:

$$\max H(\mathbf{p}, \mathbf{p}^0, \mathbf{q}, \mathbf{q}^0) = -\mathbf{p}' \ln \left(\frac{\mathbf{p}}{\mathbf{p}^0} \right) - \mathbf{q}' \ln \left(\frac{\mathbf{q}}{\mathbf{q}^0} \right) \quad (25)$$

$$\text{subject to } \mathbf{y} = \mathbf{XZp} + \mathbf{Vq} \quad (26)$$

$$\sum_l p_{ln} = 1 \quad \forall n \quad (27)$$

$$\sum_{mn} q_{mn} = 1 \quad (28)$$

GCE with autocorrelated errors The heteroscedasticity consistent formulation of the GCE problem can be seen as a special case of a more general formulation that not only allows errors shrinkage rates to be determined endogenously, but also allows the optimal re-weighted errors to combine with each other in order to create signal distortion. The constraints in case of heteroscedasticity can be seen as a special case of the constraints:

$$\mathbf{y} = \mathbf{XZp} + \mathbf{AVq}$$

where $\mathbf{A}$ is a row-standardized hypothesized error structure matrix.

To extend that the off-diagonal elements in $\mathbf{A}$ are allowed to be non-zero, the optimal re-weighted errors are allowed to combine while the distorting the signal. The matrix $\mathbf{A}$ is a row-standardized version of a spatial link, say $\mathbf{A}^*$, which can be specified in a number of different ways. To code heteroscedasticity as well as local first-order autocorrelation, we can define $\mathbf{A}^* = \mathbf{I} + \mathbf{C}$ where $\mathbf{C}$ is a first order spatial contiguity. That is:

$$a_{nj}^* = \begin{cases} 1 & \forall j = n \\ 1 & \forall j \in J_n \\ 0 & \text{for all other } j \end{cases} \quad (29)$$

where $j \in J_n$ is taken to read “all units within the neighborhood of the n th unit”. In order to include distance-based dependence for local neighbors (based on contiguity alone), we can define this matrix as $\mathbf{A}^* = \mathbf{I} + \exp(-\mathbf{D}) \odot \mathbf{C}$ where $\odot$ represents an element-by-element matrix manipulation, $\mathbf{D}$ represents an $N \times N$ matrix of distances between all pairs of data points, and $\mathbf{C}$ is defined above. Finally, allowing global dependence, albeit with some distance-based decay, can be represented by setting $\mathbf{A}^* = \exp(-\mathbf{D})$. Row-standardizing $\mathbf{A}^*$ generally yields an asymmetric matrix, i.e. $\mathbf{A}' \neq \mathbf{A}$.

3.3 Specifying the support space

In the case of binary choices, there exist natural bounds for both signal as well as the noise terms: the observed and expected outcomes in this case can only exist between 0 and 1. This means that the signals are naturally bounded by 0 and 1, i.e. $z_l \in (0, 1)$. A simple specification would be $\mathbf{z} = (0, 1)'$. Now, if we observe an outcome (i.e. $y_n = 1$) but predict it as being nearly impossible (i.e. $\hat{s}_n \approx 0$), then the error be as high as +1. Or, if the binary choice is not observed

(i.e. $y_n = 0$) but we predict it with near certainty (i.e. $\hat{s}_n \approx 1$), then the error can be as low as -1 . In others words, the errors are also naturally bounded between ± 1 i.e. $v_m \in \pm 1$ and a simple specification would be $\mathbf{v} = (1, 1)'$. If we specify the support spaces as described above and create noiseless moment constraints of:

$$\sum_n x_{kn} y_n = \sum_n x_{kn} \mathbf{z}'_n \mathbf{p}_n \quad \forall k = 1, 2, \dots, K \quad (30)$$

then the resulting Maximum Entropy solutions are identical to the Logit parameters. In fact, under this specification the Maximum Entropy dual objective function turns out be identical to the Logit Log-likelihood function. As such, all inferences derived from it, including the parameter estimates and their covariance matrix, are identical to those that would be recovered from the Logit model.

Here we are confronted with count data. Count outcomes can be thought of as a summation over a large but finite sequence of independent and identical binary choices. That is the motivation underlying a Binomial distribution and the Poisson distribution is, in fact, obtained at the limit when the number of binary choices in the sequences approach ∞ .

4 Hypothesis and specification tests

4.1 Hypothesis tests

In this section, we describe the large sample properties of the GME-GCE solutions. Although the GME-GCE solution does not have a closed-form, the dual formulation of the problem may be used to evaluate the behavior of the solutions within the context of extremum or M-estimators (Huber, 1981; Newey and McFadden, 1994). Therefore in addition to estimating the parameters, we use the dual objective function to estimate a covariance matrix for $\hat{\lambda}$. The covariance matrix for the Lagrange multipliers is defined as the inverse negative Hessian of the dual objective function evaluated at the optimal values of the Lagrange multipliers :

$$\widehat{\Sigma}_{\lambda} = \left\{ -\frac{\partial^2 L_{GCE}^D}{\partial \hat{\lambda} \partial \hat{\lambda}'} \right\}^{-1} \quad (31)$$

The square root of the elements of the above matrix are the estimated standard errors for the Lagrange multipliers. Another quantity of interest is the marginal effects of the independent variables. This quantity is evaluated at the sample mean of the predictors. In our case, the marginal effects are computed as:

$$\hat{\lambda}_k = \frac{\partial \hat{s}_*}{\partial \hat{x}_{k*}} = \hat{\lambda} \{ \mathbf{z}^{2'} \hat{\mathbf{p}}_* - (\mathbf{z}' \mathbf{p}_*)^2 \} \quad (32)$$

which are non-linear functions of the underlying Lagrange multipliers. Delta method is used in order to compute the covariance matrix of $\hat{\lambda}$ (Greene, 2000). Given the definition of $\hat{\gamma}$, the covariance matrix is written:

$$\frac{\partial \hat{\gamma}}{\partial \hat{\lambda}'} = \{\mathbf{z}^{2'} \hat{\mathbf{p}}_* - (\mathbf{z}' \hat{\mathbf{p}}_*)^2\} \cdot I + \{\mathbf{z}^{2'} \hat{p}_* + (\mathbf{z}^{2'} \hat{\mathbf{p}}_*)(\mathbf{z}' \hat{\mathbf{p}}_*) - 2(\mathbf{z}' \hat{\mathbf{p}}_*)^3\} \cdot \hat{\lambda} \mathbf{x}_*' \quad (33)$$

Using the optimized values of the objective functions, an Entropy Ratio (ER) test that is analogous to a Likelihood Ratio test is constructed to test joint hypothesis on the Lagrange multipliers. This test is defined as follow:

$$ER_R = 2\{\tilde{L}_{GCE}^D - \hat{L}_{GCE}^D\} \sim \chi_R^2 \quad (34)$$

where $\tilde{L}_{GCE}^D$ and $\hat{L}_{GCE}^D$ are respectively the optimized values of the objective function for the restricted and unrestricted models and R corresponds to the number of restrictions (Jaynes, 1979, page 67).

4.2 Specification tests

In this section, we describe how to choose among several non-nested models. Consider the model M_0 that is encompassed by the model M_1 .

The GCE models described above attempts at capturing the structure in the errors non-parametrically. If the structure hypothesized in $\mathbf{A}$ is a good approximation of reality, then must help us gain information about the error structure without giving up too much information about the signals. So the relative gain in error information is defined as:

$$H_e = \frac{H(\hat{\mathbf{q}}_0)}{H(\hat{\mathbf{q}}_1)} \quad (35)$$

where $H(\hat{\mathbf{p}}_0)$, $H(\hat{\mathbf{q}}_0)$, M_0 , $H(\hat{\mathbf{p}}_1)$ and $H(\hat{\mathbf{q}}_1)$ are respectively the quantity of uncertainty about the signal and noise term in the model M_0 and in the model M_1 . The relative loss in signal information is:

$$H_e = \frac{H(\hat{\mathbf{p}}_1)}{H(\hat{\mathbf{p}}_0)} \quad (36)$$

In order to compare if the flexibility in a given model M_1 over that provided in model M_0 is worthwhile, we compute a composite ratio as follows :

$$H_* = \frac{H_e}{H_s} \quad (37)$$

5 Data and variables

We have constructed our sample by merging three databases. The first database is the French annual firm research expenditures survey which was conducted in 2002. This survey is carried out by the Ministry of Research. It concerns the internal expenditure of research, that is to say R&D executed by the firm itself. It focuses on all the firms (having more than 20 employees) which carry out some R&D and employ at least one full time researcher. The location (region and department) is subjected to a systematic coding with a ZIP code which

allows to identify the city. This database is then appropriate to deal with the question of local dimension of technological externalities. For our study, we use the information on the firm total R&D expenditures, turnover and the sectoral decomposition. We also retain as our innovation variable the total numbers of patents granted by the firm during the year. The total numbers of patents granted is often viewed as a more appropriate measure of innovation output.

The second database we use is the Firm Annual Survey. It provides financial and accounting informations such as sales of firms, value added, capital assets... This database allows us to compute the firm's market share.

The third database is the 1999 French communal database. This file provides us the latitudinal and longitudinal coordinates which allow us to calculate the distance and contiguity matrix. The distance between two communes is calculated starting from the coordinates of the administrative center of the communes.

Innovation output is explained by the following regressors. R&D expenditures per employees, intra-sectorial communal R&D; inter-sectorial communal R&D, market share³ and sectoral effects in the sector i and in the commune j . Inter-sectorial R&D expenditures communal R&D expenditures is the difference between the communal R&D and intra-sectorial R&D. The market share of firm i is computed as the turnover of firm i in the industry k divided by the total of sales in sector k . We have only considered the domestic decomposition turnover at the "NAF level 700" of the French industrial classification. Market share is computed on all firms available in all our continuous variables are taken in logarithm. By merging these three databases, we obtain a sample of 1 566 French industrial firms. Some descriptive statistics appear in table 1 (in appendix page 17).

6 Results

We estimate five alternate models, corresponding to the five types of error structure we have previously described. Model I most closely corresponds to the baseline Poisson regression specification as $A^* = 0$. Model II allows for heteroskedasticity only. Models III, IV, V allow for heteroskedasticity, and respectively, first-order local error-correlation, first-order local error-correlation with distance-based decay and global error-correlation with distance based decay. The results are presented at the end of the article in tables 5 to 9. The comparison of the results shows that the various forms of flexibility afforded to the basic model lead to different parameter values. The reliability of predictors sometimes changes considerably across the specifications. In order to select from the various specification, as Bhati (2004) we computed the composite relative information/loss measures for each of the models. Of course, it is not possible to have them for Model I as it is never the alternate model. For model II, the encompassed model is the model I, for the three others, it is the model II. Model II is less desirable than model I as $H^* = 0,80 < 1$. In our case, allowing heteroskedasticity alone does not seem to be appropriate as Model I

³Only in the specification II.

is more desirable. Models III, IV are more desirable than model II as $H^* > 1$. The comparison of these indicators show that the model IV is largely preferred as “*the gain in information about the noise component outweighs the loss in information about the signal by a much larger proportion*” (Bhati, 2004). We conclude that “*a first-order local error-correlation with distance-based decay is the closest approximation to the underlying data generating process*”.

Now that we have identified the appropriate model, we can present the results. First, without surprise, the R&D realized inside the firm has an impact on its patent production. Firm’s size⁴ increase the number of patents obtained by a firm. In order to check the Schumpeterian hypothesis, we have introduced the market share in the regression. Our results shows that the higher is the market share, the less the firm innovate. The size of the commune has a negative and significant impact on the innovation process. When firms are localized in the Paris region, firms seem to be more innovating. We control sector-based effect in introducing dummy sector variables. The reference sector is “Manufacture of clothing articles and leather products”. All sectorial dummies variables are significant except for the sector “Publishing, printing and reproduction of recorded media”. If this result show the importance to introduce sectoral effects in the regression, it is not sufficient. So we have chosen to distinguish intra-sector based spillovers from inter-sector-based ones in order to test the role of the specialization vs. the diversity of the territory. It appears that intra-sector-based R&D in the city where firm is located has a negative impact, whereas intra-sector-based R&D done in proximate cities has a positive impact (at the 6% threshold however). Thus, the study of intra-sector based spillovers shows that their impact is not as obvious and clear as the previous studies could pretend. Indeed, we show that there is a double effect. There is a competitive effect at the very proximate level. The proximity of other firms from the same sector has a negative impact on the patent production of the firm. However, when the proximity is weaker, but still strong (as the surrounding cities) the competitive effect disappears to let the place to a spillover effect. Thus, at a very local level, we find the same result that Audrestch and Feldman (1999) on the negative effect of specialized cities. But at a further level, we reach the same conclusion than Jaffe (1989) on the importance of spillover between local firms of the same sector. It seems then that the two effects co-exist inside local area. This result confirms that in cluster, the two effects will co-exist and will likely contribute to the growth, even if we can test here. Inter-sector-based R&D realized by firms located in the same city or in the neighbourhood has a positive impact on patent production. However, surprisingly, when distance increases, the inter sectoral RD spillovers increase.

The result confirm that geographical proximity matters and that it is useful to distinguish two local levels. However, the surprising result about distant sectoral RD spillovers can be due to the structure of the local inter sectoral RD which can be very weak in the same city and much more in the surrounding. This result shows nevertheless the importance of inter-sectorial effects. Finally, if we compare intra and inter-sectorial R&D’s parameter values, we see that the

⁴Firm’s size is measured by the number of employee.

inter-sectorial spillovers are much more important than intra-sectorial spillovers. Diversity would be more important than specialization in terms of spillovers. These results go in the same way than Autant-Bernard and Massard (2004).

7 Conclusion

Our study had as objective to measure the spillovers effects of intra and inter-sector-based R&D on the patent production of the firms. The question was double: are there spillovers? And what is their spatial dimension? For that, contrary to previous study, we were able to test our model on individual data and to work at the city level as local area rather than French department or region. This level of analysis brings two interesting results in the study of spillovers. In one hand, we show that there is a competition effect in very close neighbourhood and a spillover effect in further neighbourhood. These results underline the negative effect of being located too close from competitors but the interest of being not too far. That suggests the co-existence of the two effects inside clusters. This is a relatively new result in studies on spatial spillovers. Our study shows then the importance of the geographic level of analysis. It is then very important to continue study at this very small geographic level. In another hand, there are strong inter-sector-based spillovers whose impact increases rapidly with distance. Very short proximity matters. This study has thus some implications for public policy. It clearly shows that inter-sector-based firms concentration should be preferred to intra-sectorial firms concentration. This goes in the way of inter-sector-based research, which is being developed nowadays. The “competitive pole”, in France, should go in this direction. More precisely, the public policy should encourage the geographical concentration of firms from different sectors and the implantation of firms of the same sector in the surrounding of the pole. However, it is difficult to define what is the bordering between intra and inter-sector-based R&D. Our first results are very interesting however, further work should be considered. The Poisson estimation model is characterised by only one parameter, which implies the equality between the conditional mean and variance. This assumption of equidispersion is very restrictive. It is frequent that the variance be superior to the mean. Further work will require to use negative binomial model of type I (i.e. the dispersion is function of the expected mean) (Hausman et al., 1984). Indeed, thanks to the introduction of a supplementary parameter (α), negative binomial distribution is richer than Poisson distribution. This parameter collects the unobserved heterogeneity of the explained variable. This model allows for an over-dispersion (i.e. the variance is greater than the mean).

References

- ANSELIN, L., A. VARGA, AND Z. ACS (1997): "Local geographic spillovers between university research and high technology innovations," *Journal of Urban Economics*, 42(3), 422–448.
- AUDRETSCH, D. B., AND M. P. FELDMAN (1996): "The American Economic Review," *Knowledge spillovers and the geography of innovation and production*, 86(3), 630–640.
- AUTANT-BERNARD, C. (2001): "The geography of knowledge spillovers and technological proximity," *Economics of Innovation and New Technology*, 10(4), 237–254.
- AUTANT-BERNARD, C., AND N. MASSARD (2004): "Disparités locales dans la production d'innovations : L'incidence du choix des indicateurs," Fourth Proximity Congress, Marseille.
- BOTTAZZI, L., AND G. PERI (2003): "Innovation and spillovers in regions: Evidence from European patent data," *European Economic Review*, 47(4), 687–710.
- FELDMAN, M. (1994): *The geography of innovation*. Kluwer Academic Publishers, Boston, 154 pages.
- GLAESER, E., H. KALLAL, J. SCHEINKMAN, AND A. SHEIFLER (1992): "Growth of Cities," *Journal of Political Economy*, 100, 1126–1152.
- GOLAN, A., G. JUDGE, AND D. MILLER (1996): *Maximum Entropy Econometrics : Robust Estimation with Limited Data*.
- GOOD, I. J. (1963): "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," *Annals of Mathematical Statistics*, 34, 911–934.
- GREENE, W. (2000): *Econometric Analysis*. Prentice-Hall, 4 edn.
- GRILICHES, Z. (1979): "Issues in Assessing the Contribution of R&D to Productivity Growth," *The Bell Journal of Economics*, (10), 92–116.
- HUBER, P. (1981): *Robust Statistics*. John Wiley.
- JACOBS, J. (1969): *The Economy of Cities*. London: Jonathan Cape.
- JAFFÉ, A. (1989): "The real effects of academic research," *The American Economic Review*, 79(5), 957–970.
- JAYNES, E. T. (1957): "Information Theory and Statistical Mechanics," *Physics Review*, 106, 620–630.
- JAYNES, E. T. (1979): "Where do we stand on Maximum Entropy?," in *The Maximum Entropy Formalism*, ed. by R. D. Levin, and M. Tribus, pp. 15–118. The MIT Press, Cambridge MA.

- KAPUR, J. N., AND H. K. KESAVAN (1993): *Entropy Optimization Principles with Applications*. Academic Press, New York.
- KULLBACK, J. (1959): *Information Theory and Statistics*. John Wiley and Sons.
- LE BLANC, G. (2004): “Regional Specialization, Local Externalities and Clustering in Information Technology Industries,” in *Knowledge Economy, Information Technologies and Growth*, ed. by L. Paganetto, and V. T. Burlington, pp. 453–486. Aldershot: Ashgate.
- MARSHALL, A. (1920): *Principles of Economics*. London: MacMillan.
- MORAN, P. (1948): “The interpretation of statistical maps,” *Journal of the Royal Statistical Society*, 59, 185–193.
- NEWBY, W. K., AND D. L. MCFADDEN (1994): “Large sample estimation and hypothesis testing,” in *Handbook of Econometrics*, ed. by R. F. Engle, vol. 4. Elsevier.
- SHANNON, C. (1948): “A mathematical theory of communication,” *Bell System Technical Journal*, 27, 379–423.

Appendix

Moran test

To test the existence of spatial autocorrelation, several indicators exist (Anselin, 1988). One of the oldest and best known is Moran's I indicator (Moran, 1948) defined as follow:

$$\text{Moran's I} = \frac{\frac{\sum_{i \neq j} c_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W}}{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

where $W = \sum_i c_{ij}$. The numerator is the covariance between contiguity observations (each contiguity weight $\frac{c_{ij}}{W}$). This covariance is null if there is no spatial effect, positive if there is positive spatial autocorrelation and negative if there is negative spatial autocorrelation. The covariance is normalized using the total variance of the series (denominator). The value of Moran's I is interpreted as follow: if they range from -1 to 0, there is negative spatial autocorrelation, if it is 0, there is a random distribution of the variable; if they range from 0 to 1, there is a positive spatial autocorrelation. Using this indicator requires previously assuming some null hypothesis about the lack of spatial autocorrelation. To test whether the occurrence of an event in an area follows some kind of systematic spatial pattern, this distribution can be compared with a random pattern distribution.

Tables of descriptives statistics and results

Table 1: Descriptives statistics

	Mean	Min	25%	50%	75%	Max
Firm level variables						
Number of patents	2.66	0	0	0	1	454
R&D expenditures/turnover	71.30	5.85E-05	8.98E-03	2.38E-02	6.39E-02	3.26E+04
Communal level variables						
Communal R&D expenditures	18 997.93	0	0	439,96	8 242.03	490 224.80
Intra sectorial R&D expenditures	237.58	0	0	0	0	48 103
Inter sectorial R&D expenditures	26 829.42	33.00	1 073.80	3 843.70	21 231.67	1 171 123.00
Source : Ministère de la Recherche and INSEE.						

Table 2: Percentage of firms by sector

Sector	Percentage of firm	R&D expenditure* per employee	R&D expenditure*	Mean of patent number
C1	0.007	3.84	724.62	0.46
C2	0.003	6.12	461.11	0.63
C3	0.075	15.80	7 984.17	4.63
C4	0.054	12.25	3 654.63	2.81
D0	0.039	11.12	2 2492.54	12.07
E1	0.030	13.99	46 089.83	5.30
E2	0.180	7.16	1680.10	1.52
E3	0.151	19.93	12 876.83	2.12
F1	0.032	5.06	1 852.61	1.23
F2	0.031	7.14	1 278.02	0.42
F3	0.028	2.20	1 043.42	1.78
F4	0.180	10.05	5 126.78	2.05
F5	0.088	4.43	1 491.25	1.49
F6	0.086	15.08	14 395.36	3.66
G1	0.005	14.40	9 243.31	4.35
G2	0.003	1.40	60 703.29	6.33
*: in thousand of Euros.				
Source : Ministère de la Recherche.				

Table 3: Statistics of distance

	Mean*	Min*	Median*	Max*
distance	246.47	0	209.04	848.37
* : in kilometer.				
Source : INSEE.				

Table 4: Results for Moran's I

Variable	
Patents	0.0204 (0.0240)
R&D	0.0527 (0.0161)
P-value are in brackets.	
Source : Ministère de la Recherche.	

Table 5: Specification I: A=D0

Variables	Lambda	P-value	Marginal Effects	P-value
Intercept	-10.7928	0.00	-29.1637	0.00
RDT	0.7790	0.00	2.1050	0.00
R&D intra	-0.0205	0.00	-0.0553	0.00
R&D inter	0.0413	0.00	0.1116	0.00
Market share	0.2883	0.00	0.7789	0.00
Size (effectif)	0.6024	0.00	1.6277	0.00
Size of commune	-0.1527	0.00	-0.4125	0.00
Paris	-0.0664	0.19	-0.1794	0.19
R&D intra (lag)	0.0508	0.01	0.1373	0.01
R&D inter (lag)	-0.0091	0.30	-0.0245	0.30
NAF C1	ref.	ref.	ref.	ref.
NAF C2	0.9151	0.20	2.4727	0.20
NAF C3	0.9980	0.02	2.6968	0.02
NAF C4	0.6660	0.11	1.7997	0.11
NAF D0	1.6647	0.00	4.4983	0.00
NAF E1	-0.1844	0.66	-0.4982	0.66
NAF E2	0.4770	0.25	1.2889	0.25
NAF E3	-0.2722	0.51	-0.7356	0.51
NAF F1	-0.0577	0.89	-0.1559	0.89
NAF F2	-0.6634	0.15	-1.7927	0.15
NAF F3	1.1612	0.01	3.1379	0.01
NAF F4	0.3990	0.33	1.0782	0.33
NAF F5	0.6216	0.13	1.6796	0.13
NAF F6	0.2914	0.48	0.7875	0.48
NAF G1	0.1120	0.80	0.3026	0.80
NAF G2	-0.7087	0.11	-1.9150	0.11
Model Diagnostics				
L_{GCE}^D	1 453 650.2			
Pseudo R^{2*}	55.6			
H_e	1.2388			
H_s	1.0774			
H_*	1.1498			
Observations : 1 566				
* It is defined as the proportion of observed variance in the criterion measure explained by the predictors.				
Source : Ministère de la Recherche and INSEE.				

Table 6: Specification II: A=DH

Variables	Lambda	P-value	Marginal Effects	P-value
Intercept	-9.3054	0.00	-33.2002	0.00
RDT	0.5037	0.00	1.7973	0.00
R&D intra	-0.0192	0.00	-0.0684	0.00
R&D inter	0.0096	0.00	0.0341	0.00
Market share	0.1246	0.00	0.4446	0.00
Size (effectif)	0.4692	0.00	1.6741	0.00
Size of commune	-0.0714	0.00	-0.2547	0.00
Paris	-0.0018	0.96	-0.0064	0.96
R&D intra (lag)	0.0417	0.00	0.1488	0.00
R&D inter (lag)	0.0095	0.13	0.0338	0.13
NAF C1	ref.	rref.	ref.	ref.
NAF C2	0.4918	0.11	1.7547	0.11
NAF C3	0.4488	0.02	1.6014	0.02
NAF C4	0.2438	0.19	0.8698	0.19
NAF D0	1.0595	0.00	3.7801	0.00
NAF E1	-0.2451	0.20	-0.8747	0.20
NAF E2	0.1011	0.58	0.3606	0.58
NAF E3	-0.2121	0.25	-0.7569	0.25
NAF F1	-0.1190	0.54	-0.4246	0.54
NAF F2	-0.0723	0.71	-0.2580	0.71
NAF F3	0.3990	0.04	1.4235	0.04
NAF F4	0.0969	0.59	0.3459	0.59
NAF F5	0.1681	0.36	0.5998	0.36
NAF F6	0.0117	0.95	0.0416	0.95
NAF G1	0.1115	0.62	0.3979	0.62
NAF G2	-0.5669	0.01	-2.0226	0.01
Model Diagnostics				
L_{GCE}^D	1 444 565.1			
Pseudo R^{2*}	36.9			
H_e	1.14			
H_s	1.42			
H_*	0.80			
Observations : 1 566				
* It is defined as the proportion of observed variance in the criterion measure explained by the predictors.				
Source : Ministère de la Recherche and INSEE.				

Table 7: Specification III: A=DH1C

Variables	Lambda	P-value	Marginal Effects	P-value
Intercept	-9.1473	0.00	-36.5942	0.00
RDT	0.4870	0.00	1.9481	0.00
R&D intra	-0.0210	0.00	-0.0841	0.00
R&D inter	0.0188	0.00	0.0752	0.00
Market share	0.1492	0.00	0.5967	0.00
Size (effectif)	0.4336	0.00	1.7347	0.00
Size of commune	-0.0437	0.00	-0.1750	0.00
Paris	0.2490	0.00	0.9961	0.00
R&D intra (lag)	0.0264	0.05	0.1054	0.05
R&D inter (lag)	0.0097	0.12	0.0387	0.12
NAF C1	ref.	ref.	ref.	ref.
NAF C2	0.0805	0.84	0.3222	0.84
NAF C3	0.6015	0.00	2.4062	0.00
NAF C4	0.3121	0.14	1.2485	0.14
NAF D0	1.1111	0.00	4.4448	0.00
NAF E1	-0.1328	0.53	-0.5312	0.53
NAF E2	0.1341	0.52	0.5364	0.52
NAF E3	-0.1044	0.61	-0.4176	0.61
NAF F1	-0.0487	0.82	-0.1950	0.82
NAF F2	-0.2157	0.34	-0.8630	0.34
NAF F3	0.5353	0.02	2.1413	0.02
NAF F4	0.1448	0.48	0.5794	0.48
NAF F5	0.2208	0.29	0.8832	0.29
NAF F6	0.0392	0.85	0.1570	0.85
NAF G1	0.0116	0.96	0.0464	0.96
NAF G2	-0.6398	0.01	-2.5594	0.01
Model Diagnostics				
L_{GCE}^D	1 443 337.3			
Pseudo R^{2*}	40.8			
H_e	1.26			
H_s	1.09			
H_*	1.15			
Observations : 1 566				
* It is defined as the proportion of observed variance in the criterion measure explained by the predictors.				
Source : Ministère de la Recherche and INSEE.				

Table 8: Specification IV: A=DH1D

Variables	Lambda	P-value	Marginal Effects	P-value
Intercept	-11.2927	0.00	-141.9372	0.00
RDT	0.0299	0.00	0.3760	0.00
R&D intra	-0.0169	0.00	-0.2126	0.00
R&D inter	0.0321	0.00	0.4034	0.00
Market share	-0.0191	0.01	-0.2397	0.01
Size (effectif)	0.3348	0.00	4.2085	0.00
Size of commune	-0.1044	0.00	-1.3124	0.00
Paris	0.5146	0.00	6.4679	0.00
R&D intra (lag)	0.0247	0.00	0.3106	0.00
R&D inter (lag)	0.3691	0.00	4.6397	0.00
NAF C1	ref.	ref.	ref.	ref.
NAF C2	-0.1546	0.80	-1.9438	0.80
NAF C3	1.6357	0.00	20.5591	0.00
NAF C4	2.0121	0.00	25.2899	0.00
NAF D0	2.2853	0.00	28.7238	0.00
NAF E1	1.4672	0.00	18.4411	0.00
NAF E2	0.2944	0.08	3.6999	0.08
NAF E3	0.9835	0.00	12.3613	0.00
NAF F1	1.9662	0.00	24.7132	0.00
NAF F2	-1.0650	0.00	-13.3862	0.00
NAF F3	1.2027	0.00	15.1171	0.00
NAF F4	1.5627	0.00	19.6409	0.00
NAF F5	1.6947	0.00	21.3004	0.00
NAF F6	1.0522	0.00	13.2249	0.00
NAF G1	2.1279	0.00	26.7453	0.00
NAF G2	-1.5644	0.00	-19.6631	0.00
Model Diagnostics				
L_{GCE}^D	1 407 473.2			
Pseudo R^{2*}	-250.5			
H_e	10.44			
H_s	2.29			
H_*	4.41			
Observations : 1 566				
* It is defined as the proportion of observed variance in the criterion measure explained by the predictors.				
Source : Ministère de la Recherche and INSEE.				

Table 9: Specification V: A=DHGD

Variables	Lambda	P-value	Marginal Effects	P-value
Intercept	-9.2999	0.00	-33.3450	0.00
RDT	0.5271	0.00	1.8898	0.00
R&D intra	-0.0193	0.00	-0.0692	0.00
R&D inter	0.0106	0.00	0.0380	0.00
Market share	0.1467	0.00	0.5260	0.00
Size (effectif)	0.4814	0.00	1.7260	0.00
Size of commune	-0.0830	0.00	-0.2977	0.00
Paris	-0.0499	0.17	-0.1791	0.17
R&D intra (lag)	0.0456	0.00	0.1635	0.00
R&D inter (lag)	0.0093	0.15	0.0332	0.15
NAF C1	ref.	ref.	ref.	ref.
NAF C2	0.5432	0.11	1.9478	0.11
NAF C3	0.4759	0.02	1.7065	0.02
NAF C4	0.2421	0.24	0.8679	0.24
NAF D0	1.0637	0.00	3.8140	0.00
NAF E1	-0.3072	0.14	-1.1016	0.14
NAF E2	0.0930	0.64	0.3334	0.64
NAF E3	-0.2660	0.19	-0.9536	0.19
NAF F1	-0.1409	0.51	-0.5051	0.51
NAF F2	-0.0936	0.67	-0.3357	0.67
NAF F3	0.4280	0.05	1.5347	0.05
NAF F4	0.0929	0.64	0.3331	0.64
NAF F5	0.1607	0.43	0.5764	0.43
NAF F6	0.0017	0.99	0.0061	0.99
NAF G1	0.0306	0.90	0.1098	0.90
NAF G2	-0.6685	0.01	-2.3969	0.01
Model Diagnostics				
L_{GCE}^D	1 444 664.9			
Pseudo R^{2*}	39.3			
H_e	0.9958			
H_s	0.9960			
H_*	0.9999			
Observations : 1 566				
* It is defined as the proportion of observed variance in the criterion measure explained by the predictors.				
Source : Ministère de la Recherche and INSEE.				

Table 10: Sector of activities

Code	Denomination
C1	Manufacture of clothing articles and leather products
C2	Publishing, printing and reproduction of recorded media
C3	Manufacture of pharmaceuticals products, perfumes, soap and cleaning preparation
C4	Manufacture of domestic equipment
D0	Manufacture of motor vehicles
E1	Building of ships and boats, manufacture of railway locomotives, rolling stock
E2	Manufacture of metal products, machinery and equipment
E3	Manufacture of electric and electronic equipment
F1	Mining and quarrying except energy producing materials, manufacturing of other non-metallic mineral products
F2	Manufacture of textiles
F3	Manufacture of wood, wood products, pulp, paper and paper products
F4	Manufacture of chemicals, rubber, plastic and chemical products
F5	Manufacture of basic metals and fabricated metal products
F6	Manufacture of electric and electronic components
G1	Extraction of coal, crude petroleum, gas and uranium; manufacture of coke, refined petroleum products, and nuclear fuel
G2	Electricity and gas supply; collection and distribution of water
Source : INSEE.	