© *Investigaciones Regionales. 21 – Pages 53 to 69*
Section ON THEORY AND METHODS

Investigaciones Regionales

Updating weighting matrices by Cross-Entropy

Esteban Fernández Vázquez *

ABSTRACT: The classical approach to estimate spatial models lays on the choice of a spatial weights matrix that reflects the interactions among locations. The rule used to define this matrix is supposed to be the most similar to the «true» spatial relationships, but for the researcher is difficult to elucidate when the choice of this matrix is right and when is wrong. This key step in the process of estimating spatial models is a somewhat arbitrary choice, as Anselin (2002) pointed out, and it can be seen as one of their main methodological problems. This note proposes not imposing the elements of the spatial matrix but estimating them by cross entropy (CE) econometrics. Since the spatial weight matrices are often row-standardized, each one of their rows can be approached as probability distributions. Entropy Econometrics (EE) techniques are a useful tool for recovering unknown probability distributions and its application allows the estimation of the elements of the spatial weights matrix instead of the imposition by researcher. Hence, the spatial lag matrix is not a matter of choice for researcher but of empirical estimation by CE. We compare classical with CE estimators by means of Monte Carlo simulations in several scenarios on the true spatial effect. The results show that Cross Entropy estimates outperform the classical estimates, especially when the specification of the weights matrix is not similar to the true one. This result points to CE as a helpful technique to reduce the degree of arbitrariness imposed in the estimation of spatial models.

JEL Classification: C15, C21.

Keywords: Spatial econometrics, cross entropy econometrics, spatial models specifications, Monte Carlo simulations.

Actualización de matrices de pesos espaciales por Entropía Cruzada

El enfoque clásico para estimar modelos espaciales parte de la elección de una matriz de pesos espaciales que refleje la interacción entre las diferentes zonas. Se asume que la regla para definir esta matriz es que sea lo más parecida a la «verdadera» red de relaciones espaciales, pero para el investigador es difícil dilucidar cuándo la

* *Department of Applied Economics-University of Oviedo.* Phone: (+34)985105056; *e-mail: evazquez@uniovi.es.*

Received: 13 june 2011 / Accepted: 30 november 2011.

elección de esta matriz es correcta. Este paso clave en el proceso de estimación de modelos espaciales es una elección arbitraria, como Anselin (2002) señaló, y puede ser visto como uno de sus principales problemas metodológicos. En esta nota se propone no imponer los elementos de la matriz, sino su estimación basándose en la técnica de Entropía Cruzada (CE). Como las matrices de pesos espaciales son frecuentemente normalizadas por filas, cada una de ellas se puede entender como una distribución de probabilidad. La econometría basada en medidas de entropía es una herramienta útil para la obtención de distribuciones de probabilidad desconocidas, y su aplicación permite la estimación de los elementos de la matriz de pesos espaciales. Así, la matriz ya no depende de una elección impuesta por el investigador, sino de una estimación empírica. Este artículo compara los estimadores clásicos con los basados en medidas de entropía por medio de simulaciones de Monte Carlo en varios escenarios. Los resultados muestran que estas estimaciones superan a las obtenidas por estimadores tradicionales, especialmente cuando la especificación de la matriz no es similar a la real. Este resultado destaca la utilidad de las técnicas CE a la hora de reducir el grado de arbitrariedad impuesta en la estimación de modelos espaciales.

Clasificación JEL: C15, C21.

Palabras clave: Econometría espacial, econometría basada en entropía cruzada, especificación de modelos espaciales, simulaciones de Monte Carlo.

1. Introduction

The literature distinguishes several types of spatial models depending on the assumptions made 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 lagstructure.

Traditionally, for a set of N locations and T observations in time, the so-called-spatial lag model is written as:

$$y = X\beta + \rho Wy + \epsilon \tag{1}$$

$$y = [I - \rho W]^{-1}[X\beta + \epsilon] \tag{2}$$

where y is the $(NT \times 1)$ vector with the values of the dependent variable, W is the $(N \times N)$ matrix of *a priori* spatial weights which is assumed constant along time, X is a $(NT \times H)$ matrix of exogenous variables, β is a $(H \times 1)$ vector of parameters to estimate and ϵ is a $(NT \times 1)$ stochastic error. In addition, ρ is a spatial interaction parameter that measures how the variable y is spatially influenced. The weighting matrix W represents the spatial structure of the spillovers.

The selection of a specific spatial weights matrix 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 promotes fixingthe 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 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 W that maximizes the Moran statistic. Following a similar idea, Mur and Paelinck (2010) base their specification of W on the so-called complete correlation coefficients. Two papers by Getis and Aldstadt base their specification of 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 W based on the real structure of the spatial autocovariance, while Conley (1999) proposes the direct estimation of the spatial autocovariances. This data-driven selection of W has been, however, subject to strong criticism from authors supporting the exogenous approach (see, for example, Manski, 1993).

This note explores the use of Generalized Cross Entropy (GCE) econometrics to estimate such models. The GCE approach can be considered an extension of the Generalized Maximum Entropy estimator, which 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 lag model using this technique. The present paper will use the GCE 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 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 GCE methodology and shows how it can be applied to the context of spatial lag models. Section 3 evaluates the relative performance of the GCE techniqueusing a sampling experiment under different scenarios of sample size and degrees of divergencebetween the actual spatial network and the weighting matrix W specified in

---

[1] See Anselin (2002), p. 259.

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

the estimation. Section 4 shows an empirical application that illustrates how the proposed CE estimation procedure works with a real-world example. Finally, section 5 presents the concluding remarks.

2. Generalized Cross entropy econometrics: an overview

Entropy Econometrics (EE) techniques have interesting properties when dealing with ill-conditioned estimation problems (small samples or data sets affected by large collinearity). In Golan *et al.* (1996) or Kapur and Kesavan (1992) extensive descriptions of the entropy estimation approach can be found. Generally speaking, EE techniques are used to recover unknown probability distributions of random variables that can take M different known values. The estimate $\tilde{p}$ of the unknown probability distribution p must be as similar as possible to an appropriate *a priori* distribution q, constrained by the observed data. Specifically, the Cross-Entropy (CE) procedure estimates $\tilde{p}$ by minimizing the Kullback-Leibler divergence $D(p \,||\, q)$ (Kullback, 1959):

$$\underset{p}{\text{Min}}\ D(p \,||\, q) = \sum_{m=1}^{M} p_m \ln\left(\frac{p_m}{q_m}\right) \tag{3}$$

The divergence $D(p \,||\, q)$ measures the dissimilarity of the distributions p and q. This measure reaches its minimum (zero) when p and q are identical and this minimum is reached when no constrains are imposed. If some information (for example, observations on the variable) is available, each piece of information will lead to a Bayesian update of the *a priori* distribution q.

The underlying idea of the CE methodology can be applied for estimating the parameters of general linear models, which leads us to the so-called Generalized Cross Entropy (GCE). Let us suppose a variable y that depends on H explanatory variables x_h:

$$y = X\beta + \epsilon \tag{4}$$

Where y is a $(NT \times 1)$ vector of observations for y, X is a $(NT \times H)$ matrix of observations for the x_h variables, β is the $(H \times 1)$ vector of unknown parameters $\beta = (\beta_1, ..., \beta_H)$ to be estimated, and ϵ is a $(NT \times 1)$ vector with the random term of the linear model. Each β_h is assumed to be a discrete random variable. We assume that there is some information about its $M \geq 2$ possible realizations. This information is included for the estimation by means of a support vector $b' = (b_1, ..., b_M)$ with corresponding probabilities $p_h' = (p_{h1}, ..., p_{hM})$. The vector bis based on the researcher's a priori belief about the likely values of the parameter. 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 β can be written as:

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_H \end{bmatrix} = BP = \begin{bmatrix} b & 0 & \cdots & 0 \\ 0 & b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_H \end{bmatrix} \tag{5}$$

Where B and P have dimensions ($H \times HM$) and ($HM \times 1$) respectively. Now, the value of each parameter β_h is given by the following expression:

$$\beta_h = \boldsymbol{b}' \boldsymbol{p}_{\boldsymbol{h}} = \sum_{m=1}^{M} b_m \, p_{hm}; \; \forall h = 1, \ldots, H \tag{6}$$

For the random term, a similar approach is followed. Oppositely to other estimation techniques, GCE does not require rigid assumptions about a specific probability distribution function of the stochastic component, but it still is necessary to make some assumptions. $\boldsymbol{\epsilon}$ isassumed to have mean $E[\boldsymbol{\epsilon}] = 0$ and a finite covariance matrix. Basically, we represent our uncertainty about the realizations of vector $\boldsymbol{\epsilon}$ treating each element $\boldsymbol{\epsilon}_t$ as a discrete random variable with $J \geq 2$ possible outcomes contained in a convex set $v' = (v_1, \ldots, v_J)$, which for the sake of simplicity is assumed as common for all the $\boldsymbol{\epsilon}_t$. 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 $\boldsymbol{\epsilon}$ can be defined as:

$$\varepsilon = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_{NT} \end{bmatrix} = VU = \begin{bmatrix} v & 0 & \cdots & 0 \\ 0 & v & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{NT} \end{bmatrix} \tag{7}$$

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

$$\boldsymbol{\epsilon}_{nt} = \boldsymbol{v}' \boldsymbol{u}_{\boldsymbol{nt}} = \sum_{j=1}^{J} v_j u_{ntj} \tag{8}$$

And, consequently, model (4) can be transformed into:

$$y = XBp + Vu \tag{9}$$

So we need also to estimate the elements of matrix U (denoted by $\widetilde{u}_{tj}$) and the estimation problem for the general linear model (4) is transformed into the estimation of $H + NT$ probability distributions. For this estimation, once specified the a priori

probability distributions Q and U^0 respectively for P and U, the GCE problem is written in the following terms:

$$\underset{P,U}{\text{Min}}\ D(P,U \,\|\, Q,U^0) = \sum_{h=1}^{H}\sum_{m=1}^{M} p_{hm} \ln\left(\frac{p_{hm}}{q_{hm}}\right) + \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} u_{ntj} \ln\left(\frac{u_{ntj}}{u^0_{ntj}}\right) \tag{10a}$$

subject to:

$$y_{nt} = \sum_{h=1}^{H}\sum_{m=1}^{M} b_m p_{hm} x_{nt} + \sum_{j=1}^{J} v_j u_{ntj};\ \forall n,t \tag{10b}$$

$$\sum_{m=1}^{M} p_{hm} = 1;\ \forall h \tag{10c}$$

$$\sum_{j=1}^{J} u_{ntj} = 1;\ \forall n,t \tag{10d}$$

The restrictions in (10b) ensure that the posterior probability distributions of the estimates and the errors are compatible with the observations. The equations in (10c) and (10d) are just normalization constraints [3]. In other words, the CE solutions are obtained by minimizing the Kullback-Leibler divergence $D(P\|Q)$ between the unknown p_{hm} and the *a priori* q_{hm}. Similarly, for the estimation of u_{ntj} the divergence $D(U\|U^0)$ is minimized as well. In this case, the *a priori* probabilities are usually fixed as uniform ($u^0_{ntj} = \frac{1}{J}\ \forall n,t$), which is the natural point of departure to reflect the uncertainty about ϵ.

This GCE procedure can be extended for estimating spatial lagmodels 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 $z' = (z_1, ..., z_L)$, with corresponding probabilities $s' = (s_1, ..., s_L)$. The parameter ρ, consequently, can be estimated by GCE by means of this reparametrization. A similar idea was applied by Marsh and Mittelhamer (2004) for the case of spatial autoregressive models once a matrix of spatial weights 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 W matrix. This note suggests a solution where only one single spatial parameter ρ is defined, but the elements of a spatial weights matrix W will be updated from the *a priori* values specified.

---

[3] This GCE estimation procedure can be seen as an extension of the particular Generalized Maximum Entropy (GME) principle, given that the solutions of both approaches are the same when the *a priori* probability distribution contained in Q are all uniform.

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

$$W = \begin{bmatrix} 0 & w_{12} & \cdot & w_{1N} \\ w_{21} & 0 & \cdot & w_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1} & w_{N2} & \cdot & 0 \end{bmatrix} \tag{11}$$

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

$$\boldsymbol{y} = \boldsymbol{XBp} + (\boldsymbol{s}'\boldsymbol{z})\boldsymbol{Wy} + \boldsymbol{VU} \tag{12}$$

Now the empirical GCE program estimates H+2NT+1 probability distributions, in the following terms:

$$\begin{aligned} \underset{P,s,W,U}{\text{Min}}\, D(P,s,W,U \,||\, Q, s^0, W^0, U^0) = \sum_{h=1}^{H}\sum_{m=1}^{M} p_{hm} \ln\left(\frac{p_{hm}}{q_{hm}}\right) + \sum_{l=1}^{L} s_l \ln\left(\frac{s_l}{s_l^0}\right) \\ + \sum_{n=1}^{N}\sum_{i\neq 1}^{N} w_{ni} \ln\left(\frac{w_{ni}}{w_{ni}^0}\right) + \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} u_{ntj} \ln\left(\frac{u_{ntj}}{u_{ntj}^0}\right) \end{aligned} \tag{13a}$$

subject to:

$$y_{nt} = \sum_{h=1}^{H}\sum_{m=1}^{M} b_m p_{hm} x_{hnt} + \left(\sum_{l=1}^{L} s_l z_l\right)\left(\sum_{i\neq n}^{N} w_{ni} y_i\right) + \sum_{j=1}^{J} v_j u_{ntj};\ \forall n,t \tag{13b}$$

$$\sum_{m=1}^{M} p_{hm} = 1;\ \forall h \tag{13c}$$

$$\sum_{j=1}^{J} u_{ntj} = 1;\ \forall h,t \tag{13d}$$

$$\sum_{i\neq n}^{N} w_{ni} = 1;\ \forall n = 1, \ldots, N \tag{13e}$$

$$\sum_{l=1}^{L} s_l = 1 \tag{13f}$$

The GCE program above includes the Kullback divergence associated to the spatial parameter and to the weighting matrix in the objective function (13a). Equations (13c)-(13d) are again normalization constraints. Restriction (13b) forces the recovered probabilities to fit the observations of the dependent variable. This GCE program estimates, together with the parameters of the model, the elements of the matrix of spatial weights. These estimates (namely $\hat{w}_{ni}$) are the closest to the a priori assumptions made about the elements of the W matrix (w^0_{ni}) 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 W^0.

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

$$\hat{\rho} = \sum_{l=1}^{L} \hat{s}_l z_l \tag{14}$$

3. A numerical experiment

In this section, the performance of the GCE technique will be compared with other competing techniques in a scenario where the spatial structure that generates the data is given by a distance decay matrix. Under this specification, the elements of the W^{exp} matrix are defined as the following function:

$$w_{ni}^{\exp} = \exp(-d_{ni})$$

Where d_{ni} is the distance between the locations n and i, being $w^r_{ii} = 0$. We have simulated the spatial lag model $y = X\beta + \rho W^{exp}\, y + \in$ with 1,000 replications for two lattices of $N = 15$ and $N = 47$ locations. Specifically, for the case where $N = 15$, we have taken the 15 inland Spanish regions (Autonomous Communities) and when $N = 47$, the set of locations is formed by the 47 inland Spanish provinces. d_{ni} is the distance (km. by road) between the capital cities of regions (provinces) n and i. In our experiment, the error term is generated in each simulation as a $N(0.1)$ distribution. Matrix X is composed by one constant term and one regressor x. The values for the independent variable and for the parameters (kept constant throughout the simulations) are:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.50 \end{bmatrix};\ \rho = 0.25 \tag{15a}$$

$$x_{nt} \sim U(0.10);\ n = 1, \ldots, N;\ t = 1, \ldots, T \tag{15b}$$

In the experiment, the connectivity between the two sets of locations is given by the spatial pattern contained in the matrix W^{exp}, which is not necessarily equal to the weighting matrix used to estimate the model (W). For example, using the same idea of basing our spatial weights in a distance decay function, we could have specified the elements of our matrix as $w_{ni}^{sqr} = (d_{ni})^{-2}$; which is a specification commonly used in practice as well. We have introduced the possibility of divergence between the real matrix (W^{exp}) and the one specified in the model (W) assuming that $W = (1 - \eta)$ $W^{exp} + \eta W^{sqr}$; η (bounded between 0 and 1) is a scalar that reflects the degree of divergence between the real and the used spatial weighting matrices. If $\eta = 0$ this would indicate that the real and the specified matrix are exactly the same and the higher the value of η, the larger the misspecification of the spatial weighting matrix. In the limit, if $\eta = 1$ we'll be using a matrix of spatial weights completely different of the real one W^{exp}.

In this scenario for the sampling experiment, we compare the GCE approach with other rival procedures. In order to apply the GCE procedure to estimate models like (12), it is necessary to specify some support for the set of parameters and for the errors. For β_0 and β_1 the same support $b = (-1.1.1)$ has been set. Note that 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 $z = (-1.1.1)$. Finally, the support 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 (Pulkesheim, 1994; Golan, Judge and Miller, 1996).

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

Table 1 and 2 summarize the results of the experiment for the two sets of locations assuming that we have observations for $T = 10$ time periods. For each one of the competing estimators we have computed the mean of the estimates of β_0, β_1 and ρ throughout the 1,000 simulations (columns 1, 3 and 5 respectively) and their empirical variance, the mean absolute error of the estimates of β_0 and β_1 (columns 2 and 4) and the mean absolute error (column 6, which quantifies the average absolute differences between the actual ρ and its respective estimate).

Each row of Tables 1 to 4 contains a different value for the scalar η. As expected, the deviations between the actual and the estimated parameters for both methods are relatively low for values of η close to zero. However, the performance of the two competing estimation techniques is remarkably different as η grows. When the differences between the real W^{exp} and the W used in the estimation become larger, the GCE begins to yield comparatively better estimates than ML.

Table 1. Results of the numerical experiment
($N = 47$; $T = 10$; 1,000 replications)

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
		Average $\hat{\beta}_0$ *True* $\beta_0 = 0.75$	MAE_{β^0}	*Average* $\hat{\beta}_1$ *True* $\beta_1 = 0.5$	MAE_{β^1}	*Average* $\hat{\rho}$ *True* $\rho = 0.25$	$MAE\rho$
$\eta = 0.00$	ML	0.377 [0.047]	0.393	0.498 [0.001]	0.011	0.341 [0.002]	0.091
	GCE	0.240 [0.014]	0.520	0.502 [0.003]	0.016	0.350 [0.003]	0.105
$\eta = 0.20$	ML	0.132 [0.053]	0.630	0.498 [0.001]	0.011	0.397 [0.002]	0.147
	GCE	0.073 [0.014]	0.687	0.490 [0.003]	0.016	0.350 [0.003]	0.163
$\eta = 0.40$	ML	–0.190 [0.063]	0.954	0.498 [0.001]	0.012	0.469 [0.003]	0.219
	GCE	–0.121 [0.013]	0.887	0.476 [0.003]	0.016	0.474 [0.003]	0.229
$\eta = 0.60$	ML	–0.610 [0.079]	1.371	0.500 [0.001]	0.017	0.564 [0.004]	0.314
	GCE	–0.316 [0.011]	1.090	0.462 [0.003]	0.018	0.541 [0.005]	0.296
$\eta = 0.80$	ML	–1.111 [0.106]	1.857	0.503 [0.001]	0.027	0.675 [0.006]	0.425
	GCE	–0.474 [0.009]	1.256	0.450 [0.003]	0.021	0.596 [0.005]	0.351
$\eta = 1.00$	ML	–1.470 [0.144]	2.214	0.509 [0.001]	0.031	0.748 [0.008]	0.498
	GCE	–0.558 [0.008]	1.337	0.445 [0.003]	0.028	0.624 [0.006]	0.379

Table 2. Results of the numerical experiment
($N = 15$; $T = 10$; 1,000 replications)

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
		Average $\hat{\beta}_0$ *True* $\beta_0 = 0.75$	MAE_{β^0}	*Average* $\hat{\beta}_1$ *True* $\beta_1 = 0.5$	MAE_{β^1}	*Average* $\hat{\rho}$ *True* $\rho = 0.25$	$MAE\rho$
$\eta = 0.00$	ML	0.521 [0.176]	0.364	0.487 [0.001]	0.055	0.327 [0.006]	0.091
	GCE	0.216 [0.036]	0.521	0.487 [0.003]	0.016	0.356 [0.004]	0.109
$\eta = 0.20$	ML	0.316 [0.166]	0.371	0.487 [0.004]	0.057	0.376 [0.007]	0.134
	GCE	0.144 [0.036]	0.560	0.488 [0.003]	0.016	0.403 [0.004]	0.154
$\eta = 0.40$	ML	0.060 [0.111]	0.485	0.487 [0.004]	0.074	0.437 [0.010]	0.191
	GCE	0.059 [0.036]	0.601	0.472 [0.003]	0.016	0.449 [0.005]	0.199
$\eta = 0.60$	ML	–0.241 [0.241]	0.576	0.489 [0.004]	0.088	0.507 [0.014]	0.259
	GCE	–0.016 [0.009]	0.644	0.460 [0.004]	0.018	0.488 [0.005]	0.238
$\eta = 0.80$	ML	–0.537 [0.307]	0.684	0.493 [0.005]	0.104	0.573 [0.019]	0.324
	GCE	–0.073 [0.007]	0.668	0.452 [0.004]	0.021	0.514 [0.006]	0.264
$\eta = 1.00$	ML	–0.646 [0.384]	0.801	0.487 [0.004]	0.122	0.589 [0.023]	0.339
	GCE	–0.098 [0.007]	0.725	0.505 [0.003]	0.028	0.520 [0.006]	0.270

Besides, the experiment has been repeated now assuming a cross-section data set (i.e., $T = 1$) being the results summarized in Tables 3 and 4, which present the same structure as Tables 1 and 2.

Table 3. Results of the numerical experiment
($N = 47$; $T = 1$; 1,000 replications)

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
		Average $\hat{\beta}_0$ *True* $\beta_0 = 0.75$	MAE_{β^0}	*Average* $\hat{\beta}_1$ *True* $\beta_1 = 0.5$	MAE_{β^1}	*Average* $\hat{\rho}$ *True* $\rho = 0.25$	$MAE\rho$
$\eta = 0.00$	ML	1.034 [2.536]	1.238	0.485 [0.012]	0.087	0.200 [0.172]	0.327
	GCE	0.102 [0.002]	0.638	0.476 [0.007]	0.071	0.431 [0.014]	0.187
$\eta = 0.20$	ML	0.660 [2.307]	1.233	0.481 [0.011]	0.089	0.299 [0.161]	0.327
	GCE	0.091 [0.002]	0.659	0.469 [0.007]	0.069	0.445 [0.013]	0.199
$\eta = 0.40$	ML	0.293 [2.097]	1.258	0.483 [0.011]	0.088	0.396 [0.150]	0.344
	GCE	0.081 [0.002]	0.663	0.463 [0.006]	0.071	0.458 [0.012]	0.211
$\eta = 0.60$	ML	–0.066 [1.908]	1.355	0.485 [0.012]	0.087	0.492 [0.139]	0.375
	GCE	0.081 [0.002]	0.663	0.460 [0.006]	0.071	0.468 [0.012]	0.221
$\eta = 0.80$	ML	–0.418 [1.741]	1.506	0.488 [0.012]	0.088	0.585 [0.129]	0.422
	GCE	0.070 [0.002]	0.691	0.456 [0.006]	0.073	0.471 [0.012]	0.224
$\eta = 1.00$	ML	–0.763 [1.599]	1.710	0.490 [0.012]	0.088	0.677 [0.120]	0.478
	GCE	0.048 [0.002]	0.702	0.444 [0.006]	0.077	0.496 [0.010]	0.247

In brackets, empirical variance along the simulations.

Table 4. Results of the numerical experiment
($N = 15$, $T = 1$; 1,000 replications)

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
		Average $\hat{\beta}_0$ *True* $\beta_0 = 0.75$	MAE_{β^0}	*Average* $\hat{\beta}_1$ *True* $\beta_1 = 0.5$	MAE_{β^1}	*Average* $\hat{\rho}$ *True* $\rho = 0.25$	$MAE\rho$
$\eta = 0.00$	ML	0.721 [1.540]	0.971	0.500 [0.045]	0.169	0.255 [0.074]	0.223
	GCE	0.128 [0.003]	0.622	0.475 [0.011]	0.081	0.324 [0.018]	0.131
$\eta = 0.20$	ML	0.875 [1.602]	0.989	0.499 [0.045]	0.169	0.213 [0.078]	0.230
	GCE	0.134 [0.003]	0.616	0.483 [0.011]	0.081	0.332 [0.019]	0.135
$\eta = 0.40$	ML	1.030 [1.658]	1.026	0.498 [0.045]	0.170	0.171 [0.081]	0.238
	GCE	0.140 [0.003]	0.615	0.491 [0.011]	0.082	0.332 [0.019]	0.137
$\eta = 0.60$	ML	1.185 [1.709]	1.073	0.498 [0.045]	0.170	0.130 [0.084]	0.249
	GCE	0.145 [0.003]	0.616	0.498 [0.011]	0.083	0.333 [0.020]	0.138
$\eta = 0.80$	ML	1.340 [1.754]	1.131	0.497 [0.045]	0.170	0.088 [0.084]	0.266
	GCE	0.151 [0.004]	0.599	0.506 [0.011]	0.085	0.333 [0.020]	0.138
$\eta = 1.00$	ML	1.495 [1.794]	1.194	0.496 [0.045]	0.170	0.047 [0.084]	0.287
	GCE	0.157 [0.005]	0.593	0.513 [0.011]	0.087	0.334 [0.020]	0.138

In brackets, empirical variance along the simulations.

This is because, in the GCE, the specification of 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 W^{exp}. All in all, the results suggest that with perfect certainty about the actual spatial network W^{exp}, using the GCE technique proposed does not imply gains compared with ML. On the other hand, if we do not have clear evidences for imposing the right structure in the spatial network, using a GCE estimator seems to limit the estimation errors.

4. An empirical application: modeling labor productivity for the Spanish provinces

This section illustrates the performance of the entropy-based adjustment of the W matrix with a simple real world example. The objective will be to estimate a model for the $N = 47$ Spanish inland provinces (we exclude the Canary and Balearic Island off our analysis) where the labor productivity depends on an intercept and the stock of capital per worker and a spatial autoregressive component.

Annual data from 1995 to 2006 for the 47 provinces on gross domestic product and labor have been obtained from the Regional Accounts of Spain compiled by the Spanish Statistical Institute (INE). Data of the stock of private capital have been obtained from the BDMores database elaborated by the Spanish Ministry of Economy for the same time period. All the variables are in logs and, following Holtz-Eakin and Schwartz (1995), they are measured in differences to the initial year in order to capture the long-term relationships between the variables, provided that period t is sufficiently far from the initial period.

Specifically, the model to be estimated is:

$$y = X\beta + \rho Wy + \epsilon \tag{16}$$

Where for each time period t, y is a vector containing labor productivity (gross value added divided by the amount of labor) for each province and X is a matrix with the two exogenous variables of the model, consisting in the stock of labor (L) and the stock of privatephysical capital (K) in each province. Vector β contains two of the unknown parameters of the model; namely the labor (β_L) and capital (β_K) elasticities of a Cobb-Douglas aggregate production function. The model also includes a spatial autoregressive component measured by the parameter ρ which (as well as matrix W) is assumed constant along time.

We have applied the entropy-based adjustment proposed to estimate the model, which implies that an initial specification of matrix W is required. Initially, the elements of this matrix will be based on a distance decay function as $w_{ni}^{sqr} = (d_{ni})^{-2}$, being d_{ni} the distance (km. by road) between the capital cities of two provinces n and i. For applying the CE estimation to equation (16), it is necessary to specify some supports for parameters and for the errors. For all the parameters (β_L, β_K and ρ), we have considered different ranges of plausible values with 3 points. Specifically, the supports specified for have been (–1, 0, 1) that have been later expanded to (–5, 0, 5) and (–10, 0, 10) in order to check the sensitivity of the estimates to changes in the supports [4]. The traditional three-sigma rule is applied for specifying the supporting vectors for the error terms.

[4] Note that supporting vectors centered on zero for the spatial autoregressive parameter implies assuming that sometimes a raise in a neighbor province can generate either an increase in labor productivity ora decrease in other provinces' productivity.

Besides the point estimates, the GCE procedure allows for testing some hypotheses about the model confronting our estimates with the null hypothesis that the parameters are zero. This hypotheses testing can be done with the so-called entropy ratio, which follows a limiting χ^2 distribution. Let KL_R be the Kullback'sdivergence measure of a constrained problem, where the parameter is constrained to be 0 (at the centre of its support). Now let KL_U be the Kullback's divergence measure (objective function of the GCE program) without the restriction that the parameter is equal to zero. The entropy ratio statistic ER for testing the null hypothesis that the parameter is zero is = $2[KL_R - KL_U]$, which under the null hypothesis follows a limiting χ^2 distribution with K degrees of freedom, being K the number of restriction imposed. The results are summarized in Table 5

Table 5. CE estimation of equation (16)

	(1)	*(2)*	*(3)*	*(4)*
	$\hat{\beta}_L$	$\hat{\beta}_K$	$\hat{\rho}$	*% of abs. adjustment in W*
CE estimates				
b = (–1,0,1)	0.449**	0.141*	0.652**	8.978
b = (–5,0,5)	0.468**	0.118*	0.664**	9.055
b = (–10,0,10)	0.468**	0.119*	0.664**	9.057
ML estimation	0.472**	0.128**	0.653**	

* stands for estimates significantly different from 0 at a 10% level and ** stands for estimates significantly different from 0 at a 5% level based on a χ^2 distribution

The first two columns of Table 5 show the estimates for the β_L and β_K parameters and the third one reports the estimates for ρ under the different supporting vectors considered. In order to illustrate the adjustment applied to the initial W matrix based on a squared distance-decay function, the mean percentage of change (in absolute value) between the initial and the posterior matrices are reported in column (4). On average the CE estimation procedure modifies the cells around 9%, which could be considered as a relatively modest adjustment.

Regarding the parameter estimates, the maximum likelihood results are included for comparative purposes. Note that, generally speaking, there is not much variability in the results across the different specifications assumed, all of which get significantly positive estimates of the parameters of the model. The CE estimates are close to those obtained by maximum likelihood, although the later gets a capital elasticity significant at 5% whereas the CE estimation only gets evidence of a positive estimate at 10%. Note also that supporting vectors assumed for the spatial autoregressive parameter were centered on zero, which implies assuming that sometimes a raise in a neighbor province can generate either an increase in labor productivity or a decrease in other provinces' productivity. In this case, under any of the scenarios considered, it seems to be empirical evidences of a positive and significant contagion process among the Spanish regions concerning variations in output.

5. Concluding remarks

The specification of the spatial weighting matrix has been a 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 suggest 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 GCE estimator. Focusing in the so-called spatial lagmodels, a numerical experiment compares the performance of the proposed GCE 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 GCE gets better when the divergence between the actual and the a priori elements of *W* grows. The results of the numerical experiment are complemented with an application to real data of the method proposed, obtaining empirical evidence of a positive spatial autoregressive process among the aggregate production functions on the Spanish provinces between 1995 and 2006.

Acknowledgements

This work was supported by the *Ministerio de Innovación y Ciencia* of Spain with the grant number ECO2008-01617/ECO]. The author is thankful for the comments and suggestions made by an unknown referee and the attendants to the Fourth Seminar Jean Paelinck held in Oviedo in 2010.

References

Aldstadt, J., and Getis, A. (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.

— (2003): «Spatial externalities, spatial multipliers and spatial econometrics», *International Regional Science Review,* 26 (2), 153-166.

Bhattacharjee, A., and Jensen-Butler, C. (2006): «Estimation of spatial weights matrix, with and application to diffusion in housing demand», CRIEFF Discussion Papers 0519.

Boots, B., and Dufournaud, C. (1994): «A programming approach to minimizing and maximizing spatial autocorrelation statistics», *Geographical Analysis,* 26, 54-66.

Cantos, P.; Gumbau-Albert, M., and Maudos, J. (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 Hines, J. R. (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 Rodriguez-Valez, J. (2009): «Estimating spatial autoregressive models by GCE-GCE techniques», *International Regional Science Review,* 32 (2), 148-172.

Fraser, I. (2000): «An application of maximum entropy estimation: the demand for meat in the United Kingdom», *Applied Economics,* 321, 45-59.

Getis, A., and Aldstadt, J. (2004): «Constructing the spatial weights matrix using a local statistic», *Geographical Analysis,* 36 (2), 90-104.

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

Golan, A.; Perloff, J. M., and Shen, E. Z. (2001): «Estimating a demand system with nonnegativity constraints: Mexican meat demand», *The Review of Economics and Statistics,* 833, 541-550.

Holtz-Eakin, D., and Schwartz A. (1995): «Spatial productivity spillovers from public infrastructure: Evidence from state highways», *International Tax and Public Finance 2*: 459-68.

Kapur, J. N., and Kesavan, H. K. (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 Artis, M. (2004): «Regional externalities and growth: evidence from European regions», *Journal of Regional Science,* 44 1, 43-73.

Manski, C. F. (1993): «Identification of endogenous social effects: the reflection problem», *Review of Economic Studies* 60; pp. 531-542

Marsh, T. L., and Mittelhammer, R. C. (2004): Generalized maximum entropy estimation of a first order spatial autoregressive model, Advances in Econometrics, 18, 199-234.

Mittelhammer, R. C., and Cardell, N. S. (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.

Mur, J., and Paelinck, J. H. P. (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.