# Spatial structures and spatial spillovers: a generalized maximum entropy approach

**Esteban Fernández Vázquez[1], Matías Mayor Fernández[1] and Jorge Rodríguez Vález[2]**

[1]University of Oviedo, Department of Applied Economics, Faculty of Economics, Campus del Cristo, Oviedo, 33006, Spain. e-mail: evazquez@uniovi.es; mmayorf@uniovi.es

[2]University of León and The Lawrence R. Klein Centre at Autónoma de Madrid University, Spain. e-mail: Jorge.Rodriguez@unileon.es

**ABSTRACT**:
Spatial econometric methods measure spatial interaction and incorporate spatial structure into regression analysis. The specification of a matrix of spatial weights W plays a crucial role in the estimation of spatial models. The elements $w_{ij}$ of this matrix measure the spatial relationships between two geographical locations i and j, and they are specified exogenously to the model. Several alternatives for W have been proposed in the literature, although binary matrices based on contiguity among locations or distance matrices are the most common choices. One shortcoming of using this type of matrices for the spatial models is the impossibility of estimating "heterogeneous" spatial spillovers: the typical objective is the estimation of a parameter that measures the average spatial effect of the set of locations analyzed. Roughly speaking, this is given by "ill-posed" econometric models where the number of (spatial) parameters to estimate is too large. In this paper, we explore the use of generalized maximum entropy econometrics (GME) to estimate spatial structures. This technique is very attractive in situations where one has to deal with estimation of "ill-posed" or "ill-conditioned" models. We compare by means of Monte Carlo simulations "classical" ML estimators with GME estimators in several situations with different availability of information.

**Keywords:** spatial econometrics, generalized maximum entropy econometrics, spatial spillovers, Monte Carlo simulations.

## 1. INTRODUCTION

Spatial econometrics is a subdiscipline that has gained a huge popularity in the last twenty years, not only in theoretical econometrics but in empirical studies as well. Basically, spatial econometric methods measure spatial interaction and incorporate spatial structure into regression analysis. On the one hand, the literature shows several methodological suggestions for including spatial relationships in econometric regression models. In the early 1980s Cliff and Ord (1973,1981) already provided an introduction to hypotheses testing and models of spatial process. Later, Anselin (1988) studied the performance of various estimators of spatial econometric models like least squares (LS), maximum likelihood (ML) which was first outlined by Ord (1975), instrumental variable (IV), and method of moment (MM). More recently, the generalized two-stage least squares (2SLS) and generalized moments method (GMM) have been examined by Kelejian and Prucha (1998, 1999). On the other hand, its empirical applications to several fields of economic analysis have mushroomed lately including, among others, studies in demand analysis international economics, labor economics, public economics and local public finance and agricultural and environmental economics.

Although there are other approaches to address the spatial interactions in an econometric model, the most common procedure followed in the literature is to specify a determined spatial structure by means of a spatial lag operator (Anselin, 1988). In this point is where the specification of a matrix **W**, with elements $w_{ij}$ plays a very important role. Each cell $w_{ij}$ of this matrix measures the spatial interaction between the locations $i$ and $j$ and, roughly speaking, can be interpreted as the influence that a variable located in $j$ has over other variable located in $i$. [1] It is crucial to note that the values of these elements are fixed exogenously to the model; in other words, the **W** matrix is imposed by the researcher somehow.

Various possibilities have been suggested to define **W**, although most generally they are based on some concept of geographical proximity. Following this approach, a very simple way to characterize the elements $w_{ij}$ is by defining them as binary variables that take value 1

---

[1] Most usually it is assumed that $w_{ii} = w_{jj} = 0$. Another frequent issue is that the elements $w_{ij}$ are row-standardized, and consequently $\sum_j w_{ij} = 1$. It also ensures that the spatial parameters are comparable between models because of the spatial autoregressive parameters must be constrained to the interval $1/\omega_{min}$ y $1/\omega_{max}$ where $\omega_{min}$ and $\omega_{max}$ are the smallest and largest eigenvalues of W.

when locations $i$ and $j$ have a common border and 0 otherwise. The geographical distance between locations $i$ and $j$ ($d_{ij}$) can be used in a more direct way, defining $w_{ij}$ as a function of this distance $w_{ij}(d_{ij})$, with the first derivative being negative, $w'_{ij}(d_{ij}) < 0$. Other authors claim for using not physical but economic measures of distance, based on interregional trade flows, income differences, etc.[2]

Once the values $w_{ij}$ are a priori imposed by the researcher, they are employed together with the data of the variables to estimate the model. Depending on the assumptions made about the way the spatial correlation affects the dependent variable, the literature distinguishes between several possibilities, being the so-called spatial autoregressive (SAR) structures perhaps the most commonly used. Formally, for a set of N cross-sectional data, a SAR model is expressed as

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

Where **y** is the $(N \times 1)$ vector with the values o the dependent variable, **W** is the $(N \times N)$ matrix of spatial weights, **X** is a $(N \times K)$ matrix of exogenous variables, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of parameters and $\boldsymbol{\varepsilon}$ is a $(N \times 1)$ stochastic error. In addition, $\rho$ is a spatial interaction parameter that measures how the endogenous variable $y$ is spatially influenced in average. The previous specification is a simple way to model the spatial interactions among regions, but it is possible to claim some weakness for estimate it. Firstly, the model (1) has a single parameter $\rho$. Hence, it is necessary to see the spatial interaction as an effect "in average" among regions. Furthermore, the estimated parameter $\rho$ depends on the rule followed by the researcher to define the matrix **W**, as the literature clearly shows. The election of this matrix is always in some sense a question of subjectivity introduced in the estimation. As a result, the estimation of the effect of the spatial-lag variables is a mix between data and chosen values for **W**. In other words, the previous specification is in fact a rather rudimentary way to express a much more complex spatial structure, as it follows in this system of equations

---

[2] Good examples of this other approach can be found in Case et. al(1993), Vayá et al. (1998, 1998b) and López-Bazo et al (1999). These papers define the spatial weights based on commercial relationships, while in Boarnet (1998) the weights increase with the similarity between the investigated regions. Molho (1995) and Fingleton (2001) propose a hybrid spatial weight based on economic variable and decreasing interaction force with distance.

$$y_1 = \sum_{k=1}^{k} x_{1k}\beta_k + \sum_{j=2}^{N} \rho_{1j} y_j + \varepsilon_1 \quad (2a)$$

$$y_2 = \sum_{k=1}^{k} x_{2k}\beta_k + \sum_{j\neq 2}^{N} \rho_{2j} y_j + \varepsilon_2 \quad (2b)$$

…

$$y_N = \sum_{k=1}^{k} x_{Nk}\beta_k + \sum_{j=1}^{N-1} \rho_{Nj} y_j + \varepsilon_N \quad (2c)$$

Or, in matricial terms

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega y} + \boldsymbol{\varepsilon} \quad (2d)$$

Where $\mathbf{\Omega}$ is a $\mathrm{N} \times \mathrm{N}$ matrix with zeros in its main diagonal and elements $\rho_{\mathrm{ij}}$ elsewhere; *i.e.*, the model includes a spatial parameter for each pair of regions. If this is the "real" spatial structure, the number of parameters to be estimated increases enormously. Model (1) requires the estimation of K+1 parameters from *N* observations. In contrast, in the spatial structure represented in equations (2a)-(2c) the number of parameters to be estimated now is K+N(N-1), which obviously is implausible by means of classical econometrics (OLS or ML, for example) given the negative number of degrees of freedom. Technically, this problem is labeled as an "ill-posed" econometric problem. If the number of observations N increases, this does not solve the problem but makes it worse, since the number of spatial parameters $\rho_{ij}$ to estimate also grows.[3] When several observations of the variables are available along T periods of time, the cross-section model can be transformed into a panel data model, although usually the length of the time series is not large enough to achieve efficient estimates. Even if the number of time periods was sufficient, and the problem became not "ill-posed", most probably it would be "ill-conditioned" given the high degree of multicollinearity between the variables $y_{ij}$.

These problems are circumvented estimating spatial models like (1): just one spatial parameter $\rho$ is estimated and interpreted as the average spatial effect. This means that the set of equations shown in (2a-2c) is reduced to

[3]Remember that the number of spatial parameters is N(N-1).

$$y_1 = \sum_{k=1}^{k} x_{1k}\beta_k + \rho\sum_{i=2}^{N} w_{i1} y_i + \varepsilon_1 \quad (3a)$$

$$y_2 = \sum_{k=1}^{k} x_{2k}\beta_k + \rho\sum_{i\neq 2}^{N} w_{i2} y_i + \varepsilon_2 \quad (3b)$$

…

$$y_N = \sum_{k=1}^{k} x_{Nk}\beta_k + \rho\sum_{i=1}^{N-1} w_{iN} y_i + \varepsilon_N \quad (3c)$$

In such a situation, the spatial spillover from a region *j* to other location *i* (the element $\rho_{ij}$), could be obtained as the product $\rho w_{ij}$, but then the estimated spillover is a mix between data and (exogenous) values of W. The choice of the spatial weight matrix is a key step in the spatial econometric modelling and nowadays there is not a unique method to select an appropriate specification of this matrix. In fact, this problem is suggested for future research by Anselin *et al.* (2004), and Paelink *et al.* (2004) among others. Note that if the spatial weights $w_{ij}$ are based on a measure of simply geographical distance, then the spillover from location *i* to location *j* will be exactly the same as the spillover from *j* to *i*.[4] This could turn into a strong simplification of the spatial relationships in an economy. Furthermore, if the **W** matrix is constructed as a contiguity matrix, then the spatial structure imposed is even simpler: between every pair of contiguous locations the spatial spillover is always the same and equal to $\rho$. The use of spatial weights based on some type of economic variables (instead of or besides geographical distance) could avoid the imposition of these symmetric relationships, but some problems of endogenity can emerge. Cohen and Morrison (2004) and Case et al. (1993) analyzed this problem and modified the weights in order to guarantee the ortogonality between the weights and the explanatory variables.

Note that models like (1) rely very much on the choice of matrix **W.** This issue can be considered as an important question for the estimation of the spatial econometric models, although it has not received much attention in the literature. One exception is the work by Stetzer (1982), where a numerical experiment by a series of Monte Carlo simulations is carried out to test the effects on the forecasting accuracy of misspecifying the elements of

[4] The row standardization of the W matrix implies that bcomes asymmetric even though the original matrix may nave been symmetric. Very recently, Bhattacharjee and Jensen-Butler (2005) propose the estimation of the spatial weight matrix which is consistent with a given or estimated spatial autocovariance without the non-negativity constraint on the off-diagonal elements.

**W**. More recently, Florax and Rey (1995) and Griffith (1996) made a similar exercise examining the consequences of misspecifications. [5] In a few words, one can see that all these papers agree in that a wrong specification of **W** is an important problem. Another reflection about the importance of **W** can be found in Case *et al.* (1993), where they point out that "in principle, it would be desirable to estimate the elements of the **W** matrix along with the other parameters. In practice, such an approach is out of the question because of insufficient degrees of freedom".[6] Both characteristics (excessive simplicity and too much dependence on the choice of W) can be seen as drawbacks of the classical "spatial" autoregressive models. As summary, Anselin (2002) asserts: "the specification of the weight matrix is a matter of some arbitrariness and is often cited as a major weakness of the lattice approach".

In this paper we propose the use of Generalized Maximum Entropy (GME) econometrics to estimate spatial structures. This technique is very attractive in situations where one has to deal with estimation of "ill-posed" models or "ill-conditioned" models, as those shown in equations (2a-2c). An application of GME methodology for estimating spatial models has been already proposed by Marshall and Mittelhammer (2004), but in a different fashion. The structure of the paper is as follows: in section 2 we give an overview and some intuitions of the GME methodology. In section 3 we explain how GME can be used to estimate econometric models where some spatial interrelationships are present. Section 4 compares the performance of GME estimators with the competing estimators based on Maximum Likelihood (ML) technique and the GME technique proposed by Marshal and Mittelhammer (GME-MM hereafter). A series of Monte Carlo simulations are computed to evaluate both techniques under several spatial structures. Finally, section 5 concludes.

## 2. GENERALIZED MAXIMUM ENTROPY ECONOMETRICS: AN OVERVIEW

In this section, we will give an introduction to GME econometrics, a collection of tools that can be very convenient to use scarce additional information in producing estimates for the unknown parameters of an econometric model. The aim of this section is just to give a brief introduction and some intuitions to the rationale of GME to the non-expert reader,

---

[5] Other works where the effects of misspecification are treated are Anselin (1985) or Anselin and Rey (1991). Other more recent works that study the impact of different specifications of the weight matrices are Bavaud (1998), where he introduces the possibility of using non-zero weights for the elements in the main diagonal; or Getis and Aldstadt (2004), where they search a W matrix that measures all the spatial dependence

[6] Case, A. C., H. S. Rosen and J. R. Hines Jr. (1993): "Budget spillovers and fiscal policy interdependence", Journal of Public Economics, 52, page 292.

rather than making an exhaustive review. The popularity of GME technique has increased remarkably since the comprehensive work by Golan, Judge and Miller (1996); the reader interested in a deeper analysis of this topic is strongly encouraged to read it[7].

To start with, let us assume that a random event can have $K$ possible outcomes $E_1, E_2, \ldots, E_K$ with the respective distribution of probabilities $\mathbf{p} = p_1, p_2, \ldots, p_K$ such that $\sum_{k=1}^{K} p_k = 1$. Following the formulation of Shannon (1948), the entropy of this distribution **p** will be

$$H(\mathbf{p}) = -\sum_{k=1}^{K} p_k \ln p_k \tag{4}$$

which reaches its maximum when **p** is a uniform distribution ($p_k = \frac{1}{K}, \forall k$). The entropy measure $H$ indicates the 'uncertainty' of the outcomes of the event. If some information (*i.e.*, observations) is available, it can be used to estimate an unknown distribution of probabilities for a random variable $x$ which can get values $\{x_1, \ldots, x_K\}$.

Suppose that there are $N$ observations $\{y_1, y_2, \ldots, y_N\}$ available such that

$$\sum_{k=1}^{K} p_k f_i(x_k) = y_i, \; 1 \leq i \leq N \tag{5}$$

with $\{f_1(x), f_2(x), \ldots, f_N(x)\}$ is a set of known functions representing the relationships between the random variable $x$ and the observed data $\{y_1, y_2, \ldots, y_N\}$. In such a case, the ME principle can be applied to recover the unknown probabilities. This principle is based on the selection of the probability distribution that maximizes equation (4) among all the possible probability distributions that fulfil (5). In other words, the ME principle chooses the "most uniform" distribution that agrees with the information. The following constrained maximization problem is posed:

$$\underset{\mathbf{p}}{Max}\, H(\mathbf{p}) = -\sum_{k=1}^{K} p_k \ln p_k \tag{6}$$

---

7 Kapur & Kesavan (1992) is another good reference for an extensive analysis of entropy-based econometric tools.

subject to:

$$\sum_{k=1}^{K} p_k f_i(x_k) = y_i;\ i = 1,...,N$$

$$\sum_{k=1}^{K} p_i = 1$$

In this problem, the last restriction is just a normalization constraint that guarantees that the estimated probabilities sum to one, while the first *N* restrictions guarantee that the recovered distribution of probabilities is compatible with the data for all *N* observations. It is important to note that even for *N*=1 (a situation with only one observation), the ME approach yields an estimate of the probabilities. Hence, in situations in which the number of observations is not large enough to apply econometrics based on limit theorems, this approach can be used to obtain robust estimates of unknown parameters.

For our current purposes, it is important that the above-sketched procedure can be generalized and extended to the estimation of unknown parameters for traditional linear models. Let us suppose that the problem at hand is the estimation of a linear model where a variable *y* depends on *K* explanatory variables $x_i$:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{e} \qquad (7)$$

where **y** is a $(\mathrm{N} \times 1)$ vector of observations for *y*, **X** is a $(\mathrm{N} \times \mathrm{K})$ matrix of observations for the $x_k$ variables, $\boldsymbol{\beta}$ is the $(\mathrm{K} \times 1)$ vector of unknown parameters $\boldsymbol{\beta}' = (\beta_1,...,\beta_\mathrm{K})$ to be estimated, and **e** is a $(\mathrm{N} \times 1)$ vector reflecting the random term of the linear model. For each $\beta_\mathrm{k}$, it will be assumed that there is some information about its $M \geq 2$ possible realizations by means of a 'support' vector $\mathbf{b}' = (b_1,...,b^*,...,b_M)$, the elements of which are symmetrically distanced around a central value $\beta_\mathrm{k} = \mathrm{b}^*$ (the prior expected value of the parameter), with corresponding probabilities $\mathbf{p}'_\mathrm{k} = (\mathrm{p}_{\mathrm{k}1},...,\mathrm{p}_{\mathrm{kM}})$. The construction of the vector **b** is based on the researcher's prior knowledge (or beliefs) about the parameter. Golan *et al.* (1996, chapter 8) devote more attention to consequences of choices concerning the elements of the vector **b**. For the sake of convenient exposition, it will be assumed that the *M* values are the same for every parameter, although this assumption can easily be relaxed. Now, vector $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{bmatrix} = \mathbf{Bp} = \begin{bmatrix} \mathbf{b'} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{b'} & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{b'} \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \dots \\ \mathbf{p}_K \end{bmatrix} \tag{8}$$

where $\mathbf{B}$ and $\mathbf{p}$ have dimensions $(K \times KM)$ and $(KM \times 1)$, respectively. The value for each parameter is then given by

$$\beta_k = \mathbf{b'p}_k = \sum_{m=1}^{M} b_m p_{km} \; ; \; k = 1,\dots,K \tag{9}$$

For the random terms, a similar approach is chosen. To express the lack of information about the actual values contained in $\mathbf{e}$, we assume a distribution for each $e_i$, with a set of $R \geq 2$ values $\mathbf{v'} = (v_1,\dots,v_R)$ with respective probabilities $\mathbf{q}'_i = (q_{i1}, q_{i2},\dots,q_{iR})$.[8] Hence, we can write

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_N \end{bmatrix} = \mathbf{Vq} = \begin{bmatrix} \mathbf{v'} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{v'} & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{v'} \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \dots \\ \mathbf{q}_N \end{bmatrix} \tag{10}$$

and the value of the random term for an observation *i* equals

$$e_i = \mathbf{v'q}_i = \sum_{r=1}^{R} v_r q_{ir} \; ; \; i = 1,\dots,N \tag{11}$$

And, consequently, equation (7) can be transformed into

$$\mathbf{y} = \mathbf{XBp} + \mathbf{Vq} \tag{12}$$

Now, the estimation problem for the unknown vector of parameters $\boldsymbol{\beta}$ is reduced to the estimation of $N + K$ probability distributions, and the following maximization problem (similar to problem (6)) can be solved to obtain these estimates

$$\underset{\mathbf{p},\mathbf{q}}{Max}\, H(\mathbf{p},\mathbf{q}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km} \ln(p_{km}) - \sum_{i=1}^{N}\sum_{r=1}^{R} q_{ir} \ln(q_{ir}) \tag{13}$$

subject to:

[8] Usually, the distribution for the errors is assumed symmetric and centered about 0, therefore $v_1 = -v_R$.

$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{ki} b_m p_{km} + \sum_{r=1}^{R} v_r q_{ir} = y_i ;\ i = 1,\dots,N$$

$$\sum_{m=1}^{M} p_{km} = 1;\ k = 1,\dots,K$$

$$\sum_{r=1}^{R} q_{ir} = 1;\ i = 1,\dots,N$$

By solving this GME program, we recover the estimated probabilities that allow us to obtain estimates for the unknown parameters.[9] The estimated value of $\beta_{\mathrm{k}}$ will be

$$\hat{\beta}_k = \sum_{m=1}^{M} \hat{p}_{km} b_m \ ; k = 1,\dots,K \tag{14}$$

Note that the solution of the constrained maximization problem (13) without additional information yields estimates equal to the expected value $b^*$ of the prior distribution, since in such a situation the recovered distribution would be uniform.

## 3. THE GME APPROACH FOR ESTIMATING SPATIAL STRUCTURES

### *3.1. THE GENERAL MODEL*

In this section, we suggest the use of GME to estimate spatial models with the general structure described in equation (2d). As commented previously, this is not the first proposal of using GME in this context: Marshall and Mittelhammer (2004) already proposed the use of GME data constrained estimator (GME-D) and GME normalized moment constrained estimator (GME-NM) in the context of spatial models, but only for estimating spatial structures expressed as equation (1). Our aim is to extent the use of GME estimators for more complex spatial structures.[10]

The starting point is the linear model of equation (7) where a spatial autoregressive term is added and, consequently, transformed into equation (2d)

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega y} + \boldsymbol{\varepsilon} \tag{2d}$$

---

[9] Golan *et al.* (1996, Chapter 6) show that these estimators are consistent and asymptotically normal. In Golan *et al.* (1996, Chapter 7) the finite sample behaviour of the GME estimators is numerically compared to traditional least squares and maximum likelihood estimators. In experimental samples with limited data, the ME estimators are found to be superior.

[10] For the shake of simplicity in this paper we focus only on the GME-D estimator. More details about their properties for linear models can be found in Golan *et al.* (1996, chapter 6) or Mittelhammer and Cardell (1998).

The GME procedure for the $\beta_k$ parameters and the $e_i$ error terms is the same as explained in section 3. Following this same reasoning, for each $\rho_{ij}$, it will be assumed that there are $L \geq 2$ possible realizations (assumed the same for all $\rho_{ij}$) that appear in a support vector $\mathbf{z}' = (z_1,...,z_L)$, with corresponding probabilities $\mathbf{s}'_{ij} = (s_{ij1},...,s_{ijL})$. Therefore, the matrix $\mathbf{\Omega}$ with elements $\rho_{ij}$ will be expressed as

$$\mathbf{\Omega} = \begin{bmatrix} 0 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 0 & \cdots & \rho_{2N} \\ \bullet & \bullet & \bullet & \bullet \\ \rho_{N1} & \rho_{N2} & \cdots & 0 \end{bmatrix} = \mathbf{z}' \otimes \mathbf{S} = \mathbf{z}' \otimes \begin{bmatrix} 0 & \mathbf{s_{12}} & \bullet & \mathbf{s_{1N}} \\ \mathbf{s_{21}} & 0 & \bullet & \mathbf{s_{2N}} \\ \bullet & \bullet & \bullet & \bullet \\ \mathbf{s_{N1}} & \mathbf{s_{N2}} & \bullet & 0 \end{bmatrix} \tag{15}$$

Where $\otimes$ denotes the Kronecker product. Consequently, equation (2d) can be rewritten as

$$\mathbf{y} = \mathbf{XBp} + \mathbf{z}' \otimes \mathbf{Sy} + \mathbf{Vq} \tag{16}$$

Now, the GME program for the unknown set of parameters $\boldsymbol{\beta}$ and $\mathbf{\Omega}$ is turned into the estimation of K+N(N-1)+N probability distributions, in the following terms:

$$\underset{\mathbf{p,s,q}}{Max}\, H(\mathbf{p},\mathbf{q},\mathbf{s}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km} \ln(p_{km}) - \sum_{i\neq j}^{N}\sum_{j\neq i}^{N}\sum_{l=1}^{L} s_{ijl} \ln(s_{ijl}) - \sum_{i=1}^{N}\sum_{r=1}^{R} q_{ir} \ln(q_{ir}) \tag{17}$$

subject to:

$$\sum_{k=1}^{K}\sum_{m=1}^{M} x_{ki} b_m p_{km} + \sum_{j\neq i}^{N}\sum_{l=1}^{L} y_j z_l s_{ijl} + \sum_{r=1}^{R} v_r q_{ir} = y_i;\ i = 1,...,N$$

$$\sum_{m=1}^{M} p_{km} = 1;\ k = 1,...,K$$

$$\sum_{r=1}^{R} q_{ir} = 1;\ i = 1,...,N$$

$$\sum_{l=1}^{L} s_{ijl} = 1;\ i = 1,...,N;\ j = 1,...,N;\ \forall \mathrm{i} \neq \mathrm{j}$$

By solving this GME program, we recover the estimated probabilities that allow us to obtain estimates for the unknown parameters. The estimated value of the spatial spillovers will be:[11]

[11] The expressions of estimators for $\beta$ parameters would be exactly as in the general linear model (12) described in the previous section.

$$\hat{\rho}_{ij} = \sum_{l=1}^{L} \hat{s}_{ijl} z_l \; ; \forall i \neq j \tag{18}$$

*3.2. THE USE OF ADDITIONAL "A PRIORI" INFORMATION*

The spatial model written in equation (2d) is the "most general" structure within a wide range of first-order spatial autoregressive processes. We speak about the "most general" because we are not imposing any prior belief that constraints the presence of spatial spillovers among the locations, which implies that many spatial parameters have to be estimated. Note that we allow the presence of spatial spillovers between any pair of locations, depending their magnitude or sign on the **z** values: the only prior information we are using refers to the values of the supporting vectors of the parameters.

A more restricted spatial structure can be estimated by means of GME, however, including some extra a priori information in the model; basically this can be done by making some extra assumptions. A natural way to exemplify this is referring to the type of spatial models estimated by Marshall and Mittelhammer (2004). Basically, as commented previously, they estimate autoregressive models like (1) using GME to obtain estimates of the parameter $\rho$. Since they use a contiguity matrix for the spatial weights $w_{ij}$, they assume that the spatial spillovers between any two locations with a common border are symmetric and with identical value. This a priori information included in the GME procedure reduces the number of spatial parameters to estimate, just 1 in such a case, and obviously the complexity of the computations is also decreased. But other not so straightforward spatial models can be estimated by using different prior information. The possibilities of incorporating prior beliefs are almost infinite and vary very much depending on the specific problem analyzed; in the following sub-sections we will consider two different sources of this information a priori: assumptions about the properties of the spatial spillovers and the use of a spatial weight matrix.

*3.2.1 Assumptions about the properties of the $\rho_{ij}$'s*

One way for reducing the complexity of models like (2d) would be that the researcher assumed that the spatial spillovers from a region *j* are exactly the same, not depending on the region they are going to. In other words, imposing that $\rho_{ij} = \rho_{hj} = \rho_j$; $\forall i,h \neq j$. This would transform the $\mathbf{\Omega}$ matrix in a new matrix $\mathbf{\Pi}$ such as

$$\mathbf{\Pi} \equiv \begin{bmatrix} 0 & \rho_2 & \dots & \rho_N \\ \rho_1 & 0 & \dots & \rho_N \\ \cdot & \cdot & \cdot & \cdot \\ \rho_1 & \rho_2 & \dots & 0 \end{bmatrix} \tag{19}$$

Obviously, in contrast with the general equation (2d) the number of spatial parameters to estimate reduces to N. The structure of the spatial autoregressive model looks

$$y_1 = \sum_{k=1}^{k} x_{1k}\beta_k + \sum_{j=2}^{N} \rho_j y_j + \varepsilon_1 \tag{20a}$$

$$y_2 = \sum_{k=1}^{k} x_{2k}\beta_k + \sum_{j\neq 2}^{N} \rho_j y_j + \varepsilon_2 \tag{20b}$$

…

$$y_N = \sum_{k=1}^{k} x_{Nk}\beta_k + \sum_{j=1}^{N-1} \rho_j y_j + \varepsilon_N \tag{20c}$$

Or in a more compact form as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Pi y} + \mathbf{\varepsilon} \tag{20d}$$

A similar prior, but in a different direction, can be incorporated if the researcher believes that a region *i* receives exactly the same spillover from any other location, *i. e*., supposing that $\rho_{ij} = \rho_{il} = \rho_i$; $\forall j,l \neq i$. In such a situation the matrix $\mathbf{\Omega}$ would become $\mathbf{\Theta}$, being this new matrix

$$\mathbf{\Theta} \equiv \begin{bmatrix} 0 & \rho_1 & \dots & \rho_1 \\ \rho_2 & 0 & \dots & \rho_2 \\ \cdot & \cdot & \cdot & \cdot \\ \rho_N & \rho_N & \dots & 0 \end{bmatrix} \tag{21}$$

In such a case we would have a set of equations as

$$y_1 = \sum_{k=1}^{k} x_{1k}\beta_k + \rho_1 \sum_{j=2}^{N} y_j + \varepsilon_1 \tag{22a}$$

$$y_2 = \sum_{k=1}^{k} x_{2k}\beta_k + \rho_2 \sum_{j\neq 2}^{N} y_j + \varepsilon_2 \tag{22b}$$

…

$$y_N = \sum_{k=1}^{k} x_{Nk}\beta_k + \rho_N \sum_{j=1}^{N-1} y_j + \varepsilon_N \quad (22c)$$

Or

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Theta y} + \boldsymbol{\varepsilon} \quad (22d)$$

Again, the number of spatial parameter to estimate is N. The form of the GME programs to estimate both types of models (20d) and (22d) would be very similar to (17), but with some minor changes in the objective function and the constraints. Evidently, the type of spatial model depicted in (1) is a stronger assumption than structures as (20d) or (22d) because it supposes that $\rho_{ij} = \rho;\ \forall i \neq j$.

*3.2.2 Using the W matrix as prior information*

In the previous subsection it has been explained how the GME methodology to estimate spatial models can be implemented without necessarily using a matrix of spatial weights **W**. However, if the researcher firmly believes that the $w_{ij}$ elements chosen truly reflect the spatial structure examined, this belief can be incorporated to the GME estimation procedure as prior information that may reduce the complexity of the model. A straightforward way to do this is modifying the form of equations (2a-2c) and transforming them into

$$y_1 = \sum_{k=1}^{k} x_{1k}\beta_k + \sum_{j=2}^{N} w_{1j}\rho_{1j} y_j + \varepsilon_1 \quad (23a)$$

$$y_2 = \sum_{k=1}^{k} x_{2k}\beta_k + \sum_{j\neq 2}^{N} w_{2j}\rho_{2j} y_j + \varepsilon_2 \quad (23b)$$

…

$$y_N = \sum_{k=1}^{k} x_{Nk}\beta_k + \sum_{j=1}^{N-1} w_{Nj}\rho_{Nj} y_j + \varepsilon_N \quad (23c)$$

Or, in matricial terms, as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{w}\mathbf{y} + \boldsymbol{\varepsilon} \quad (23d)$$

Where

$$\mathbf{\Omega}^{w} \equiv \begin{bmatrix} 0 & w_{12}\rho_{12} & \dots & w_{1N}\rho_{1N} \\ w_{21}\rho_{21} & 0 & \dots & w_{2N}\rho_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1}\rho_{N1} & w_{N2}\rho_{N2} & \dots & 0 \end{bmatrix} \tag{24}$$

Consider the case when the **W** matrix is binary (a contiguity matrix, for example), so the $w_{ij}$ elements can only take values 1 or 0. In such a situation is quite clear that the number of spatial parameters to be estimated will almost certainly decreases: the number of spatial parameters to estimate (non-zero cells of matrix $\mathbf{\Omega}^{w}$) would be equal to the number of cells of **W** with value 1, let say $S$, and evidently $S \leq N(N-1)$.

Of course, both types of information considered in these two subsections can be combined (or enhanced with other possible sources of prior beliefs). Although the use of this prior information can be helpful to alleviate the computational problems given by estimating a large number of parameters, note that the same problems commented in section 1 concerning the use of a misspecified weight matrix **W** or an excessively simple (non-realistic) spatial structure hold now.

## 4. MONTE CARLO SIMULATIONS

In this section, a numerical experiment will be carried out to compare the performance of GME methodology with other rival estimators in several scenarios, changing the features of the spatial first-order autoregressive process, as well as the a priori information incorporated to the GME programs.

### *4.1. DESIGN OF THE EXPERIMENT*

The model to be simulated for a grid of $N = 15$ artificially generated locations will be

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j \neq i}^{N} \rho_{ij} y_j + \varepsilon_i;\ i = 1,\dots,N \tag{25}$$

Or equation (2d) in matricial terms, where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} \tag{26a}$$

$$\varepsilon_i \approx \mathbf{N}[0,1];\ i = 1,\dots,N \tag{26b}$$

and $x_i \approx \mathbf{U}[0,10]; \; i = 1,...,N$, which are kept constant along the simulations (26c)

For simulating several spatial structures, the elements $\rho_{ij}$ of matrix $\mathbf{\Omega}$ have been generated in different scenarios

$$\rho_{ij} \approx \mathbf{U}[0,1]; \; \forall i \neq j \tag{27a}$$

$$\rho_{ij} \approx \mathbf{U}[-0.5,0.5]; \; \forall i \neq j \tag{27b}$$

In the first case (27a) the spatial spillovers are generated uniformly and constrained to take only positive values not greater than 1. In (27b) they can take negative or positive values either, with the limit of 0.5 in absolute value. In both cases they are generated from a uniform distribution and they both keep constant along the simulations. For the situation (27a) we denote the $\mathbf{\Omega}$ matrix as $\mathbf{\Omega^{F1}}$ and for (27b) as $\mathbf{\Omega^{F2}}$. The superscript F is used to call the attention to the point that all the off- diagonal elements of the matrix are not zero, so the matrix is completely "filled". In contrast to these situations, we additionally simulate two alternative scenarios where just some cells of the matrix (out of its trace) are allowed to be non-zero; specifically

$$\begin{cases} \rho_{ij} \approx \mathbf{U}[0,1] & \text{if } i \text{ and } j \text{ have a common border} \\ \rho_{ij} = 0 & \text{otherwise} \end{cases} \tag{27c}$$

$$\begin{cases} \rho_{ij} \approx \mathbf{U}[-0.5,0.5] & \text{if } i \text{ and } j \text{ have a common border} \\ \rho_{ij} = 0 & \text{otherwise} \end{cases} \tag{27d}$$

In order to decide when two locations *i* and *j* can be considered as neighbors, a rook criterion has been applied to our grid of 15 simulated locations.[12] The remaining characteristics are the same as in the two previous scenarios. The spillovers matrices simulated for cases (27c) and (27d) are labeled as $\mathbf{\Omega^{R1}}$ and $\mathbf{\Omega^{R2}}$ respectively. Clearly, the spatial processes generated by matrices $\mathbf{\Omega^{F1}}$ and $\mathbf{\Omega^{F2}}$ are more complex than those produced by $\mathbf{\Omega^{R1}}$ and $\mathbf{\Omega^{R2}}$, in the sense that the number of spatial relationships among the locations is greater in the former cases. Summing up, four different spatial autoregressive processes will be simulated 100 times, namely

[12] If a contiguity matrix is specified, two cells of the regular grid are contiguous if they have a common border of non-zero length, but the common border may be defined in different ways. The rook criterions consider as *common border* the common edge. Following a queen criterion, the common borer would a common vertex.

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{R1}}\mathbf{y} + \boldsymbol{\varepsilon} \tag{28a}$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{R2}}\mathbf{y} + \boldsymbol{\varepsilon} \tag{28b}$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{F1}}\mathbf{y} + \boldsymbol{\varepsilon} \tag{28c}$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{F2}}\mathbf{y} + \boldsymbol{\varepsilon} \tag{28d}$$

### *4.2. COMPARING THE ESTIMATORS*

Next, we will compare the performance of various spatial GME estimators proposed along this paper with other more classical proposals that will be taken as a benchmark. Specifically, our yardstick will be the Maximum Likelihood (ML) estimator and the GME estimator (GME-MM) proposed in Marshall and Mittelhammer (2004). Note that both estimation procedures suggest estimating models like that depicted in equation (1). Therefore, in order to implement them, it is necessary to construct a matrix of spatial weights **W** for the grid of 15 locations. Among all the wide range of possibilities we have considered two very simple and popular binary configurations for this matrix, being both of them based on a contiguity criterion: one is defined following a rook criterion and another following a queen criterion, labeled respectively as $\mathbf{W^R}$ and $\mathbf{W^Q}$. So we have models like

$$\mathbf{y} = \rho\mathbf{W}^{\mathbf{R}}\mathbf{y} + \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{29a}$$

$$\mathbf{y} = \rho\mathbf{W}^{\mathbf{Q}}\mathbf{y} + \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{29b}$$

As an alternative, we have considered the GME estimators for the models shown in equations (2d), (20d) and (22d). Following the reasoning of the GME procedure, it will be necessary to specify some support for the set of parameters to estimate and for the errors. Obviously, this is also required for obtaining the GME-MM estimators. For all them, we have chosen the following supporting vectors: $\mathbf{b} = [0,1,2]$ will be the discrete common support for $\beta_0$ and $\beta_1$, $\mathbf{s} = [-1,0,1]$ will be the discrete common support for every $\rho_{ij}$, and finally the support **v** for the error will be generated as a three-point vector centered about 0 following the 3-sigma rule of variable $y$ in each trial of the simulation, which is the most common practice.[13]

[13] A deeper discussion about the choice of these supports will be realized in the following subsection, where a sensitivity analysis is made.

For the GME estimators proposed in this paper it is not strictly necessary to employ a **W** matrix, although it can be incorporated into the GME programs in the form of prior information. This information can be integrated in those models in the form of a belief provided by the researcher. Consequently, besides equations (2d), (20d) and (22d), we have taken into account the following models[14]

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Pi}^{\mathbf{R}}\mathbf{y} + \mathbf{\varepsilon} \quad (30a)$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Pi}^{\mathbf{Q}}\mathbf{y} + \mathbf{\varepsilon} \quad (30b)$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Theta}^{\mathbf{R}}\mathbf{y} + \mathbf{\varepsilon} \quad (31a)$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Theta}^{\mathbf{Q}}\mathbf{y} + \mathbf{\varepsilon} \quad (31b)$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{R}}\mathbf{y} + \mathbf{\varepsilon} \quad (32a)$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\Omega}^{\mathbf{Q}}\mathbf{y} + \mathbf{\varepsilon} \quad (32b)$$

Which are basically extensions of the model (23d).

All this battery of models will be used to estimate the spatial structures simulated by equations from (28a) to (28d) and their estimates will be compared with the ML and GME-MM estimators under the two described configurations of matrix **W**. To realize the comparison we have computed along the 100 simulations the mean of several measures of error: the bias when estimating $\beta_0$ and $\beta_1$, the squared error (MSE) when estimating $\beta_0$, $\beta_1$ and the spatial parameters $\rho_{ij}$,[15] and the squared forecasting error (MSFE). The following tables summarize the results of this comparison:

[14] Again, the superscript R and Q are used to indicate the criterion followed (rook or queen) to define the matrix **W** used as prior information in the GME programs.

[15] In the case of MSE for $\rho_{ij}$ spillovers, we show the average computed for every $i \neq j$.

**Table 1. Comparison of the estimators in scenario (28a), true matrix is $\Omega^{R1}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | -1.956 | -3.456 | 0.908 | 0.407 | 12.210 | 0.181 | 4.304 | 1169.431 |
| **ML with $W^Q$** | -5.656 | -7.156 | 0.893 | 0.393 | 54.936 | 0.216 | 14.502 | 732.333 |
| **GME-MM with $W^R$** | 0.894 | -0.606 | 0.606 | 0.106 | 0.368 | 0.015 | 6.081 | 612.916 |
| **GME-MM with $W^Q$** | 0.871 | -0.629 | 0.561 | 0.061 | 0.396 | 0.008 | 14.417 | 730.154 |
| **GME $\Pi^R$ (30a)** | 0.911 | -0.589 | 0.537 | 0.037 | 0.347 | 0.003 | 10.872 | 949.624 |
| **GME $\Pi^Q$ (30b)** | 0.903 | -0.597 | 0.523 | 0.023 | 0.560 | 0.002 | 15.025 | 1101.191 |
| **GME $\Theta^R$ (31a)** | 0.954 | -0.546 | 0.767 | 0.267 | 0.298 | 0.074 | 16.545 | 639.978 |
| **GME $\Theta^Q$ (31b)** | 0.935 | -0.565 | 0.624 | 0.124 | 0.319 | 0.020 | 17.810 | 1068.749 |
| **GME $\Omega^R$ (32a)** | 0.907 | -0.593 | 0.396 | -0.104 | 0.352 | 0.013 | 12.383 | 682.534 |
| **GME $\Omega^Q$ (32b)** | 0.915 | -0.585 | 0.460 | -0.040 | 0.342 | 0.003 | 13.998 | 900.810 |
| **GME $\Pi$ (20d)** | 0.918 | -0.582 | 0.516 | 0.016 | 0.339 | 0.006 | 16.597 | 672.380 |
| **GME $\Theta$ (22d)** | 0.985 | -0.515 | 0.793 | 0.293 | 0.265 | 0.008 | 21.302 | 846.583 |
| **GME $\Omega$ (2d)** | 0.950 | -0.550 | 0.778 | 0.278 | 0.303 | 0.008 | 13.596 | 272.706 |

**Table 2. Comparison of the estimators in scenario (28b), true matrix is $\Omega^{R2}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | 2.104 | 0.604 | 0.483 | -0.017 | 1.296 | 0.026 | 3.272 | 60.129 |
| **ML with $W^Q$** | -0.428 | -1.928 | 0.511 | 0.011 | 18.709 | 0.086 | 5.772 | 98.179 |
| **GME-MM with $W^R$** | 1.112 | -0.388 | 0.478 | -0.022 | 0.152 | 0.006 | 3.041 | 39.186 |
| **GME-MM with $W^Q$** | 1.061 | -0.439 | 0.447 | -0.053 | 0.195 | 0.007 | 3.154 | 39.661 |
| **GME $\Pi^R$ (30a)** | 0.951 | -0.549 | 0.560 | 0.060 | 0.303 | 0.006 | 4.093 | 6.066 |
| **GME $\Pi^Q$ (30b)** | 0.952 | -0.548 | 0.564 | 0.064 | 0.300 | 0.006 | 4.096 | 3.536 |
| **GME $\Theta^R$ (31a)** | 0.953 | -0.547 | 0.616 | 0.116 | 0.300 | 0.018 | 6.544 | 22.264 |
| **GME $\Theta^Q$ (31b)** | 0.959 | -0.541 | 0.572 | 0.072 | 0.315 | 0.009 | 4.578 | 35.661 |
| **GME $\Omega^R$ (32a)** | 0.970 | -0.530 | 0.474 | -0.026 | 0.282 | 0.002 | 3.419 | 9.089 |
| **GME $\Omega^Q$ (32b)** | 0.922 | -0.578 | 0.610 | 0.110 | 0.334 | 0.014 | 3.441 | 5.228 |
| **GME $\Pi$ (20d)** | 0.985 | -0.515 | 0.906 | 0.406 | 0.265 | 0.165 | 3.673 | 2.022 |
| **GME $\Theta$ (22d)** | 1.015 | -0.485 | 0.619 | 0.119 | 0.236 | 0.025 | 6.839 | 25.692 |
| **GME $\Omega$ (2d)** | 0.959 | -0.541 | 0.677 | 0.177 | 0.293 | 0.074 | 3.296 | 1.908 |

**Table 3. Comparison of the estimators in scenario (28c), true matrix is $\Omega^{F1}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | -4.311 | -5.811 | 0.827 | 0.327 | 34.959 | 0.169 | 67.558 | 41.187 |
| **ML with $W^Q$** | -4.191 | -5.691 | 0.802 | 0.302 | 32.233 | 0.154 | 66.631 | 39.178 |
| **GME-MM with $W^R$** | 0.343 | -1.157 | 0.003 | -0.497 | 1.354 | 0.250 | 57.498 | 87.437 |
| **GME-MM with $W^Q$** | 0.456 | -1.044 | 0.011 | -0.489 | 1.098 | 0.240 | 57.082 | 43.757 |
| **GME $\Pi^R$ (30a)** | 0.498 | -1.002 | 0.026 | -0.474 | 1.106 | 0.225 | 66.447 | 41.935 |
| **GME $\Pi^Q$ (30b)** | 0.416 | -1.084 | 0.004 | -0.496 | 1.183 | 0.246 | 63.568 | 37.405 |
| **GME $\Theta^R$ (31a)** | 0.389 | -1.111 | 0.016 | -0.484 | 1.240 | 0.235 | 66.504 | 30.420 |
| **GME $\Theta^Q$ (31b)** | 0.492 | -1.008 | 0.098 | -0.402 | 1.021 | 0.214 | 69.328 | 29.266 |
| **GME $\Omega^R$ (32a)** | 0.443 | -1.057 | 0.024 | -0.476 | 1.122 | 0.227 | 64.837 | 35.655 |
| **GME $\Omega^Q$ (32b)** | 0.505 | -0.995 | 0.023 | -0.477 | 0.992 | 0.228 | 64.426 | 22.588 |
| **GME $\Pi$ (20d)** | 0.667 | -0.833 | 0.054 | -0.446 | 0.695 | 0.200 | 45.204 | 13.369 |
| **GME $\Theta$ (22d)** | 0.811 | -0.689 | 0.121 | -0.379 | 0.476 | 0.146 | 77.781 | 24.323 |
| **GME $\Omega$ (2d)** | 0.760 | -0.740 | 0.103 | -0.397 | 0.549 | 0.158 | 65.490 | 21.246 |

**Table 4. Comparison of the estimators in scenario (28d), true matrix is $\Omega^{F2}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | 5.841 | 4.341 | -2.160 | -2.660 | 65.287 | 18.198 | 19.099 | 7718.992 |
| **ML with $W^Q$** | 8.730 | 7.230 | -2.705 | -3.205 | 78.772 | 18.488 | 18.353 | 10361.023 |
| **GME-MM with $W^R$** | 0.946 | -0.554 | 0.754 | 0.254 | 0.307 | 0.069 | 19.693 | 648.593 |
| **GME-MM with $W^Q$** | 0.975 | -0.525 | 0.897 | 0.397 | 0.276 | 0.163 | 19.120 | 338.425 |
| **GME $\Pi^R$ (30a)** | 0.913 | -0.587 | 0.312 | -0.188 | 0.345 | 0.039 | 18.893 | 371.365 |
| **GME $\Pi^Q$ (30b)** | 0.360 | -1.140 | 0.008 | -0.492 | 0.360 | 0.008 | 19.761 | 210.943 |
| **GME $\Theta^R$ (31a)** | 0.938 | -0.562 | 0.491 | -0.009 | 0.316 | 0.003 | 20.258 | 314.012 |
| **GME $\Theta^Q$ (31b)** | 0.975 | -0.525 | 0.677 | 0.177 | 0.276 | 0.033 | 25.215 | 283.881 |
| **GME $\Omega^R$ (32a)** | 0.896 | -0.604 | 0.484 | -0.016 | 0.365 | 0.004 | 18.697 | 202.635 |
| **GME $\Omega^Q$ (32b)** | 0.936 | -0.564 | 0.664 | 0.164 | 0.319 | 0.030 | 18.141 | 182.675 |
| **GME $\Pi$ (20d)** | 0.960 | -0.540 | 0.754 | 0.254 | 0.292 | 0.073 | 23.282 | 82.636 |
| **GME $\Theta$ (22d)** | 1.011 | -0.489 | 0.712 | 0.212 | 0.239 | 0.047 | 21.657 | 290.845 |
| **GME $\Omega$ (2d)** | 0.958 | -0.542 | 0.743 | 0.243 | 0.295 | 0.060 | 18.320 | 136.584 |

Table 1 shows the average results for all these estimation alternatives for a scenario where the spillovers are bounded between 0 and 1 and they have been generated for every pair of locations with a common border following a rook criterion; *i.e.*, the situation shown by equation (28a) where the matrix of spatial spillovers employed to simulate the results is $\mathbf{\Omega}^{R1}$. Analogously, Tables 2, 3 and 4 do the same for the remaining 3 different scenarios generated by, respectively, spatial matrices $\mathbf{\Omega}^{\mathbf{R2}}$, $\mathbf{\Omega}^{\mathbf{F1}}$ and $\mathbf{\Omega}^{\mathbf{F2}}$. In every table the first group of results (first four rows) refers to the performance of ML and GME-MM estimators under the two configurations of **W** considered. The following six rows are connected with the GME estimators of the spatial models in equations from (30a) to (32b). Note that these models also impose the spatial structure specified in **W**. Finally, the set of the last three rows refers to the GME estimators for models where no a priori information about **W** has been considered.

The first two tables refer to scenarios where the $\mathbf{\Omega}^{\mathbf{R}}$ matrices were generated following a rook criterion. Consequently, a rational feeling would be that the models that include the belief that the **W** matrix is like $\mathbf{W}^{\mathbf{R}}$ are going to yield lower measures of error than those that impose a spatial structure derived from a $\mathbf{W}^{\mathbf{Q}}$ matrix or those that do not use at all any configuration of the spillovers as a priori information. If we examine the results of the simulation, it can be observed how the imposition of the right spatial configuration has special transcendence only in the case when we use a ML estimator. The first two rows of Tables 1 and 2 show how, for ML estimators, if we make wrong choice in the design of **W** matrix, then the consequences over the accuracy of our estimates and/or the forecasting capabilities of the model can be serious.[16]

The importance of this choice decreases if we use some of the GME based models. This can be seen as an advantage of using these techniques instead of more classical ML estimators since it seems that the gravity of a misspecification in **W** is reduced. Even if we do not include any a priori specified spatial structure, as in models (20d), (22d) or (2d) the measures of error present much smaller variability than for ML estimators. Note that this pattern holds for all the GME based estimators, including the GME-MM. Actually, in such scenarios where the spatial configuration can be more or less well described by the prior information included in the GME programs, there are not clear gains derived of using the type of GME estimators proposed in the paper (taking the GME-MM estimators as benchmark). Only models like (2d), which imply a considerable increase in the computational complexity, improve the forecasting accuracy of the GME-MM model, but they do not yield unquestionably better estimates for the $\beta$ or $\rho_{ij}$ parameters.

[16] This numerical result agrees with the conclusions of some previously mentioned papers, like Stetzer (1982) or Florax and Rey (1995).

The question now is: what happen if the actual spatial structure is more complex than the configuration of the **W** matrix we are specifying for our model? Tables 3 and 4 can give some clues about the answer. For the last two scenarios (28c) and (28d), one would expect that the GME estimators that do not include the structure contained in the **W** matrices somehow outperformed the ML and GME-MM estimators (since these impose a spatial structure derived from a rook or queen **W** matrix). The reason for this thought is given by the fact that these two scenarios are characterized by matrices of spatial spillovers $\mathbf{\Omega}^{\mathbf{F1}}$ and $\mathbf{\Omega}^{\mathbf{F2}}$, which implies spatial structures with a higher number of correlations among locations than are not taken into account when we use the rook or queen criterion. In other words, the type of models that uses a rook or queen **W** matrix includes "wrong" prior information, which forces the model to estimate a much more simple spatial structure than the actual one.

The results of our Monte Carlo simulations do not disagree with this idea: in general terms the results of the MSE for the parameters and the MSFE measure present the lowest values for models (2d), (20d) and (22d). Note that the gains are notable if we refer to the ML estimator. Even if we consider the GME-MM estimator as a yardstick the gains are more modest but still remarkable. The most important ones refer to the estimation of the $\beta$ parameters and to the forecasting capabilities of the models (in all the cases the squared errors are lower) rather than to the estimation of spatial spillovers $\rho_{ij}$. An in-between possibility between models (2d), (20d) and (22d) and ML and GME-MM estimators is the use of models like those expressed in equations (30a) to (32b): they contain the spatial structure imposed by the **W** matrices considered (like ML and GME-MM models), but they avoid the assumption that just one single average spatial parameter $\rho$ describes well the spatial configuration analyzed (unlike ML and GME-MM procedures). The figures of Tables 3 and 4 show clear improvements in the estimate of the $\beta$ parameters and in the forecasting errors with respect to the ML estimators, but some doubts with their performance when estimating the spatial spillovers $\rho_{ij}$. Compared to the GME-MM estimator, this same pattern holds although the gains derived from decreases in the squared errors are more moderate.

All in all, the results of the simulation suggest that it may be better not imposing any spatial structure in the estimation than considering an excessively simple one. The use of models like those in equations (2d), (20d) or (22d) do not require the imposition of a prior belief about the exact configuration of the spatial structure analyzed, but they estimate all the possible spatial relationships with no more assumptions than the functional form considered and the values included in the support vectors. The procedure proposed could be used successfully when there is not a clear certainty about what is the right specification for matrix **W**.

### *4.3. TESTING THE SENSITIVITY OF THE RESULTS*

A potential drawback of the GME estimators is an excessively high dependence of the estimates on the support vectors specified. This is an important issue since when we compared the performance of GME with ML in the previous subsection we were not being completely "fair", since we gave supports **b** and **z** that were quite well specified given how we simulate the different scenarios. For example, the GME estimates of spatial spillovers $\beta$ parameters should necessarily lay between 0 and 2, which limits the potential error that we can yield compared with ML technique (which does not restrict their values a priori). In order to check if the relatively good performance of the proposed GME estimators is just a consequence of this correct prior belief included in the supports, a sensitivity analysis is required.

To do that, we have taken the maximum and minimum estimates of $\beta_0$, $\beta_1$ and $\rho$ obtained along the 100 simulations by the ML procedure. In the cases where the spillovers were generated between 0 and 1 these bounds were:

| | | | | | |
|---|---|---|---|---|---|
| $\hat{\beta}_0$ max. | 0.439 | $\hat{\beta}_1$ max. | 1.605 | $\hat{\rho}$ max. | 0.511 |
| $\hat{\beta}_0$ min. | -11.326 | $\hat{\beta}_1$ min. | 0.297 | $\hat{\rho}$ min. | -0.178 |

And when the spillovers were generated between -0.5 and 0.5:

| | | | | | |
|---|---|---|---|---|---|
| $\hat{\beta}_0$ max. | 25.535 | $\hat{\beta}_1$ max. | 6.467 | $\hat{\rho}$ max. | 0.452 |
| $\hat{\beta}_0$ min. | -10.664 | $\hat{\beta}_1$ min. | -13.215 | $\hat{\rho}$ min. | -0.260 |

If we take these extreme estimates as the bounds for new support vectors **b'** and **z'** note that we will augment the wideness of these vectors and we will increase, therefore, the uncertainty about the plausible values of the parameters. More important, we are providing the GME programs with "bad" information since the central points of the new support are far from being the true values of the parameters; in contrast with the original supports chosen (this is especially clear for the case of the $\beta$ parameters). Furthermore, note that the true $\beta$ parameters are out of the range of the maximum and minimum values specified in the first case.

Considering the same measures of error to evaluate all the rival estimating procedures we obtain the following results:[17]

**Table 5. Sensitivity analysis, scenario (28a), true matrix is $\Omega^{R1}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | -1.956 | -3.456 | 0.908 | 0.407 | 12.210 | 0.181 | 4.304 | 1169.431 |
| **ML with $W^Q$** | -5.656 | -7.156 | 0.893 | 0.393 | 54.936 | 0.216 | 14.502 | 732.333 |
| **GME-MM with $W^R$** | -4.934 | -6.434 | 0.887 | 0.387 | 41.400 | 0.150 | 5.496 | 833.770 |
| **GME-MM with $W^Q$** | -4.852 | -6.352 | 0.753 | 0.253 | 40.473 | 0.068 | 14.231 | 1128.866 |
| **GME $\Pi^R$ (30a)** | -5.543 | -7.043 | 0.750 | 0.250 | 49.619 | 0.063 | 8.408 | 1303.408 |
| **GME $\Pi^Q$ (30b)** | -5.422 | -6.922 | 0.896 | 0.396 | 47.932 | 0.157 | 14.302 | 1138.529 |
| **GME $\Theta^R$ (31a)** | -4.812 | -6.312 | 0.895 | 0.395 | 39.858 | 0.156 | 14.544 | 686.022 |
| **GME $\Theta^Q$ (31b)** | -5.374 | -6.874 | 0.875 | 0.375 | 47.269 | 0.142 | 14.255 | 1115.650 |
| **GME $\Omega^R$ (32a)** | -4.944 | -6.444 | 0.798 | 0.298 | 41.554 | 0.089 | 8.524 | 805.401 |
| **GME $\Omega^Q$ (32b)** | -5.274 | -6.774 | 0.830 | 0.330 | 45.910 | 0.110 | 14.246 | 1220.308 |
| **GME $\Pi$ (20d)** | -4.506 | -6.006 | 0.921 | 0.421 | 36.098 | 0.178 | 13.094 | 1287.435 |
| **GME $\Theta$ (22d)** | -4.753 | -6.253 | 0.866 | 0.366 | 39.122 | 0.134 | 13.863 | 1329.451 |
| **GME $\Omega$ (2d)** | -2.345 | -3.845 | 0.973 | 0.473 | 15.091 | 0.225 | 13.002 | 62.3409 |

[17] Obviously, the results obtained by ML estimators are identical to those obtained previously.

### Table 6. Sensitivity analysis, scenario (28b), true matrix is $\Omega^{R2}$

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | 2.104 | 0.604 | 0.483 | -0.017 | 1.296 | 0.026 | 3.272 | 60.129 |
| **ML with $W^Q$** | -0.428 | -1.928 | 0.511 | 0.011 | 18.709 | 0.086 | 5.772 | 98.179 |
| **GME-MM with $W^R$** | 1.701 | 0.201 | 0.290 | -0.210 | 0.779 | 0.050 | 3.381 | 38.441 |
| **GME-MM with $W^Q$** | 2.068 | 0.568 | 0.322 | -0.178 | 0.676 | 0.036 | 3.370 | 28.305 |
| **GME $\Pi^R$ (30a)** | 1.503 | 0.003 | 0.192 | -0.308 | 0.322 | 0.101 | 4.328 | 11.632 |
| **GME $\Pi^Q$ (30b)** | 0.65 | -0.850 | 0.294 | -0.206 | 0.938 | 0.049 | 3.691 | 9.311 |
| **GME $\Theta^R$ (31a)** | 0.671 | -0.829 | 0.412 | -0.088 | 0.949 | 0.018 | 5.435 | 36.588 |
| **GME $\Theta^Q$ (31b)** | -0.430 | -1.930 | 0.533 | 0.033 | 3.886 | 0.009 | 5.122 | 44.990 |
| **GME $\Omega^R$ (32a)** | 0.045 | -1.455 | 0.533 | 0.033 | 2.272 | 0.008 | 3.913 | 28.150 |
| **GME $\Omega^Q$ (32b)** | 0.302 | -1.198 | 0.307 | -0.193 | 1.608 | 0.043 | 3.531 | 29.100 |
| **GME $\Pi$ (20d)** | -0.482 | -1.982 | 0.005 | -0.495 | 4.052 | 0.252 | 5.420 | 2.522 |
| **GME $\Theta$ (22d)** | -1.109 | -2.609 | 0.415 | -0.085 | 6.630 | 0.019 | 5.125 | 38.876 |
| **GME $\Omega$ (2d)** | -2.699 | -4.199 | 0.302 | -0.198 | 18.084 | 0.055 | 5.481 | 10.914 |

### Table 7. Sensitivity analysis, scenario (28c), true matrix is $\Omega^{F1}$

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | -4.311 | -5.811 | 0.827 | 0.327 | 34.959 | 0.169 | 67.558 | 41.187 |
| **ML with $W^Q$** | -4.191 | -5.691 | 0.802 | 0.302 | 32.233 | 0.154 | 66.631 | 39.178 |
| **GME-MM with $W^R$** | -3.105 | -4.605 | 0.329 | -0.171 | 21.260 | 0.029 | 64.841 | 98.721 |
| **GME-MM with $W^Q$** | -1.980 | -3.480 | 0.305 | -0.195 | 12.158 | 0.038 | 68.710 | 50.024 |
| **GME $\Pi^R$ (30a)** | -3.208 | -4.708 | 0.487 | -0.013 | 22.296 | 0.003 | 59.228 | 47.413 |
| **GME $\Pi^Q$ (30b)** | -2.071 | -3.571 | 0.332 | -0.168 | 12.792 | 0.029 | 56.070 | 62.529 |
| **GME $\Theta^R$ (31a)** | -1.989 | -3.489 | 0.331 | -0.169 | 12.196 | 0.030 | 40.667 | 36.741 |
| **GME $\Theta^Q$ (31b)** | -2.440 | -3.940 | 0.376 | -0.124 | 15.567 | 0.016 | 42.339 | 105.203 |
| **GME $\Omega^R$ (32a)** | -2.341 | -3.841 | 0.392 | -0.108 | 14.826 | 0.014 | 58.846 | 42.114 |
| **GME $\Omega^Q$ (32b)** | -2.131 | -3.631 | 0.331 | -0.169 | 13.251 | 0.029 | 55.422 | 52.947 |
| **GME $\Pi$ (20d)** | -1.624 | -3.124 | 0.384 | -0.116 | 9.791 | 0.015 | 43.048 | 40.054 |
| **GME $\Theta$ (22d)** | -1.976 | -3.476 | 0.331 | -0.169 | 12.141 | 0.030 | 42.870 | 40.053 |
| **GME $\Omega$ (2d)** | -1.026 | -2.526 | 0.443 | -0.057 | 6.427 | 0.005 | 38.475 | 46.694 |

**Table 8. Sensitivity analysis, scenario (28d), true matrix is $\Omega^{F2}$**

| Average results | $\hat{\beta}_0$ | Bias $\hat{\beta}_0$ | $\hat{\beta}_1$ | Bias $\hat{\beta}_1$ | MSE $\beta_0$ | MSE $\beta_1$ | MSE $\rho_{ij}$ | MSFE |
|---|---|---|---|---|---|---|---|---|
| **ML with $W^R$** | 5.841 | 4.341 | -2.160 | -2.66 | 65.287 | 18.198 | 19.099 | 7718.992 |
| **ML with $W^Q$** | 8.730 | 7.223 | -2.705 | -3.2045 | 78.772 | 18.488 | 18.353 | 10361.023 |
| **GME-MM with $W^R$** | -0.195 | -1.695 | 0.582 | 0.082 | 3.186 | 0.011 | 18.034 | 667.669 |
| **GME-MM with $W^Q$** | -0.518 | -2.018 | 0.832 | 0.332 | 5.541 | 0.132 | 18.088 | 333.742 |
| **GME $\Pi^R$ (30a)** | 4.930 | 3.430 | -0.699 | -1.199 | 13.185 | 1.467 | 18.619 | 419.983 |
| **GME $\Pi^Q$ (30b)** | 0.359 | -1.141 | 0.138 | -0.362 | 2.338 | 0.540 | 19.455 | 333.416 |
| **GME $\Theta^R$ (31a)** | 1.576 | 0.076 | -0.009 | -0.509 | 0.304 | 0.269 | 20.352 | 369.882 |
| **GME $\Theta^Q$ (31b)** | 1.731 | 0.231 | -0.154 | -0.654 | 0.231 | 0.434 | 20.351 | 476.251 |
| **GME $\Omega^R$ (32a)** | 1.677 | 0.177 | 0.235 | -0.265 | 0.672 | 0.048 | 18.662 | 452.720 |
| **GME $\Omega^Q$ (32b)** | 0.046 | -1.454 | 0.218 | -0.282 | 2.802 | 0.093 | 19.322 | 390.753 |
| **GME $\Pi$ (20d)** | -0.215 | -1.715 | 0.185 | -0.315 | 3.944 | 0.135 | 19.883 | 284.965 |
| **GME $\Theta$ (22d)** | 1.806 | 0.306 | 0.052 | -0.448 | 0.360 | 0.209 | 18.928 | 403.105 |
| **GME $\Omega$ (2d)** | 0.081 | -1.419 | -0.013 | -0.513 | 2.200 | 0.276 | 19.925 | 203.191 |

Tables 5 to 8 show the behavior of the GME estimators do under these new support vectors. Obviously, the measure errors for the $\beta$ parameters increase and the forecasting errors are also larger almost in all the situations. The change in the MSE's for parameters $\rho_{ij}$ is not so important, since the new supports are not radically different from the true range specified of these parameters. Even so, the general proposal explained in the previous subsection still remains: from tables 7 and 8 we can observe how the GME models (2d), (20d) and (22d) that do not employ a **W** matrix still outperform competing estimators based on models that consider a wrong (too simple) configuration of the actual spatial structure.

When one wants to estimate a spatial econometric model it is necessary to assume some prior information. One possibility is using a classical a approach and specifying a matrix **W** of spatial weights: this could imply important consequences for the accuracy of the estimates if this belief is not correct. Other possibility is using some of the GME estimators assuming that the support vectors that we have to define for the parameters really bound their actual value. One might think that, in most cases, for the researcher is easier to define plausible values of the economic parameters rather than giving an accurate description of spatial structure by means of defining a matrix **W**. The basic idea that suggest the results of this sensitivity analysis is that the performance of the spatial models are more vulnerable to wrong priors of the first type than to bad specifications of the vectors used as support by the GME estimators.

## 5. Concluding remarks

Generalized Maximum Entropy (GME) econometrics is an attractive methodology in situations where one has to deal with estimation of "ill-posed" or "ill-conditioned" models. In this paper we propose the use of this technique to estimate complex spatial structures, which fit with these "ill-behaved" situations where the number of observations is not large enough to estimate the desired number of parameters. To compare the performance of the proposed technique to other more traditional estimation methodologies a series of Monte Carlo simulations are carried out under different scenarios. The outcomes of the simulations suggest that the proposed GME technique outperforms other competing estimators if the actual spatial structure is different from the assumptions specified in the **W** matrix, which is inevitably used by these other methodologies.

The two most important advantages of the proposed GME procedure are: 1) the possibility of obtaining "individual" estimates of $\rho_{ij}$ spatial parameters for each pair of locations (instead of a single "average" spatial parameter $\rho$) , and 2) it does not requires necessarily the assumption of an (exogenously specified) matrix of spatial weights **W**. On the other hand, it requires the specification of priors for the values of the parameters to be estimated. Consequently, the use of the GME procedure implies switching from assumptions about the underlying spatial structure to beliefs about the values of the parameters. However, our feeling is that for the researcher is generally easier to make more accurate assumptions about the plausible values of the parameters than about the structure of the spatial relationships among the locations studied. Nevertheless, this paper must be seen just as a first approximation to an approach that potentially can be very useful for the estimation of spatial models. However, much further research in this direction must be done with the GME technique proposed. Its performance has to be evaluated under more sophisticated definitions of **W,** different types of spatial correlation, sizes of sample, etc.

## References

Anselin, L. (1988): *Spatial Econometrics: Methods and models*. Boston: Kluwer Academic.

Anselin, L. (1985): "Estimation and inference in misspecified spatial autoregressive models: some small sample results", Columbus, Ohio State University Research Foundation.

Anselin, L. and Rey, S. (1991): "Properties of Test for Spatial Dependence in Linear Regression Models", *Geographical Analysis*, vol.23, nº 2, págs. 112-131.

Anselin, L.; Bera, A.K. (1998): "Spatial dependence in linear regression models", in Ullah, A. and Giles, D. Eds, *Handbook of Applied Economic Statistics*. Marcel Dekker, New York.

Anselin, L.; R.J.G.M Florax, and S.J. Rey, (2004): "Econometrics for Spatial Models: Recent Advances" in Anselin,L.;Florax, R.J.G.M and Rey, S.J (eds): *Advances in Spatial Econometrics*, p. 1-25. Berlin:Springer_Berlag.

Anselin, L (2002): "Under the hood: Issues in the specification and interpretation of spatial regression models", *Agricultural Economics*, vol.27, p.247-267.

Bhattacharjee, A and Jensen-Butler, C. (2005): "Estimation of Spatial Weights Matrix in a Spatial Error Model, with an Application to Diffusion in Housing Demand", *CRIEFF Discussion Papers* nº0519

Bavaud, F. (1998): "Models for spatial weights: A systematic look", *Geographical Analysis*, 30, p.153-171.

Boarnet, M.G. (1998): "Spillovers and the Locational Effects of Public Infrastructure", *Journal of Regional Science*, vol. 38, p. 381–400.

Case, A.. C. (1991): "Spatial Patterns in Household Demand", *Econometrica*, vol.59, p.953-965.

Case, A. C., Rosen, H. and Hines J.R. (1993): "Budget Spillovers and Fiscal policy interdependence: Evidence from the States", *Journal of Public Economics*, 52, p.285-307.

Cliff, A.D. and Ord, J.K. (1973): *Spatial autocorrelation*, Pion, London,UK.

Cliff, A.D.; Ord, J.K. (1981): *Spatial processes: models and applications*. Pion Limited.

Cohen, J. P and Morrison, C. J. (2004): "Public infrastructure investment, interstate spatial spillovers, and manufacturing cost", *The Review of Economic and Statistics*, May 2004, 86(2): 551-560.

Conley, T. G. (1999): "GMM estimation with cross sectional dependence", *Journal of Econometrics*, 92, p.1-45.

Dubin, R. A. (1998): "Spatial autocorrelation: A Primer", *Journal of Housing Economics,* nº 7, p. 304-327.

Fingleton, B. (2001): "Equilibrium and Economic Growth: Spatial Econometric Models and Simulations", *Journal of Regional Science*, vol. 41, nº1, p. 117–147.

Florax, R.J.G.M. and Van der Vlist, A.J. (2003): "Spatial econometric data analysis: moving beyond traditional models", *International Regional Science Review* 26,3: 223-243.

Fraser, I., (2000): "An application of maximum entropy estimation: the demand for meat in the United Kingdom", *Applied Economics*, Vol. 32(1), pp. 45-59.

Gardebroek, C. and A. Oude Lansink, (2004): "Farm-specific Adjustment Costs in Dutch Pig Farming", *Journal of Agricultural Economics*, Vol. 55(1), pp. 3-24.

Golan, A., Judge, G. and D. Miller, (1996): *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York, John Wiley & Sons.

Golan, A., Perloff, J.M. and E.Z. Shen, (2001): "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand", *The Review of Economics and Statistics*, Vol. 83(3), pp. 541-550.

Getis, A. and Aldstadt (2004): "Constructing the spatial weights matrix using a local statistic", *Geographical Analysis* vol.36, nº 2, p. 90-104.

Griffith, D.A. (1996): "Some Guidelines for Specifying the Geographic Weight Matrix Contained in Spatial Statistical Models", in *Practical Handbook of Spatial Statistics* edited by S.S. Arlinghaus. Boca Raton: CRC.

Kapur, J.N and H.K. Kesavan (1992): *Entropy Optimization Principles with Applications*, academic Press, New York.

Kelejian, H.H. and Prucha, I.R. (1998): "A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances", *Journal of Real Estate Finance and Economics*, 17, p. 99-121.

Kelejian, H.H. and Prucha, I.R. (1999): "A generalized moments estimator for the autoregressive parameter in a spatial model", *International Economic Review*, 40, p. 509-533.

López-Bazo, E.; Vayá, E.; Artís, M. (2004): "Regional externalities and growth: evidence from European regions", *Journal of Regional Science,* vol. 44, No.1, p. 43-73.

Marsh T. L. and Mittelhammer R. C. (2004): "Generalized Maximum Entropy Estimation of a First Order Spatial Autoregressive Model", *Advances in Econometrics*, 18, pp. 199-234

Mittelhammer, R. C. and N. S. Cardell (1998): "The Data-Constrained GME Estimator of the GLM: Asymptotic Theory and Inference" *Mimeo*, Washington State University.

Molho, I. (1995): "Spatial autocorrelation in British unemployment", *Journal of Regional Science*, vol. 36, nº 4, p. 641-658.

Ord, J. K. (1975): "Estimation methods for models of spatial interaction", *Journal of the American Statistical Association*, 95, p.120-126.

Paelink, J.H.P; Mur,J. and Trívez, J. (2004): “Spatial Econometrics: More Lights than Shadows”, *Estudios de Economía Aplicada*, Vol.22-3, p.1-19.

Paris, Q. and R.E. Howitt, (1998): "An Analysis of Ill-Posed Production Problems Using Maximum-Entropy", *American Journal of Agricultural Economics*, Vol. 80, pp. 124-138.

Stetzer, F. (1982): “Specifying weights in spatial forecasting models: the results of some experiments”, *Environment and Planning A,* 14, p.571-584.