

Fernandez-Vazquez, Esteban

Conference Paper

Empirical versus exogenous spatial weighting matrices: An entropy-based intermediate solution

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Fernandez-Vazquez, Esteban (2010) : Empirical versus exogenous spatial weighting matrices: An entropy-based intermediate solution, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118909>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Empirical versus exogenous spatial weighting matrices: an entropy-based intermediate solution

Esteban Fernández Vázquez

Department of Applied Economics-University of Oviedo

Phone: (+34)985105056; e-mail: evazquez@uniovi.es.

Abstract:

The classical approach to estimate spatial models lays on the choice of a spatial weights matrix that reflects the interactions among locations. The two main alternative proposals to specify the structure of these weights may be called exogenous and endogenous procedures. In this paper, we propose an intermediate solution by using entropy econometrics. The basic idea is that, once an exogenous weighting matrix is specified, we could modify our initial specification using information contained in the sample. We evaluate the comparative performance of this entropy-based estimation by means of Monte Carlo simulations and also show an illustration with an empirical example.

Keywords: spatial econometrics, maximum entropy econometrics, weighting matrix specification, Monte Carlo simulations.

JEL Classification: C15, C21

1. INTRODUCTION

The selection of a specific spatial weights matrix \mathbf{W} is a key issue when estimating spatial models, but at the same time there is not a unanimous criterion to choose the most appropriate spatial weights for a given empirical application.¹ Basically, there are two alternative approaches to the problem of the specification of spatial weights. One of the streams promote fixing the \mathbf{W} matrix exogenously to the model basing on some concept of geographical proximity. For example, a very simple way to characterize their elements w_{ij} is by defining them as binary variables that take value 1 when locations i and j are neighbor and 0 otherwise (depending on the existence or not of a common border, for example). The geographical distance between locations i and j can be used in a more direct way, defining w_{ij} as a distance decay function. Other authors prefer using some economic measure of distance based on interregional trade flows, income differences, etc.²

Some other authors, on the contrary, propose the construction of \mathbf{W} matrices based on some “empirical” evidence about the variables of the model. They are critical of the “exogenous approach”, because the spatial lag operator imposed can be very different from the real spatial structure underlying in the data. For example, Kooijman (1976) or Boots and Dufornaud (1994) define as one criterion the choice of \mathbf{W} that maximizes the Moran statistic. Following a similar idea, Mur and Paelinck (2010) base their specification of \mathbf{W} on the so-called complete correlation coefficients. Two papers by Getis and Aldstadt base their specification of \mathbf{W} on the values of the G_i^* local statistic (Getis and Aldstadt, 2004) and on the use of a multi-directional algorithm (Aldstadt and Getis, 2006). Bhattacharjee and Jensen-Butler (2006) suggest a method to estimate \mathbf{W} based on the real structure of the spatial autocovariance, while Conley (1999) proposes the direct estimation of the spatial autocovariances.

¹ See Anselin (2002), page 259.

² Some examples of these other approaches can be found in Molho (1995), Fingleton (2001) or López-Bazo, Vayá and Artís (2004).

The literature distinguishes several spatial models depending on the assumptions about the way in which spatial correlation affects the dependent variable. Specifically, Anselin (2003) presents a wide taxonomy of different types of spatial models. Although it can be easily extended to other situations, in this paper we focus on a situation where the externalities spread across space through a spatial cross-regressive (SCR) structure. This type of model is applied when some spatial spillovers are included in the model, but their influence is assumed to be limited to local effects (Anselin, 2003, page 161). In other words, these models are appropriate when some of the explanatory variables of the model produce spatial spillovers with a spatial range delimited by the location and its immediate neighbors, but not beyond.

Traditionally, for a set of N cross-sectional data, a SCR model is written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{x}^* + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{y} is the $(N \times 1)$ vector with the values of the dependent variable, \mathbf{W} is the $(N \times N)$ matrix of *a priori* spatial weights, \mathbf{X} is a $(N \times K)$ matrix of exogenous variables, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of parameters to estimate and $\boldsymbol{\varepsilon}$ is a $(N \times 1)$ stochastic error. In addition, ρ is a spatial interaction parameter that measures how the variable y is spatially influenced. The weighting matrix \mathbf{W} represents the spatial structure of the spillovers.

In this paper we propose using Generalized Maximum Entropy (GME) econometrics to estimate such models. The GME approach has been applied recently to spatial regression models by Marsh and Mittelhammer (2004) or Fernandez-Vazquez et al. (2009), who estimated a first order spatial auto-regression model using this technique. The present paper will use the GME technique to define spatial lag operators that can be seen to lie in an intermediate position between the “exogenous” and “empirical” approaches.

The basic idea is that we initially fix an exogenous a priori \mathbf{W} matrix but, once this is specified, we could modify our initial specification.

The structure of the paper is the following: Section 2 provides an overview of the GME methodology and shows how it can be applied to the context of SCR models. Section 3 evaluates the relative performance of the GME technique using a sampling experiment under different scenarios of divergence between the actual spatial network and the weighting matrix \mathbf{W} specified in the estimation. In section 4 we illustrate how the proposed GME approach can be applied to an empirical estimation problem. Finally, section 5 presents the concluding remarks.

2. GENERALIZED MAXIMUM ENTROPY ECONOMETRICS: AN OVERVIEW³

Let us assume that a discrete random event can take K possible outcomes E_1, E_2, \dots, E_K with the respective distribution of probabilities $\mathbf{p}' = p_1, p_2, \dots, p_K$ such that $\sum_{k=1}^K p_k = 1$. Following the formulation proposed by Shannon (1948), the entropy of \mathbf{p} is:

$$H(\mathbf{p}) = - \sum_{k=1}^K p_k \ln(p_k) \quad (2)$$

The entropy function H measures the ‘uncertainty’ of the outcomes of the event. This function reaches its maximum when \mathbf{p} follows a uniform distribution: $p_k = \frac{1}{K}; \forall k$. On the other hand, this function takes a value of zero (no uncertainty) when the probability of one of the outcomes goes to one. If some information about the variable (*i.e.*, observations) is available, it can be used to estimate an unknown distribution of probabilities for a random variable \mathbf{x} that takes values x_1, x_2, \dots, x_K . Suppose that there are N observations y_1, y_2, \dots, y_N available such that:

³ This section summarizes the process to estimate the parameters of a linear model. See Golan, Judge and Miller (1996) and Kapur and Kesavan (1992) for further details.

$$\sum_{k=1}^K p_k f_n(x_k) = y_n ; \forall n \quad (3)$$

Where $f_n(\mathbf{x})$ is a generic function that represents the relationships between the random variable \mathbf{x} and the observed data y_1, y_2, \dots, y_N . In this situation, 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 (2) among all of the possible probability distributions that fulfill (3). In other words, the ME principle chooses the “most uniform” distribution that corresponds with the information. The following constrained maximization problem is posed:

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

subject to:

$$\sum_{k=1}^K p_k f_n(x_k) = y_n ; n = 1, \dots, N \quad (4b)$$

$$\sum_{k=1}^K p_k = 1 \quad (4c)$$

In this problem, the last constraints is simply guarantees that the estimated probabilities add up to one, while the first N restrictions force the recovered distribution of probabilities to be compatible with the data for all N observations.

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 to face is the estimation of a linear model where a variable \mathbf{y} depends on K explanatory variables x_k :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5)$$

where \mathbf{y} , \mathbf{X} , $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are defined exactly the same way as for equation (1). Each parameter β_k is assumed to be a discrete random variable with $M \geq 2$ possible realizations that are included into a ‘support’ vector $\mathbf{b}' = (b_1, \dots, b_M)$ with corresponding probabilities $\mathbf{p}'_k = (p_{k1}, \dots, p_{kM})$. The vector \mathbf{b} is based on the researcher’s a priori belief about the likely values of the parameter. For the ease of exposition, it will be assumed that the M values are the same for every k parameter, although this assumption can easily be relaxed. Now, vector $\boldsymbol{\beta}$ can be written as:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} = \mathbf{B}\mathbf{P} = \begin{bmatrix} \mathbf{b}' & 0 & \dots & 0 \\ 0 & \mathbf{b}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{b}' \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{bmatrix} \quad (6)$$

Where \mathbf{B} and \mathbf{P} have dimensions $(K \times KM)$ and $(KM \times 1)$ respectively. Now, the value of each parameter β_k is given by the following expression:

$$\beta_k = \mathbf{b}' \mathbf{p}_k = \sum_{m=1}^M b_m p_{km}; \quad \forall k = 1, \dots, K \quad (7)$$

For the random term, a similar approach is followed. In contrast to other estimation techniques, GME does not require rigid assumptions about a specific probability distribution function of the stochastic component, but it still is necessary to make some assumptions. $\boldsymbol{\varepsilon}$ is assumed to have mean $E[\boldsymbol{\varepsilon}] = 0$ and a finite covariance matrix. Basically, we represent our uncertainty about the realizations of vector $\boldsymbol{\varepsilon}$ treating each element ε_n as a discrete random variable with $J \geq 2$ possible outcomes contained in a convex set $\mathbf{v}' = \{v_1, \dots, v_J\}$, which for the sake of simplicity is assumed as common for all the ε_n . We also assume that these possible realizations are symmetric around zero ($-v_1 = v_J$). The traditional way of fixing the upper and lower limits of this set is to apply the three-sigma rule (see Pukelsheim, 1994). Under these conditions, vector $\boldsymbol{\varepsilon}$ can be defined as:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} = \mathbf{V}\mathbf{U} = \begin{bmatrix} \mathbf{v}' & 0 & \dots & 0 \\ 0 & \mathbf{v}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{v}' \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix} \quad (8)$$

and the value of the random term for an observation n equals:

$$\varepsilon_n = \mathbf{v}' \mathbf{u}_n = \sum_{j=1}^J v_j u_{nj}; \quad \forall n = 1, \dots, N \quad (9)$$

Consequently, model (5) can be transformed into:

$$\mathbf{y} = \mathbf{X} \mathbf{B} \mathbf{p} + \mathbf{V} \mathbf{U} \quad (10)$$

So we need also to estimate the elements of matrix \mathbf{U} (denoted by \tilde{u}_{nj}) and the estimation problem for the general linear model (5) is transformed into the estimation of $K + N$ probability distributions. Therefore, the GME problem is written in the following terms:

$$\text{Max}_{\mathbf{P}, \mathbf{U}} H(\mathbf{P}, \mathbf{U}) = - \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln(p_{km}) - \sum_{n=1}^N \sum_{j=1}^J u_{nj} \ln(u_{nj}) \quad (11a)$$

subject to:

$$y_n = \sum_{k=1}^K \sum_{m=1}^M b_m p_{km} x_{kn} + \sum_{j=1}^J v_j u_{nj}; \quad \forall n = 1, \dots, N \quad (11b)$$

$$\sum_{m=1}^M p_{km} = 1; \quad \forall k = 1, \dots, K \quad (11c)$$

$$\sum_{j=1}^J u_{nj} = 1; \quad \forall n = 1, \dots, N \quad (11d)$$

By solving this GME program, we recover the estimated probabilities that allow us to obtain estimates for the unknown parameters. Further details about the statistical properties of the GME estimator can be found in Appendix A. The estimated value of β_k will be:

$$\hat{\beta}_k = \mathbf{b}' \hat{\mathbf{p}}_k = \sum_{m=1}^M b_m \hat{p}_{km}; \quad \forall k = 1, \dots, K \quad (12)$$

This GME procedure can be extended for estimating spatial cross-regressive models such as (1). Following the same procedure explained above for the β_k parameters, it will be assumed that there are $L \geq 2$ possible realizations for the spatial parameter ρ in a support vector $\mathbf{z}' = (z_1, \dots, z_L)$, with corresponding probabilities $\mathbf{q}' = (q_1, \dots, q_L)$. The parameter ρ , consequently, can be also estimated by GME by means of this reparametrization.

This idea was applied by Marsh and Mittelhamer (2004) for the case of spatial autoregressive models once a matrix of spatial weights \mathbf{W} is specified. Fernandez-Vazquez et al. (2009) extended this idea and proposed estimating all the ρ_{ij} elements of a matrix of spatial parameters instead of using a predetermined \mathbf{W} matrix. In this paper we suggest a solution that lies in an intermediate position between these two previous approaches. In our proposal, only one single spatial parameter ρ is defined, but the elements of a spatial weights matrix \mathbf{W} will be also estimated.

Note that the GME can be naturally applied in this context, given that the elements of matrix \mathbf{W} are typically row-standardized and are non-negative. Consequently, each row of \mathbf{W} can be taken as a probability distribution with unknown elements w_{ni} to be recovered:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \cdot & w_{1N} \\ w_{21} & 0 & \cdot & w_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1} & w_{N2} & \cdot & 0 \end{bmatrix} \quad (13)$$

This means that equation (1) can be rewritten as:

$$\mathbf{y} = \mathbf{XBp} + \mathbf{s}'\mathbf{zWx}^* + \mathbf{VU} \quad (14)$$

Now the empirical GME program estimates $K+2N+1$ probability distributions, in the following terms:

$$\begin{aligned} \text{Max}_{\mathbf{P}, \mathbf{s}, \mathbf{U}} H(\mathbf{P}, \mathbf{s}, \mathbf{W}, \mathbf{U}) = & - \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln(p_{km}) - \sum_{n=1}^N \sum_{j=1}^J u_{nj} \ln(u_{nj}) \\ & - \sum_{l=1}^L s_l \ln(s_l) - \sum_{n=1}^N \sum_{i \neq n}^N w_{ni} \ln(w_{ni}) \end{aligned} \quad (15a)$$

subject to:

$$y_n = \sum_{k=1}^K \sum_{m=1}^M b_m p_{km} x_{kn} + \left(\sum_{l=1}^L s_l z_l \right) \left(\sum_{i \neq n}^N w_{ni} x_i^* \right) + \sum_{j=1}^J v_j u_{nj}; \quad \forall n = 1, \dots, N \quad (15b)$$

$$\sum_{m=1}^M p_{km} = 1; \quad \forall k = 1, \dots, K \quad (15c)$$

$$\sum_{j=1}^J u_{nj} = 1; \quad \forall n = 1, \dots, N \quad (15d)$$

$$\sum_{i \neq n}^N w_{ni} = 1; \quad \forall n = 1, \dots, N \quad (15e)$$

$$\sum_{l=1}^L s_l = 1 \quad (15f)$$

The GME program above includes the entropy associated to the spatial parameter and to the weighting matrix in the objective function (15a). Equations (15c)-(15d) are again normalization constraints. Restriction (15b) forces the recovered probabilities to fit with the observations of the dependent variable. This GME program estimates, together with the parameters of the model, the elements of the matrix of spatial weights. These estimates (namely \hat{w}_{ni}) are those that maximize our uncertainty about the \mathbf{W} matrix and that, simultaneously, are compatible with the available information. In other words, we choose as elements of the matrix those \hat{w}_{ni} that, being consistent with the observed data, diverge least with our prior assumption on \mathbf{W} .

Additionally, the estimated value of the spatial spillovers will be:

$$\hat{\rho} = \sum_{l=1}^L s_l z_l \quad (16)$$

3. A NUMERICAL EXPERIMENT: ESTIMATING A SCR MODEL WITH A BINARY CONTIGUITY CRITERION

In this section, the performance of the GME technique will be compared with other competing techniques in a scenario where the spatial structure that generates the data is given by a binary matrix of proximity. This type of contiguity criterion is relatively common in the empirical applications of spatial models and is very simple to define. Under this specification, the elements of the \mathbf{W} matrix are defined as the following function:

$$\begin{cases} w_{ni} = 1 & \text{if locations } n \text{ and } i \text{ are neighbor regions} \\ w_{ni} = 0 & \text{otherwise} \end{cases}$$

We have simulated the cross-regressive model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}^r\mathbf{x}^* + \boldsymbol{\varepsilon}$ with 1,000 replications for a lattice of $N = 15$ artificially generated locations. We also suppose that we have observations for $T=20$ time periods. In our experiment, the error term is generated in each simulation as a $N(0,1)$ distribution. Matrix \mathbf{X} is composed by a constant term and two independent variables x_1 and x_2 . Moreover, we assume that there is a variable \mathbf{x}^* that produces spillovers between neighbor regions. More specifically, we suppose that only the independent variable x_2 generates these spatial externalities on other regions. The values for the independent variables and for the parameters (kept constant throughout the simulations) are:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 1.5 \\ 0.5 \end{bmatrix}; \rho = 0.2 \quad (17a)$$

$$x_{1nt} \sim U(-10,10); n = 1, \dots, N; t = 1, \dots, T \quad (17b)$$

$$x_{2nt} \sim U(-5,5); n = 1, \dots, N; t = 1, \dots, T \quad (17c)$$

$$x_{nt}^* = x_{2nt}; n = 1, \dots, N; t = 1, \dots, T \quad (17d)$$

In the experiment, the connectivity between the 15 locations is given by the spatial pattern contained in the matrix \mathbf{W}^r , which is not necessarily equal to the weighting matrix used to estimate the model (\mathbf{W}). We have introduced the possibility of divergence between the real matrix (\mathbf{W}^r) and the one specified in the model (\mathbf{W}) assuming that $w_{ni}^r - w_{ni} = u_{ni}$; where $u_{ni} \sim N(0, \sigma)$. σ is a scalar that reflects the degree of divergence between the real and the used spatial weighting matrices and if $\sigma = 0$ this would indicate that the real and the specified matrix are exactly the same.

From this scenario for the sampling experiment, we compare the GME approach with other rival procedures. In order to apply the GME procedure to estimate models like (14), it is necessary to specify some support for the set of parameters and for the errors. For β_0 , β_1 and β_2 the same support $\mathbf{b}' = (-2, 0, 2)$ has been set. Note that although the actual values of the parameters are quite different among them, the support chosen contains the same range of possible realizations. In addition, the support is not centered on the true value of any of the parameters, which means that we are including not very good prior information for the estimation of the β parameters. The support vector for the spatial parameter ρ was set as $\mathbf{z}' = (-1, 0, 1)$. Finally, the support \mathbf{v} for the error has been generated as a three-point vector centered about 0 following the common procedure of the 3-sigma rule of variable y in each trial of the experiment (Pulkeshim, 1994; Golan, Judge and Miller, 1996).

The benchmark for the comparison will be the estimation by maximum likelihood (ML). One basic difference is that in ML we specify a matrix \mathbf{W} and we apply it directly in the estimation. In contrast, using GME we take \mathbf{W} as an *a priori* approximation to \mathbf{W}^r , but then we let the data speak for themselves and we could use spatial weights \hat{w}_{ni} (estimates of the elements on \mathbf{W}^r) different from our initial assumptions.

Table 1 summarizes the results of this first experiment. For each one of the competing estimators we have computed the mean of the estimates of β_0 , β_1 and β_2 throughout the 1,000 simulations (columns 1 to 3) and the empirical variance, the overall mean squared error of the estimates of β_0 , β_1 and β_2 (MSE β , column 4) and the so-called “mean squared spatial error” (MSSE, column 5). This measure of deviation in the estimates quantifies the aggregate squared differences between the actual ρ and its respective estimate.

<<Table 1 about here>>

Each row of Table 1 contains a different value for the scalar σ . As expected, the deviations between the actual and the estimated parameters for both methods increase for large values of σ . Moreover, when the matrix \mathbf{W} is correctly specified ($\sigma = 0$) both GME and ML yield relative low error measures, although GME slightly outperforms ML.⁴ However, the performance of the two competing estimation techniques is remarkably different as σ grows. When the differences between the real \mathbf{W}^r and the \mathbf{W} used in the estimation become larger, the GME begins to yield comparatively better estimates than ML. Figures 1 and 2 illustrates this idea both for the β and ρ parameters.

<<Figure 1 about here>>

<<Figure 2 about here>>

This is because, in the GME, the specification of \mathbf{W} can be seen as an a priori assumption that can be modified by the information contained in the sample. In other words, the data in the sample help to alleviate a wrong assumption about \mathbf{W}^r . All in all, the results suggest that with perfect certainty about the actual spatial network \mathbf{W}^r , using the GME technique proposed does not imply large gains compared with ML. On the other hand,

⁴ Actually, this is the case analyzed in Marsh and Mittelhammer (2004) for SAR models.

if we do not have clear evidences for imposing the right structure in the spatial network, using a GME estimator seems to limit the estimation errors.

4. An empirical illustration: regional spillovers from public capital in Spain.

This section applies the entropy-based estimator presented previously to a real world example. The main objective will be to illustrate how to estimate the spatial spillovers present in a model such (1) for the N=15 Spanish inland regions (we exclude the Canary and Balearic Island in our analysis). The dependent variable \dot{y}_{it} is labor productivity (gross value added per worker), which is assumed to depend basically on private capital per worker (\dot{k}_{it}), the internal stock of public capital (\dot{g}_{it}) and the stock of public capital in other regions ($\dot{g}_{nt}, n \neq i$), which are spatially weighted by matrix \mathbf{W} . The elements of this matrix are defined as $w_{ni} = 1$ when regions n and i share a common border and are null in any other case. In other words, we will assume that the stock of public capital in one region can produce changes in the labor productivity in adjacent regions. The specific equation to be estimated is:

$$\dot{y}_{it} = \alpha \dot{k}_{it} + \beta \dot{g}_{it} + \rho \sum_{n \neq i}^{15} w_{ij} \dot{g}_{nt} + \varepsilon_{it} \quad (18)$$

where $i = 1, \dots, 15$ (the 15 Spanish inland regions) and $t = 1980, \dots, 2000$. Regional data on gross value added, number of workers and stock of private and public capital have been obtained from the BDMores database elaborated by the Spanish Ministry of Economy at constant prices of 1980. All the variables are measured in logs and expressed in differences to their respective regional average along the period 1980-2000 (the dots indicate growth rates with respect to this average).

We have applied the entropy-based estimation procedure proposed in this paper, which requires specifying some supports for the set of parameters to

estimate and for the errors. For the parameters α and β we have considered the same support vector \mathbf{b} with 3 points $(-0.5, 0, 0.5)$ for all the parameters included in the model. To test the sensitivity of the results to the specification of this support, we repeated the estimation fixing the elements of vector \mathbf{b} in $(-1, 0, 1)$ and $(-10, 0, 10)$. The usual three-sigma rule applies for specifying the supporting vectors for the error terms in all the cases.

Besides the point estimates, the GME procedure facilitates diagnosis and inference. In particular, it is possible to obtain standard deviation of the point estimates, as described in the Appendix. A global measure of goodness of fit has been also obtained. The results are summarized in Table 2:

<<Table 2 about here>>

The first two columns of Table 2 show the estimates for the parameters and the third one reports a pseudo- R^2 obtained from the variance of the errors of the model as in Arnd et al. (2002). The results obtained under the three specifications for the supporting vectors would suggest a positive contribution of the stock of public capital to regional productivity, revealing also the importance of the endowment of this factor in the adjacent regions. The estimates of the elasticities of both own and external public capital are in line with those obtained in earlier works applied to Spain like, for example, Mas et al. (1996, p. 645), Cantos et al. (2005, p. 42) Moreno and Lopez-Bazo (2007, p. 58) or Marquez et al (2009, p. 29).

An interesting result comes from the estimates of the cells of \mathbf{W} matrix, which also provide information on the regional characteristics of the public capital spillovers. Note that, originally, the specified contiguity matrix assigned uniform weights to each region i that shares a common border with other region n . However, the GME procedure applied here allows for updating this initial assumption given the observed sample: some of the weights will be increased whereas others will be reduced. Table 3 shows, for

each weight in the matrix, the rate of variation from their a priori uniform value:⁵

<<Table 3 about here>>

Obviously, the row sums of these variations should equal zero in all the cases, since the estimates of w_{ni} must be row-standardized, but the column sums inform of the positive or negative update assigned to the spatial weights on each region. A positive column sum will indicate that the GME estimates assign more weight to this specific region in the generation of public capital spillovers than the a priori uniform specification. The opposite case happens when the column sum is negative. The regional distribution of these column sums are shown in Map 1:

<<Map 1 about here>>

The regions marked in green are those with a net positive update of the spatial weights, while those marked in brown are the regions with a negative update. Even when this is a highly aggregated regional classification that does not allow for a detailed analysis, some geographical patterns can be observed in this map. All the regions located in the coast (with the exception of Murcia) receive larger spatial weights than originally assigned in the contiguity matrix, whereas the weights of the interior regions situated in the North of the country are reduced. This seems to indicate that, generally speaking, the interior Northern regions receive more intense public capital spillovers from the coastal regions than the other way around.

⁵ The results shown in Table 3 and Map 1 correspond to the model with the supporting vectors specified at the intermediate situation when $\mathbf{b} = (-1, 0, 1)$, but the general picture does not change for the other two supports.

5. CONCLUDING REMARKS

The specification of the spatial weighting matrix has been an important issue in the field of spatial econometric analysis that has received considerable attention. The main problem is that there is not a unique approach to define the spatial weights and two alternative streams can be distinguished in the literature. One of the proposals supports using weighting matrices determined exogenously to the model, while other authors prefer to use some empirical evidence to specify them. This paper suggests a sort of intermediate way between these two proposals where the W matrix is a priori specified exogenously, but in a second stage the weights are updated by means of the GME estimator. Focusing in the so-called spatial cross-regressive models, a numerical experiment compares the performance of the proposed GME with a traditional ML estimator, and the results suggest that the possibility of updating the prior assumptions made in the W matrix facilitates more accurate estimates. Not surprisingly, the comparative performance of GME gets better when the divergence between the actual and the a priori elements of W grows.

In spite of the encouraging results found, much work in this line of research has still to be done. Note that the present version of the paper just focuses on the case of binary W matrices, but the basic idea can be easily extended to other possible (more complex) configurations. Basing on the relationship between the Shannon's entropy measure and the Kullback divergence with respect to any given a priori distribution, the GME problem can be transformed in terms of a Generalized Cross Entropy (GCE) program in order to consider W matrices different from the binary case.

ACKNOWLEDGEMENTS

This research was partially financed by the *Ministerio de Innovación y Ciencia* with the project ECO2008-01617/ECO.

REFERENCES

- Ardnt C., Robinson S. and F. Tarp, 2002, Parameter estimation for a computable general equilibrium model: a maximum entropy approach, *Economic Modelling*, 19, 375-398.
- Aldstadt, J. and A. Getis, 2006, Using AMOEBA to create a spatial weights matrix and identify spatial clusters, *Geographical Analysis*, 38 (3), 327-343.
- Anselin, L., 2002: Under the hood: Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 27, 247-267.
- Anselin, L., 2003: Spatial externalities, spatial multipliers and spatial econometrics, *International Regional Science Review*, 26 (2), 153-166.
- Bhattacharjee, A. and C. Jensen-Butler (2006), Estimation of spatial weights matrix, with an application to diffusion in housing demand, *CRIEFF Discussion Papers 0519*.
- Boots, B. and C. Dufournaud, (1994), A programming approach to minimizing and maximizing spatial autocorrelation statistics, *Geographical Analysis*, 26, 54-66.
- Cantos, P., Gumbau-Albert M. and J. Maudos, 2005, Transport infrastructures, spillover effects and regional growth: evidence of the Spanish case, *Transport Reviews*, 25 1, 25-50.
- Case, A.C., Rosen, H. and J.R. Hines, 1993, Budget spillovers and fiscal policy interdependence: Evidence from the States, *Journal of Public Economics*, 52, 285-307.
- Conley, T.G. (1999), GMM estimation with cross sectional dependence, *Journal of Econometrics* 92, 1-45.
- Fernandez-Vazquez, E., Mayor-Fernandez M. and J. Rodriguez-Valez, 2009, Estimating spatial autoregressive models by GME-GCE techniques, *International Regional Science Review*, 32 (2), 148-172.
- Fingleton, B., 2001, Equilibrium and economic growth: Spatial econometric models and simulations, *Journal of Regional Science*, 41 (1), 117–147.

- Fraser, I., 2000, An application of maximum entropy estimation: the demand for meat in the United Kingdom, *Applied Economics*, 32(1), 45-59.
- Getis, A. and J. Aldstadt, 2004, Constructing the spatial weights matrix using a local statistic, *Geographical Analysis*, 36 (2), 90-104.
- Golan, A., Judge, G. and D. Miller, 1996, *Maximum Entropy Econometrics: Robust Estimation with Limited Data* (John Wiley & Sons-New York).
- 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*, 83(3), 541-550.
- Kapur, J.N and H.K. Kesavan, 1992, *Entropy optimization principles with applications*, New York, Academic Press.
- Kooijman, S.A.L.M., 1976, Some remarks on the statistical analysis of grids especially with respect to ecology, *Annals of Systems Research*, 5, 113-132.
- Lopez-Bazo, E.; Vaya, E. and M. Artis, 2004, Regional externalities and growth: evidence from European regions, *Journal of Regional Science*, 44 1, 43-73.
- Marquez, M. A., Ramajo, J., and G.J.D. Hewings, 2009, Domestic and cross-border effects of public capital: a SVAR approach for the Spanish regions, Discussion Paper Series of REAL, 09-T-2.
- Mas, M., Maudos, J., Perez F. and E. Uriel, 1996, Infrastructures and productivity in the Spanish regions, *Regional Studies*, 30 7, 641-649.
- Marsh T. L. and R.C. Mittelhammer, 2004, Generalized maximum entropy estimation of a first order spatial autoregressive model, *Advances in Econometrics*, 18, 199-234.
- Mittelhammer R.C. and N.S. Cardell, 1997, On the consistency and asymptotic normality of data-constrained GME estimator in the general linear model, Mimeo, University of Washington.
- Molho, I., 1995, Spatial autocorrelation in British unemployment, *Journal of Regional Science*, 36 4, 641-658.

Moreno, R., and E. Lopez-Bazo, 2007, Returns to local and transport infrastructure under regional spillovers, *International Regional Science Review*, 30 (1), 47-71.

Mur, J. and J.H.P. Paelinck, 2010, Deriving the W-matrix via p-median complete correlation analysis of residuals, *Annals of Regional Science*, (forthcoming).

Pukelsheim F, (1994) The three sigma rule, *The American Statistician*, 48, 88-91.

APPENDIX A: STATISTICAL PROPERTIES OF THE GME ESTIMATOR

The large sample properties of the ME estimators are analyzed in Golan, Judge and Miller (1996; chapter 6). ME estimators are shown to be consistent and asymptotically normal. These authors analyze also the small sample properties using Monte Carlo simulation. They compare numerically the GME estimators to traditional least squares and maximum likelihood estimators. Their results show a good performance in terms of the accuracy of the estimates.

In order to do inference in the GME approach, the procedure suggested by Mittelhammer and Cardell (1997), Fraser (2000) or Golan, Perloff and Shen (2001) can be followed. Under some assumptions on the behavior of model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ that guarantee the consistency and asymptotical normality of the estimator, the distribution of the estimates follows $\widehat{\boldsymbol{\beta}} \rightarrow N\left[\boldsymbol{\beta}, \frac{\sigma_\lambda^2}{\kappa^2} (\mathbf{X}'\mathbf{X})^{-1}\right]$, where σ_λ^2 is the variance of the Lagrange multipliers of (13b) or (17b). It is possible to estimate with consistency σ_λ^2 as $\sigma_\lambda^2 = \frac{\sum_{n=1}^N \widehat{\lambda}_n^2}{N}$, where $\widehat{\lambda}_n$ is the Lagrange-multiplier associated to observation n . κ^2 is a scalar related to the variance of the error term. The parameter κ can be estimated consistently as $\widehat{\kappa} = \frac{1}{N \sum_{n=1}^N \text{Var}(\widehat{e}_n)}$; where $\text{Var}(\widehat{e}_n) = \left[\sum_{j=1}^J v_j^2 \widehat{u}_{ij}\right] - \left[\sum_{j=1}^J v_j \widehat{u}_{ij}\right]^2$. Hence, it is possible to estimate the variance of GME estimators and obtain the t-ratios as $\frac{\widehat{\boldsymbol{\beta}}}{\sqrt{\widehat{\text{Var}}(\widehat{\boldsymbol{\beta}})}}$.

Table 1. Results of the numerical experiment (1,000 replications)

		(1)	(2)	(3)	(4)	(5)
Estimator		Average $\hat{\beta}_0$ True $\beta_0 = 0.25$	Average $\hat{\beta}_1$ True $\beta_1 = 1.5$	$\hat{\beta}_2$ True $\beta_2 = 0.5$	MSE$_{\beta}$	MSSE
$\sigma = 0.00$	ML	0.246 [0.002]	1.497 [0.001]	0.490 [0.006]	0.008	0.006
	GME	0.175 [0.001]	1.470 [0.001]	0.523 [0.001]	0.008	0.003
$\sigma = 0.05$	ML	0.251 [0.002]	1.497 [0.001]	0.545 [0.006]	0.010	0.008
	GME	0.179 [0.001]	1.470 [0.001]	0.532 [0.001]	0.008	0.004
$\sigma = 0.10$	ML	0.254 [0.002]	1.498 [0.001]	0.592 [0.006]	0.016	0.015
	GME	0.182 [0.001]	1.471 [0.001]	0.540 [0.001]	0.009	0.005
$\sigma = 0.15$	ML	0.255 [0.002]	1.498 (0.001)	0.637 [0.006]	0.026	0.025
	GME	0.183 [0.001]	1.471 [0.001]	0.547 [0.001]	0.010	0.007
$\sigma = 0.20$	ML	0.255 [0.002]	1.498 [0.001]	0.682 [0.006]	0.041	0.040
	GME	0.184 [0.001]	1.471 [0.001]	0.555 [0.001]	0.010	0.008
$\sigma = 0.25$	ML	0.255 [0.002]	1.498 [0.001]	0.732 (0.006)	0.061	0.060
	GME	0.184 [0.001]	1.472 [0.001]	0.564 ([0.001])	0.010	0.009

The empirical variances of the estimates are shown in brackets.

$$\text{MSE}_{\beta} = \text{MSE } \beta_0 + \text{MSE } \beta_1 + \text{MSE } \beta_2.$$

Table 2. Estimates of the empirical model

	(1)	(2)	(3)	(4)
	α	β	ρ	Pseudo-R ²
$\mathbf{b} = (-0.5, 0, 0.5)$	0.316** [0.016]	0.084** [0.015]	0.182** [0.018]	0.861
$\mathbf{b} = (-1, 0, 1)$	0.413** [0.016]	0.066** [0.015]	0.161** [0.018]	0.866
$\mathbf{b} = (-10, 0, 10)$	0.459** [0.016]	0.066** [0.015]	0.141** [0.018]	0.866

The standard deviation of the estimates are shown in brackets. ** means that the estimate differs significantly from zero at a 5% level.

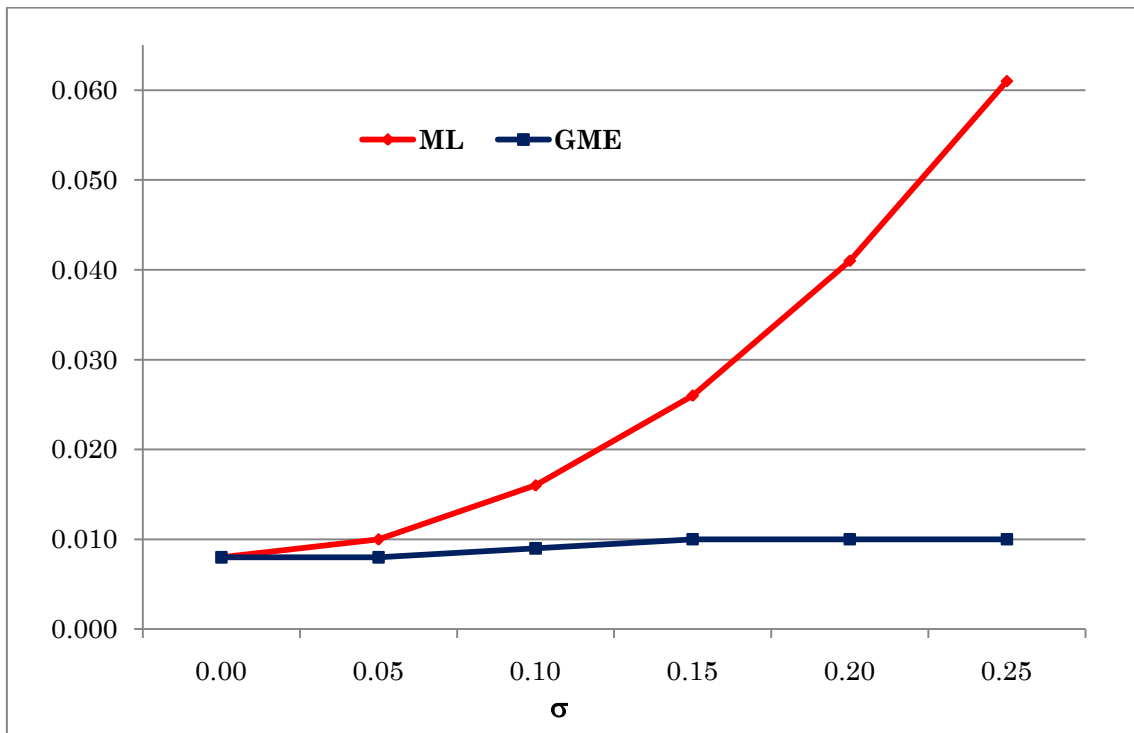
Figure 1. MSE for the β parameters

Figure 2. MSE for the ρ parameter

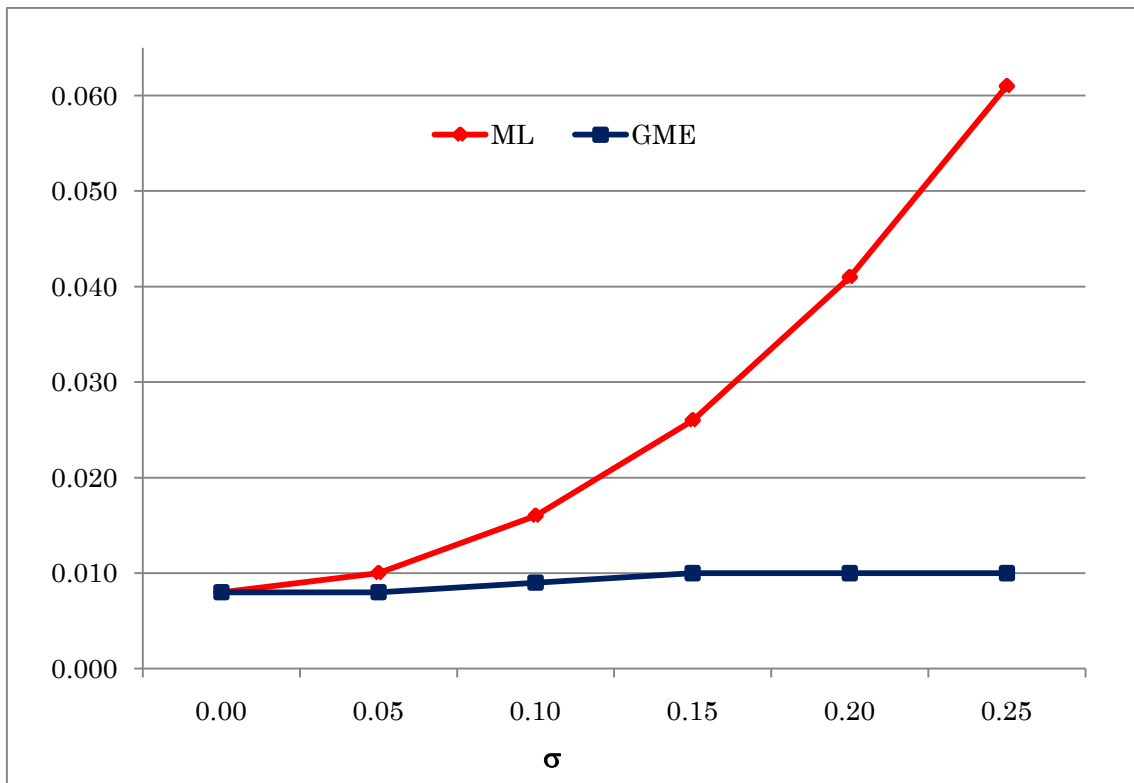


Table 3. Percentage of variation between the initial and the estimated weights

	And	Ara	Ast	Cant	CLeon	CLM	Cat	Val	Ext	Gal	Mad	Mur	Nav	BC	Rio
Andalucia						4.49			-2.50			-2.00			
Aragon					-1.93	8.92	1.91	7.00					-1.75		-14.16
Asturias				2.16	-5.59					3.44					
Cantabria			4.72		-10.66									5.94	
C. Leon		-5.66	1.37	4.28		2.16			5.55	3.34	-2.73		-2.40	3.28	-9.20
C. La Mancha	2.63	-2.97			-1.95			-0.09	2.19		-2.05	2.24			
Catalonia		2.04						-2.04							
Valencia						0.07	6.24					-6.31			
Extremadura	13.13				-14.58	1.46									
Galicia			3.13		-3.13										
Madrid					7.32	-7.32									
Murcia	-7.88					4.10		3.78							
Navarra		-0.80			-2.94									-4.32	8.07
Basque Country				10.53	-0.03								0.82		-11.32
Rioja		-10.36			-1.84								-0.39	12.58	

Map 1. Net re-allocation on spatial weights by region

