INSTITUTO DE ESTUDIOS LABORALES Y DEL DESARROLLO ECONÓMICO (ielde)
Facultad de Ciencias Económicas, Jurídicas y Sociales
Universidad Nacional de Salta (UNSa)
Salta
Argentina

Documentos de Trabajo

A Comparison Study on Criteria to Select the Most Adequate Weighting Matrix

Marcos Herrera
Jesús Mur
Manuel Ruiz

Otoño de 2017
Nº 18

ielde **– Facultad de Ciencias Económicas, Jurídicas y Sociales - UNSa**
http://www.economicas.unsa.edu.ar/ielde
UNSa: Av. Bolivia 5150, A4408FVY, Salta, Argentina
ISSN 1852-1118 (impreso), ISSN 1852-1223 (en línea)

A Comparison Study on Criteria to Select the Most Adequate Weighting Matrix

Marcos Herrera[*], Jesús Mur[†], Manuel Ruiz[‡]

Abstract

The practice of spatial econometrics revolves around a weighting matrix, which is often supplied by the user on previous knowledge. This is the so called **W** issue. Probably, the aprioristic approach is not the best solution altough, nowadays, there few alternatives for the user. Our contribution focuses on the problem of selecting a **W** matrix from among a finite set of matrices, all of them deemed appropriate for the case. We develop a new and simple method based on the Entropy corresponding to the distribution of probability estimated for the data. Other alternatives, which are common in current applied work, are also reviewed. The paper includes a large Monte Carlo resolved in order to calibrate the effectiveness of our approach compared to the others. Two well-known case study are also included.

JEL Classification: C4, C5, R1.

Keywords: Weights matrix, Model Selection, Entropy, Monte Carlo.

[*]CONICET-IELDE, National University of Salta, Salta, 4400, Argentina (e-mail: mherreragomez@conicet.gov.ar.)
[†]Department of Economic Analysis, University of Zaragoza, Zaragoza, Spain (e-mail: jmur@unizar.es)
[‡]Department of Computing and Quantitative Methods, Technical University of Cartagena, Cartagena, Spain (e-mail: manuel.ruiz@upct.es)

1 Introduction

Let us begin with a mantra: the weighting matrix is the most characteristic element in a spatial model. This commonplace is probably true and we have no problem with that. As known, spatial models deal primarily with phenomena such as spillovers, trans-boundary competition or cooperation, flows of trade, migration, knowledge, etc. in very complex networks. Rarely does the user know about how these events operate in practice. Indeed, they are mostly unobservable phenomena which are, however, required to build the model. Traditionally the gap has been solved by providing externally this information, in the form of a weighting matrix. However, we must recognize that the **W** alternative is not the unique solution to deal with such kind of unobservables (Oud and Folmer, 2008, for example, develop a latent variables approach that does not need of **W**), but is the most simple.

Hays et al. (2010) give a sensible explanation about the preference for a predefined **W**. Network analysts are more interested in the formation of networks, taking units attributes and behaviors as given. This is spatial dependence due to selection, where relations of homophily and heterophily are crucial. The spatial econometricians are more interested in what they call *'computing the effects of alters actions on ego's actions through the network*'; in this case, the patterns of connectivity are taken as given. This form of spatial dependence is due to influence and the notions of contagion and interdependence are capital. So, if the network is predefined, why not supplying it externally?

However, beyond this point, the **W** issue han been frequent cause of dispute. In the early stages, terms like 'join' or 'link' were very common (for instance, in Moran, 1948, or Whittle, 1954). The focus at that time was mainly on testing for the presence of spatial effects, for which is not so important the specification of a highly detailed weighting matrix; contiguity, nearness, rough measures of separation are appropriate notions for that purpose. The work of Ord (1975) is a milestone in the evolution of this issue because of its strong emphasis on the task of modelling spatial relationships. It is evident that the weights matrix needs more attention if we want to avoid estimation biases and wrong inference. Anselin (1988, 2002) puts the **W** matrix in the center of the debate about specification of spatial models, but, after decades of practicing, the question is still rather obscure.

The meaning of the so-called **W** is clear: we need to '*determine which ... units in the spatial system have an influence on the particular unit under consideration ... expressed in notions of neighborhood and nearest neighbor*' relations, in words of Anselin (1988, p.16) or '*to define for any set of points or area objects the spatial relationships that exist between them*' as stated by Haining (2003, p. 74). The problem is how should it be done.

Roughly speaking, we may distinguish two approaches: (i) specifying **W** exogenously; (ii) estimating **W** from data. The exogenous approach is by far the most common and includes, for example, use of a binary contiguity criterion, k-nearest neighbours, kernel functions based on distance, etc. The second approach uses the topology of the space and the nature of the data, and takes many forms.We find ad-hoc procedures in which a predefined objective guides the search such as the maximization of Moran's I in Kooijman (1976) or the local statistical model of Getis and Aldstadt (2004). Benjanuvatra and Burridge (2015) develop a quasi maximum-likelihood, QML, algorithm to estimate the weights in $\mathbf{W}$ assuming partial knowledge about the form of the weights. More flexible approaches are possible if we have panel information such as in Bhattacharjee and Jensen-Butler (2013) or Beenstock and Felsenstein (2012). Endogeneity of the weight matrix is another topic introduced recently in the field (i.e., Qu and f. Lee, 2015), which connects with the concept of *coevolution* put forward by Snijders et al. (2007) whose basis is difficult to object: in the long run, network connectivity must evolve with the model, that is, with the endogenous variable. Part of the recent literature on spatial econometrics revolves around endogeneity, but our contribution pertains to the exogenous approach where remains most part of the applied research.

Before continue, we may wonder if the $\mathbf{W}$ issue, even in our context of pure exogeneity, is really a problem to worry for or it is the *biggest myth* of the discipline as claimed by LeSage and Pace (2014). Their argument is that only dramatic different choices for $\mathbf{W}$ would lead to significant differences in the estimates or in the inference. We partly agree with them in the sense that is a bit silly to argue whether it is better the 5 or the 6 nearest-neighbor matrix; surely there will be almost no difference between the two. However, there is consistent evidence, obtained mainly by Monte Carlo (Florax and Rey, 1995; Franzese Jr and Hays, 2007; Lee and Yu, 2012; Debarsy and Ertur, 2016) showing that the misspecification of $\mathbf{W}$ (bigger errors than just marginal mismatches) tends to attenuate, or inflate, the estimates of the coefficients of spatial dependence with an inverse impact on the slope coefficients. Moreover, the magnitude of the bias increases for the estimates of the marginal direct/indirect effects. So, we are not pretty sure that '*far too much effort has gone into fine-tunning spatial weight matrices*' as stated by LeSage and Pace (2014). Our impression is that any useful result is welcomed in this field and especially we need practical, clear guides to approach the problem.

Another question of concern are the criticisms of Gibbons and Overman (2012). As said, it is common in spatial econometrics to assume that the weighting matrix is known, which is the cause of (weak) identification problems in the models; this flaw extends to the instruments, moment conditions, etc. There is little to say in relation to this point. In fact, spatial parameters (i.e., ρ) and weighting matrix, $\mathbf{W}$, are jointly identified (we do estimate $\rho\mathbf{W}$). Hays et al. (2010) and Bhattacharjee and Jensen-Butler (2013) agree in this point.

Bavaud (1998, p. 153), given this state of confusion, was very skeptic, '*there is no such thing as "true", "universal" spatial weights, optimal in all situations*' and continues by stating that the weighting matrix '*must reflect the properties of the particular phenomena, properties which are bound to differ from field to field*'. We share his skepticism about the concept of truth. Perhaps it would suffice with a 'reasonable' weighting matrix, the best among those that we are considering. In practical terms, this means that the problem of selecting a weighting matrix can be interpreted as a problem of model selection. In fact, different weighting matrices result in different spatial lags of the endogenous or the exogenous variables included in the model. Finally, different equations with different regressors, or different structures, amounts to a model selection problem.

Our impression is that there is a general agreement in the current applied literature in that approach to the **W** issue; however, we would like to extend the discussion a bit further. Section 2 revises four selection criteria that fit well to the problem of selecting a weighting matrix from among a finite set of them. Section 3 presents the main features of the Monte Carlo solved in the fourth Section. Section 5 includes two well known case studies which are revised in the light of our findings. Sixth Section concludes.

2 Criteria to select a W matrix from among a finite set

As said, the **W** problem has been present in the literature since very early. However the case of choosing one matrix from among a finite set of them is relatively recent.

Anselin (1984) poses formally the problem suggesting a *Cox* statistic derived in a framework of non-nested models. Leenders (2002), on this basis, elaborates a J-test using classical augmented regressions. Later on, Kelejian (2008) extends the approach of Leenders to a SAC model, with spatial lags of the endogenous variable and in the error terms, using GMM estimates. Piras and Lozano (2012) confirm the adequacy of the J-test to compare different weighting matrices stressing that we should make use of a full set of instrument to increase GMM accuracy. Burridge and Fingleton (2010) show that the Chi-square asymptotic approximations for the J-tests produces irregular results, excessively liberal or conservative in a series of leading cases; they suggest a bootstrap resampling approach. Burridge (2012) focuses on the propensity of the spatial GMM algorithm to deliver spatial parameter estimates lying outside the invertibility region which, in turn, affects the bootstrap; he suggest the use of a QML algorithm to remove the problem. Kelejian and Piras (2011) generalized and modify the original version of Kelejian to account for all the available information, according to the findings of Piras and Lozano. Finally, Kelejian and Piras (2016) adapt the J test to a panel data setting with unobserved fixed effects and additional endogenous variables which reinforces the adequacy of the GMM framework.

Another milestone in the J test literature is Hagemann (2012), who copes with the reversion problem originated by the lack of a well defined null hypothesis in the test. He introduces the minimum J test, MJ. His approach is based on the idea that if there is a finite set of competing models, only the model with the smallest J statistic can be the correct one. In this case, the J statistic will converge to the Chi-square distribution but will diverge if none of the models is correct. The author proposes a wild bootstrap to test if the model with the minimum J is correct. This approach has been applied by Debarsy and Ertur (2016) to a spatial setting with good results.

Our intention is to use only the first part of the procedure of Hagemann, given that we know that there is a correct model in the Monte Carlo that follows. So assuming that we have a collection of m competing weighting matrices, such as: $\mathcal{W} = \{\mathbf{W}_1; \mathbf{W}_2; ...; \mathbf{W}_m\}$ for the same model, then:

1. We are going to estimate the m models; each one corresponds to a different weighting matrix belonging to $\mathcal{W}$. Following Burridge (2012) and given that our interest lies on the small sample case, the models are estimated by QML.

2. For each model, we obtain the battery of J statistic as usual, after estimating, also by QML, the corresponding extended equations.

3. The chosen matrix is the one associated to the minimum J statistic. We do not test if this matrix is really the correct matrix.

Another popular method for choosing between models deals with the so-called *Information Criteria.* Most are developed around a loss function, such as the *Kullback-Leibler*, KL, quantity of information which measures the closeness of two density functions. One of them corresponds to the true probability distribution that generated the data, obviously not known, the other is the distribution estimated from the data. The criteria differ in the role assigned to the aprioris and in the way of solving the approximation to the unknown true density function (Hansen, 2005). The two most common procedures are the AIC (Akaike, 1973) and the Bayesian BIC criteria (Schwarz et al., 1978). The first compares the models on equal basis whereas the second incorporates the notion of model of the null. Both criteria are characterized by their lack of specificity in the sense that the selected model is the closest to the true model in terms of KL without any other consideration (i.e., a good global fit does not mean that the model is the best alternative to estimate the parameters of interest; Pötscher, 1991). AIC and BIC lead to single expressions that depend on the accuracy of the ML estimation of the models (in this sense, they are parametric methods) plus a penalty term related to the number of parameters entering the model, such as the following:

$$\left.\begin{array}{ll} AIC(k): & -2l(\widetilde{\gamma}) + 2k, \\ BIC(k): & -2l(\widetilde{\gamma}) + k\log(n), \end{array}\right\} \tag{1}$$

where $l(\widetilde{\gamma})$ means the estimated log-likelihood at the ML estimates, $\widetilde{\gamma}$, k is the number of non-zero parameters in the model and n the number of observations. For the case that we are considering the models only differ in the weighting matrix, so k and n are the same for every case. This means that the decision depends on the estimated log-likelihood or, what is the same, on the balance between the estimated variance and the Jacobian them. Note that, for a standard spatial model of, i.e., SLM type we can write: $l(\widetilde{\gamma}) \propto \log\left[\frac{1}{\widetilde{\sigma}^n}\left|I-\widetilde{\rho}\mathbf{W}\right|\right]$, being σ the standard deviation and ρ the corresponding spatial dependence coefficient. To minimize any of the two statistics in (1) the objective is to maximize the concentrated estimated log-likelihood, $l(\widetilde{\gamma})$. The same as before, the *Information Criteria* approach implies:

1. Estimate by ML each one of the m models corresponding to each weighting matrix in $\mathcal{W}$.
2. For each model, we obtain the corresponding AIC statistic (BIC produces the same results).
3. The matrix in the model with minimum AIC statistic should be chosen.

An important part of the recent literature on spatial econometrics has Bayesian basis, which extends to the topic of choosing a weighting matrix. The point of departure, once again, is to recognize that differences in the weighting matrix, everything else constant, amounts to different models. The Bayesians are well equipped to cope with these type of problems using the concept of *posterior probability* as the basis for taking a decision. There are excellent reviews that can be consulted such as Hepple (1995a,b, 2004), Besag and Higdon (1999) and especially, LeSage and Pace (2009). For the sake of completeness, let us highlight the main points in this approach.

The analysis is made conditional to a model, which is not under discussion. Thus, we have a collection of m weighting matrices in $\mathcal{W}$, a set of k parameter in γ, some of which are of dispersion, σ, others of position, β, and others of spatial dependence, ρ and θ, and a panel data set with nT observations in y. The point of departure is the joint probability of all these elements represented as:

$$p\left(\mathbf{W}_i;\gamma;y\right)=\pi\left(\mathbf{W}_i\right)\pi\left(\gamma\mid\mathbf{W}_i\right)L\left(y\mid\gamma;\mathbf{W}_i\right), \tag{2}$$

where $\pi\left(\cdot\right)$ are the prior distributions and $L\left(y\mid\gamma;\mathbf{W}_i\right)$ the likelihood for y conditional on the parameters and the matrix. Bayes' rule leads to the posterior joint probability for matrices and parameters:

$$p\left(\mathbf{W}_i;\gamma\mid y\right)=\frac{\pi\left(\mathbf{W}_i\right)\pi\left(\gamma\mid\mathbf{W}_i\right)L\left(y\mid\gamma;\mathbf{W}_i\right)}{L\left(y\right)}, \tag{3}$$

whose integration over the space of parameters, $\gamma\in\Upsilon$, produces the posterior probability for matrix $\mathbf{W}_i$:

$$p\left(\mathbf{W}_i\mid y\right)=\int\limits_{\Upsilon}p\left(\mathbf{W}_i;\gamma\mid y\right)d\gamma. \tag{4}$$

The presence of spatial structures in the model complicates the resolution of (4) which usually requires of numerical integration. The priors are always a point of concern in this approach and, usually, practitioners prefer diffuse priors. LeSage and Pace (2009, Section 6.3) suggest $\pi\left(\mathbf{W}_i\right)=\frac{1}{m}\ \forall i$, a NIG

conjugate prior for β and σ where $\pi_\beta(\beta \mid \sigma) \sim N\left(\beta_0; \sigma^2 (\kappa X'X)^{-1}\right)$, being X the matrix of the exogenous variables in the model, and $\pi(\sigma)$ a inverse gamma, $IG(a,b)$. For the parameter of spatial dependence they suggest a $Beta(d,d)$ distribution, being d the amplitude of the sampling space of ρ. For example, the defaults in the MATLAB codes of LeSage (2007) are $\beta_0 = 0$, $\kappa = 10^{-12}$ and $a = b = 0$. In sum, the *Bayesian* approach implies the following:

1. Fix the priors for all the terms appearing in the equation. In this point, we have followed the suggestions of LeSage and Pace.

2. For each matrix, obtain the corresponding posterior probability of (4) for which we need (i) solve the ML estimation of the corresponding model and (ii) solve the numerical integration of (4).

3. The matrix chosen will be that associated with the highest posterior probability.

Our own proposal to deal with the selection problem is based on the notion of *Entropy*, who dates back to Shannon and Weaver (1949). In our statistical framework, *Entropy* is viewed as a measure of the information carried by a distribution of probability. Let us assume an univariate continuous variable, y, whose probability density function is $p(y)$; then, *Entropy* is defined as:

$$h(p) = -\int_I p(y) \log p(y) dy, \tag{5}$$

being I the domain of the random variable y. Higher *Entropy* means less information or, what is the same, more uncertainty about y. Our case fits well with Shannon's framework: we observe a random variable, y, and there are a finite set of rival distribution functions capable of having generated the data. The limits of our decision problem are also well defined because each distribution function differs from the others only in the weighting matrix; the other elements are the same. However, we are not interested in the Laplacian principle of indifference (select the density with maximum *Entropy*, or less informative, to avoid the use of unwarranted information). Quite the opposite: given that in our case there is no unwarranted information, we are looking for the more informative probability distribution so our objective is to minimize *Entropy*.

As with the other three cases, the application of this principle requires the complete specification of the distribution function, which means that the user knows the form of the model (equations 7 to 9 below, except the **W** matrix); additionally we add a Gaussian distribution. Next, we should remind that for the case of a $(n \times 1)$ multivariate normal random variable, $y \sim N(\mu; \Sigma)$, the entropy is: $h(y) = \frac{1}{2}\left(n + \log((2\pi)^n |\Sigma|)\right)$. This measure does not depend, directly, on first order moments (parameters of position of the model) but on second order moments (dependence and dispersion parameters). For example, in the case of the SLM of (9) below, the entropy is:

$$h(y)_{SDM} = \frac{1}{2}\left(nT + \log((2\pi\sigma^2)^{nT}\left|(B'B)^{-1}\right|\right)), \tag{6}$$

where $B = (I - \rho\mathbf{W})$. Note that the covariance matrix for y in the SDM is $V(y) = B^{-1}V(u)B^{'-1}$. If u is indeed a white noise random term with variance σ^2, the covariance matrix of y is simply $V(y) = \sigma^2(B'B)^{-1}$. Let us note that the covariance matrix of y in the $SDEM$ of (8) coincides with that of the SLM case; same happens with the SDM of (7). In sum, the expression of the *Entropy* for the three models (SLM, SDM and $SDEM$) is formally the same as that in (6).

In order to apply the *Entropy* criterion we must must go through the following steeps:

1. Estimate each one of the m versions of the model that we are considering. As said, each models differs only in the weighting matrix. We maintain the ML estimation algorithm for reasons given above.

2. For each model, we obtain the corresponding value of the *Entropy*, in the $h_i;\ i = 1, 2, ..., m$ statistic.

3. The decision criterion consists in choosing the weighting matrix corresponding to the model with minimum value of the *Entropy*.

3 Description of the Monte Carlo

This part of the paper is devoted to the design of the Monte Carlo conducted in the next Section in order to to calibrate the performance of the four criteria presented so far for selecting $\mathbf{W}$: the MJ test, the *Bayesian* approach, the AIC criterion and the *Entropy* measure. The objective of the analysis is to identify and select the true matrix, which intervened in the generation of the data. Moreover, our focus is on small sample sizes. As will be clear in the next Section, the four criteria have good behaviour even in small samples; so it is not necessary to simulate very large sample sizes

We are going to simulate a panel setting, with three of the most common *DGPs* in the current literature on applied spatial econometrics; namely, the spatial Durbin Model, SDM of (7), the spatial Durbin error model, $SDEM$ in expression (8) and the spatial lag model of (9), SLM.[1]

[1]Main conclusions can be extended to other processes like the spatial error model, which are not replicated here (details on request from the authors).

$$y_{it} = \beta_0 + \rho \sum_{j=1}^{n} \omega_{ij} y_{jt} + x_{it}\beta_1 + \theta \sum_{j=1}^{n} \omega_{ij} x_{jt} + \varepsilon_{it}, \quad (7)$$

$$y_{it} = \beta_0 + x_{it}\beta_1 + \theta \sum_{j=1}^{n} \omega_{ij} x_{jt} + u_{it},\ u_{it} = \rho \sum_{j=1}^{n} \omega_{ij} u_{jt} + \varepsilon_{it}. \quad (8)$$

$$y_{it} = \beta_0 + \rho \sum_{j=1}^{n} \omega_{ij} y_{jt} + x_{it}\beta_1 + \varepsilon_{it}, \quad (9)$$

Only one exogenous regressor, x variable, appears in the right hand side of the equations whose observations are obtained from a normal distribution, $x_{it} \sim i.i.d.N\left(0;\sigma_x^2\right)$, where $\sigma_x^2 = 1$; the same applies with respect to the error terms: $\varepsilon_{it} \sim i.i.d.N\left(0;\sigma_\varepsilon^2\right)$, where $\sigma_\varepsilon^2 = 1$. The two variables are not related, $E\left(x_{it}\varepsilon_{it}\right) = 0$. Our space is made of hexagonal pieces which are arranged regularly, one next to the others without discontinuities nor empty spaces.

At least one weighting matrix appears in the three equations, which plays a central role in the functioning of the model. As said before, the weighting matrix is not observable and the user must take actions to resolve the uncertainty. The decision problem consists in choosing one matrix from among a finite set of alternatives which in our simulation are composed by only three candidates: $\mathbf{W}_1$ is built using the traditional contiguity criterion between spatial units; the weights in $\mathbf{W}_2$ are the inverse of the distance between the centroids of the spatial units, $\mathbf{W}_2 = \left\{\omega_{ij} = \frac{1}{d_{ij}}; i \neq j\right\}$; whereas $\mathbf{W}_3$ incorporates a cut-point in the networks of connections of $\mathbf{W}_2$, so that $\mathbf{W}_3 = \left\{\omega_{ij} = \frac{1}{d_{ij}}; i \neq j\ if\ j \in N_8(i);\ 0\ otherwise\right\}$ being $N_8(i)$ the set of 8 nearest neighbors to i. Following usual practice, every matrix has been row-standardized. To keep things simple, the same weighting matrix intervenes with the endogenous and exogenous variables in (7) and with the exogenous and error terms in (8). Due to the row-standardization, the three matrices are non nested in the sense that the weights are different among them.

Only three different small cross-sectional sample sizes, n, have been used $n \in \{25, 49, 100\}$; that is enough because, as shown later, higher values of this parameter adds nothing to the information about the $\mathbf{W}$ dilemma. For the same reason, the number of cross-sections in the panel, T, are limited to only three, $T \in \{1, 5, 10\}$. The values for the coefficient of spatial dependence, ρ, ranges from negatives to positives, $\rho = \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$. Other global parameters are those associated with the constant term, $\beta_0 = 1$, the x variable, $\beta_1 \in \{1, 5\}$, and its spatial lag, $\theta \in \{1, 5\}$.

In sum, each case consists in:

- Generate the data using a given weighting matrix, $\mathbf{W}_k$, $k = 1, 2, 3$ and a spatial equation, SLM, SDM or $SDEM$. There are 216 cases of interest for each equation (6 values in ρ, 3 values in n, 3 values in T, 2 values in β_1 and 2 values in θ).

- The spatial equation is assumed to be known so the model can be estimated by maximum likelihood, ML, once the user supplies a $\mathbf{W}$ matrix.
- Compute the four selection criteria, MJ, *Posterior probability*, *Entropy* and AIC for the three alternative weighting matrices for each draw.
- Select the corresponding matrix according to each criterion and compare the result with the *true* matrix in the DGP.
- The process has been replicated $1,000$ times.

Note that the selection of the matrix is made conditional on a correct specification of the equation. We are perfectly aware that this dichotomy is artificial; in fact, both decisions are intimately related because the same matrix, but in different equations, plays different roles and bears different information. However, this point is not further developed in the present paper. In order to give some intuition, we include the results corresponding to the case of a wrong selection of the spatial equation (i.e, estimate a SDM model whereas the true model in the DGP is a $SDEM$).

4 Results of the Monte Carlo

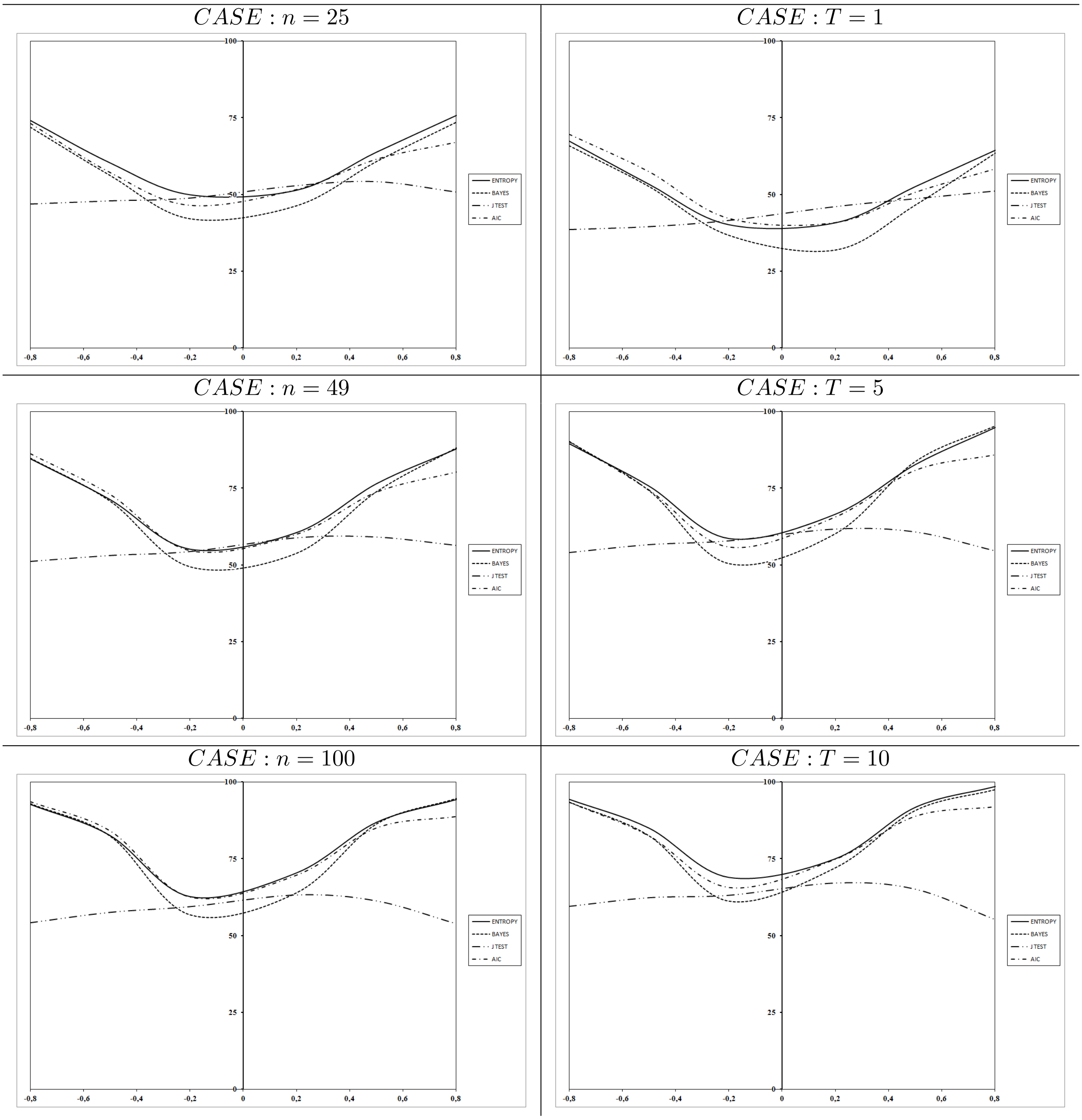
The Monte Carlo has provided us with a lot of information; the purpose of this Section is to summarize main results. Let us advance an little spicy: in strictly quantitative terms, the *Entropy* measure is the best criterion. What is more surprising, the *Bayesian* approach is only marginally better than the AIC but only when the amount of information is large and there is positive spatial correlation; finally, the MJ test is the worse alternative among the four criteria. The last two observations are really striking for us given the strong support that the two criteria have received in the literature. Table 1, where appears the percentage of correct selections attained by each criterion after aggregating by n, T and design of the DGP, offers a quick overview of our results. A cell in bold indicates that the respective criterion reaches the maximum rate of correct selections.

Table 1: Percentage of correct selections. Aggregated results

ρ	$h(y)$	*Bayes*	MJ	AIC
-0.8	83.8	83.2	50.7	**84.4**
-0.5	**71.4**	69.7	52.8	**71.4**
-0.2	**55.9**	49.4	54.2	54.6
0.2	**60.8**	54.6	58.3	60.5
0.5	**75.7**	73.6	58.2	73.5
0.8	**85.9**	85.4	53.6	78.7

Entropy dominates in 5 out of the 6 cases presented in the Table, and is the second in the sixth case; *AIC* leads in two cases, is second in two and third in another two cases. *Bayes* does bad for small values of the spatial coefficient (is fourth in ± 0.2) and the *power curve* of MJ test is very flat. If we disaggregate by number of cross-sections or number of spatial units, the conclusion has to be only slightly modified as is clear in Figure 1.

Figure 1: Percentages of correct selections by n or T



Note the asymmetry in all the curves and the weird behaviour of the MJ test going down at the extremes of the interval for ρ. The other three criteria react positively to increases in the sample size (both in n or in T). Overall, the improvement appear quicker according to T, number of cross-sections, than to n, specially for high values of the spatial coefficient.

Tables 2 to Table 5 present the details by type of *DGP*. A quick look at the Tables reveals that bold figures are concentrated, first, in the *Entropy* criterion columns, then in the *AIC* columns.

The predominance of the *Entropy* criterion extends quite regularly (the exception is the *SDEM* process where *AIC* has better results). That prevalence remains for correctly specified models, as in Tables 2, 3 and 4 and also for misspecified equations, as in Table 5; for negative but especially for positive values of the spatial coefficient, for small and also for big cross-sections and for simple to large panels. Overall, *Entropy* shows the highest rates in 48% of the cases in the 4 Tables.

Table 2: Average percentage of correct selections. DGP: SDM. Equation estimated: SDM.

Aggregated by cross-section, sample size (n)						**Aggregated by time series, sample size (T)**					
	ρ	$h(y)$	*Bayes*	MJ	AIC		ρ	$h(y)$	*Bayes*	MJ	AIC
$n = 25$	−0.8	78.1	77.8	52.4	**79.6**	$T = 1$	−0.8	67.4	66.2	39.4	**68.8**
	−0.5	**62.9**	62.5	52.0	61.8		−0.5	54.4	54.3	38.5	**57.5**
	−0.2	**53.5**	48.7	53.1	50.2		−0.2	41.1	38.4	40.0	**41.7**
	0.2	61.5	59.8	**65.0**	61.2		0.2	43.2	35.8	**48.4**	40.8
	0.5	**74.7**	56.5	50.8	72.1		0.5	**56.5**	50.8	55.2	54.3
	0.8	**84.3**	81.7	74.5	75.5		0.8	**69.7**	68.1	63.4	63.8
$n = 49$	−0.8	88.9	88.7	57.6	**90.1**	$T = 5$	−0.8	91.9	93.0	62.3	**93.5**
	−0.5	76.4	77.5	58.6	**78.7**		−0.5	79.4	**80.2**	63.7	79.5
	−0.2	**59.6**	55.5	58.6	58.8		−0.2	**63.3**	57.3	62.4	60.1
	0.2	71.0	67.9	**73.1**	70.0		0.2	**79.1**	76.6	78.1	76.1
	0.5	**84.1**	81.7	81.6	82.0		0.5	92.3	**92.5**	87.1	89.8
	0.8	93.3	**93.8**	88.1	87.4		0.8	**98.4**	98.3	88.2	89.9
$n = 100$	−0.8	**94.4**	94.3	63.9	95.2	$T = 10$	−0.8	97.3	97.4	69.2	**97.9**
	−0.5	87.3	87.2	66.6	**88.7**		−0.5	**88.8**	88.7	72.1	87.9
	−0.2	**67.6**	61.9	62.8	66.6		−0.2	**72.3**	66.5	68.2	69.8
	0.2	**80.5**	76.4	79.4	77.1		0.2	86.6	87.7	86.4	**87.4**
	0.5	**91.9**	90.5	85.6	89.7		0.5	97.0	**97.5**	92.9	95.4
	0.8	**97.3**	96.3	89.5	92.4		0.8	**99.8**	**99.8**	94.6	95.8

The complete percentages for the 216 cases corresponding to the four pairs of *DGP*/Estimated-equation appear in Tables A1 to A4 in the Appendix. Using these results, it is worth noting the good results attained in the case of small samples ($n = 25$ and $T = 1$) where the average rate of correct selections fluctuates above 40% for *Entropy* and *AIC* criteria (a little worse for the other two). The percentage exceeds 60% at the extremes of the spatial parameter interval, ± 0.8. The average rate improves by 50%, until 65% to 75%, for the case of $n = 25$ and $T = 5$ and almost doubles when $T = 10$, where we find a majority of cases with a rate of correct selections above 90%. In general, the rate of correct selections is nearly 100%, using 5 to 10 cross-sections.

In a similar vein, the increase in the cross-sectional size, n, maintaining fixed the number of cross-sections, T, also has positive effects in the four criteria. The rate of correct selections for the case of a hundred of spatial units is above 70%, on average, for the case of a single cross-section, but these percentages improve quickly if the time dimension of the panel increases.

The value of parameter β_1, as expected, has an almost negligible impact in the four criteria, but the signal of θ_1 plays a crucial role in the $SDEM$ case. Another interesting feature is the asymmetry of the selection curves. Negative spatial dependence helps to better detect the correct weighting matrix, especially when the number of panel cross-sections is also low. The asymmetry appears in all criteria, except in the MJ test whose behavior, contrary to the others, is worse in case of negative values in parameter ρ. The impact of the asymmetry disappears with higher values of T as well as of θ_1.

Table 3: Average percentage of correct selections. DGP: SDEM. Equation estimated: SDEM.

Aggregated by cross-section, sample size (n)						**Aggregated by time series, sample size (T)**					
	ρ	$h(y)$	*Bayes*	MJ	AIC		ρ	$h(y)$	*Bayes*	MJ	AIC
$n = 25$	−0.8	80.5	77.3	56.7	**82.5**	$T = 1$	−0.8	66.7	65.3	42.5	**70.4**
	−0.5	**69.6**	65.2	57.5	**69.6**		−0.5	55.4	55.6	44.0	**62.1**
	−0.2	**59.6**	52.5	56.5	58.2		−0.2	42.5	42.8	43.8	**49.3**
	0.2	55.6	52.4	56.9	**57.7**		0.2	39.5	36.1	**45.7**	43.4
	0.5	63.5	62.6	55.7	**63.7**		0.5	**49.5**	45.4	46.5	48.7
	0.8	**74.4**	73.8	54.0	67.0		0.8	**59.3**	58.1	48.9	53.2
$n = 49$	−0.8	88.1	88.5	64.5	**91.0**	$T = 5$	−0.8	94.0	94.7	71.1	**95.4**
	−0.5	78.2	78.8	65.6	**81.9**		−0.5	84.1	84.8	72.5	**84.9**
	−0.2	64.6	62.6	65.2	**66.3**		−0.2	70.2	67.1	**71.6**	69.6
	0.2	64.8	61.4	65.2	**65.9**		0.2	71.2	69.3	70.3	**73.1**
	0.5	**78.0**	75.7	64.5	75.1		0.5	83.1	85.5	67.8	**83.7**
	0.8	**88.0**	87.1	64.0	79.6		0.8	94.7	**95.4**	64.6	86.8
$n = 100$	−0.8	95.1	95.8	75.1	**96.4**	$T = 10$	−0.8	97.7	98.2	78.9	**98.6**
	−0.5	88.9	91.3	76.4	**92.1**		−0.5	**92.4**	91.6	79.3	91.7
	−0.2	74.2	75.1	76.4	**77.2**		−0.2	**81.5**	77.4	78.8	78.3
	0.2	75.6	74.9	75.1	**78.3**		0.2	81.4	80.3	77.3	**81.7**
	0.5	87.9	**89.1**	72.9	88.1		0.5	93.3	**93.5**	75.2	91.3
	0.8	94.3	**95.6**	69.4	90.7		0.8	99.1	**99.4**	70.0	93.9

Table 4: Average percentage of correct selections. DGP: SLM. Equation estimated: SLM.

Aggregated by cross-section, sample size (n)						Aggregated by time series, sample size (T)					
	ρ	$h(y)$	*Bayes*	MJ	AIC		ρ	$h(y)$	*Bayes*	MJ	AIC
$n = 25$	-0.8	58.7	**59.7**	27.3	58.3	$T = 1$	-0.8	54.0	53.1	23.7	**54.5**
	-0.5	**41.9**	36.2	27.0	38.4		-0.5	36.3	33.3	24.3	**37.7**
	-0.2	**28.3**	15.8	**28.3**	26.5		-0.2	22.6	14.2	**28.3**	22.5
	0.2	33.4	21.0	30.2	**33.5**		0.2	30.8	12.6	**32.5**	30.0
	0.5	54.0	49.6	31.8	**54.2**		0.5	**46.3**	37.4	34.6	45.4
	0.8	**72.4**	70.8	31.9	70.0		0.8	**61.0**	**61.0**	36.2	56.6
$n = 49$	-0.8	73.6	73.2	22.1	**74.4**	$T = 5$	-0.8	79.7	81.4	19.9	**80.7**
	-0.5	**53.2**	47.7	25.4	51.5		-0.5	**57.9**	53.0	24.1	55.9
	-0.2	**32.3**	17.9	28.9	30.5		-0.2	**32.8**	15.1	28.1	30.2
	0.2	**41.7**	24.9	31.5	39.9		0.2	**44.4**	27.4	29.2	43.0
	0.5	**68.8**	64.3	26.9	67.7		0.5	**73.0**	72.8	24.4	71.5
	0.8	86.8	**87.1**	26.2	82.1		0.8	**93.5**	93.3	24.8	88.3
$n = 100$	-0.8	86.7	87.0	12.0	**87.6**	$T = 10$	-0.8	**85.4**	85.3	17.7	85.1
	-0.5	68.0	65.2	18.0	**68.3**		-0.5	**68.8**	62.8	22.0	64.7
	-0.2	**37.8**	22.2	27.1	36.3		-0.2	**43.0**	26.6	28.0	40.6
	0.2	**51.3**	35.4	27.6	50.6		0.2	**51.1**	41.3	27.6	**51.1**
	0.5	**81.4**	79.8	20.6	79.0		0.5	**84.9**	83.4	20.2	84.0
	0.8	92.3	**92.9**	20.2	86.9		0.8	**97.0**	96.5	17.3	94.0

Table 5: Average percentage of correct selections. DGP: SDEM. Equation estimated: SDM.

Aggregated by cross-section, sample size (n)						Aggregated by time series, sample size (T)					
	ρ	$h(y)$	*Bayes*	MJ	AIC		ρ	$h(y)$	*Bayes*	MJ	AIC
$n = 25$	-0.8	**79.2**	77.2	51.4	77.6	$T = 1$	-0.8	66.2	66.0	38.1	**68.9**
	-0.5	**66.5**	64.1	55.1	62.3		-0.5	54.3	55.5	40.8	**58.1**
	-0.2	**57.9**	52.4	57.4	52.7		-0.2	42.2	40.2	42.3	**42.7**
	0.2	54.8	54.0	**59.1**	55.4		0.2	38.2	32.3	**44.8**	37.7
	0.5	**62.8**	62.4	55.9	59.4		0.5	**46.4**	42.3	45.4	43.6
	0.8	71.9	**72.2**	42.5	59.4		0.8	**55.1**	54.1	42.5	47.3
$n = 49$	-0.8	87.9	88.6	60.4	**89.9**	$T = 5$	-0.8	**92.3**	92.3	64.3	91.7
	-0.5	77.5	78.9	62.7	**79.8**		-0.5	**83.3**	81.6	67.4	80.0
	-0.2	64.2	61.3	**64.7**	63.1		-0.2	**69.4**	63.4	69.5	64.5
	0.2	64.8	60.7	**65.5**	64.1		0.2	**71.9**	68.5	69.4	70.9
	0.5	**75.1**	73.5	63.2	70.2		0.5	82.2	**83.9**	65.7	78.2
	0.8	**84.3**	**84.3**	47.1	72.1		0.8	92.2	**93.7**	44.3	77.9
$n = 100$	-0.8	94.9	95.0	67.0	**95.8**	$T = 10$	-0.8	**97.2**	**97.2**	72.6	96.9
	-0.5	87.9	88.6	70.4	**89.8**		-0.5	**89.9**	**89.9**	76.1	88.8
	-0.2	**72.4**	68.9	71.5	71.3		-0.2	**78.6**	74.9	77.7	75.6
	0.2	**74.9**	70.0	71.1	73.4		0.2	**80.8**	80.2	77.0	80.5
	0.5	**85.9**	**85.9**	68.2	83.3		0.5	91.8	**92.5**	72.0	87.9
	0.8	92.7	**93.3**	39.6	84.4		0.8	98.3	**98.8**	38.3	87.6

To complete the picture, we are going to estimate, for each DGP/Estimated-equation combination, a *response-surface* to model the empirical probability of choosing the correct weighting matrix using the corresponding criterion, p_i. As usual, a logit transformation of the empirical probabilities is carried out, so the estimated equation is:

$$\log\left(\frac{p_i + (2r)^{-1}}{1 - p_i + (2r)^{-1}}\right) = p_i^* = \eta + z_i\varphi + \epsilon_i, \tag{10}$$

where p_i^* is the logit transformation, often known as the *logit*, r the number of replications of each experiment (1000 in all the cases; $(2r)^{-1}$ assures that the *logit* is defined even when the probability of correct selection is 0 or 1; Maddala, 1983); η is an intercept term, z_i the design matrix whose columns reflect the conditions of each experiment, φ is a vector of parameters and ϵ_i the error term assumed to be independent and identically distributed (this assumption is reasonable if all experiments come from the same study, as ours, and are obtained under identical circumstances; Florax and De Graaff, 2004). Let us remind that the number of observations for each *response-surface* equation is 216 (so $i = 1, 2, ..., 216$). Table 6 shows the results for the three DGP/Estimated-equation combination combination.

Table 6: Estimated response surfaces.

SDEM case	constant	n	T	β_1	θ	$\|\rho\|$	R^2	F_{AV}
Entropy	-5.9410	0.0037	0.0566	0.0005	0.0748	0.5568	0.74	117.90
	(0.0000)	(0.0000)	(0.0000)	(0.9402)	(0.0000)	(0.0000)		(0.0000
Bayes	-6.2233	0.0051	0.0660	-0.0017	0.0904	0.6813	0.66	81.57
	(0.0000)	(0.0000)	(0.0000)	(0.8553)	(0.0000)	(0.0000)		(0.0000
MJ test	-6.1295	0.0044	0.0520	0.0106	0.1569	-0.0377	0.82	196.74
	(0.0000)	(0.0000)	(0.0000)	(0.0910)	(0.0000)	(0.4612)		(0.0000
AIC	-5.9177	0.0043	0.0506	0.0044	0.0795	0.4590	0.67	87.21
	(0.0000)	(0.0000)	(0.0000)	(0.5407)	(0.0000)	(0.0000)		(0.0000
SDM case	constant	n	T	β_1	θ	$\|\rho\|$	R^2	F_{AV}
Entropy	-5.8902	0.0033	0.0481	0.0053	0.0702	0.06348	0.66	83.35
	(0.0000)	(0.0000)	(0.0000)	(0.4614)	(0.0000)	(0.0000)		(0.0000
Bayes	-6.1117	0.0033	0.0548	0.0052	0.0861	0.8116	0.60	63.33
	(0.0000)	(0.0000)	(0.0000)	(0.5974)	(0.0000)	(0.0000)		(0.0000
MJ test	-5.8998	0.0024	0.0476	0.0186	0.1036	0.1668	0.47	36.74
	(0.0000)	(0.0004)	(0.0000)	(0.0813)	(0.0000)	(0.0552)		(0.0000
AIC	-5.9339	0.0034	0.0479	0.0092	0.0722	0.6301	0.67	83.61
	(0.0000)	(0.0000)	(0.0000)	(0.2051)	(0.0000)	(0.0000)		(0.0000
SLM case	constant	n	T	β_1		$\|\rho\|$	R^2	F_{AV}
Entropy	-6.3435	0.0049	0.0613	-0.0390		1.2505	0.81	113.60
	(0.0000)	(0.0000)	(0.0000)	(0.0001)		(0.0000)		(0.0000
Bayes	-7.0854	0.0054	0.0786	-0.0709		2.2207	0.83	122.53
	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)		(0.0000
MJtest	1.3129	-0.0131	-0.0896	-0.2215		-1.4089	0.40	16.92
	(0.0000)	(0.0000)	(0.0006)	(0.0000)		(0.0004)		(0.0000
AIC	-6.3808	0.0050	0.0599	-0.0396		1.2678	0.79	96.74
	(0.0000)	(0.0000)	(0.0000)	(0.0003)		(0.0000)		(0.0000
MIX case	constant	n	T	β_1	θ	$\|\rho\|$	R^2	F_{AV}
Entropy	-5.9736	0.0039	0.0583	-0.0004	0.0745	0.5505	0.72	109.13
	(0.0000)	(0.0000)	(0.0000)	(0.9511)	(0.0000)	(0.0000)		(0.0000
Bayes	-6.1882	0.0040	0.0648	-0.0001	0.0916	0.7103	0.67	85.13
	(0.0000)	(0.0000)	(0.0000)	(0.9887)	(0.0000)	(0.0000)		(0.0000
MJ test	-5.6677	0.0020	0.0379	0.0007	0.1162	-0.3854	0.55	50.92
	(0.0000)	(0.0000)	(0.0000)	(0.9431)	(0.0000)	(0.0000)		(0.0000
AIC	-5.9741	0.0043	0.0558	-1.9169	0.0696	0.4728	0.68	88.38
	(0.0000)	(0.0000)	(0.0000)	(0.9979)	(0.0000)	(0.0000)		(0.0000

Note: pvalue appear between brackets. F_{AV} means F test of the null that all coefficients are zero except the constant.

In general, the estimates confirm previous results. The main factor influencing the empirical probability of choosing the correct weights matrix is the spatial parameter, absolute value of ρ in Table 6. Its contribution is crucial in the case of the *Bayesian* criteria and, to a lesser extend, also in the cases of *Entroy* and AIC. This parameter is not significant, for the case of the MJ test, in $SDEM$ processes, hardly significant for SDM models and its contribution is significative, but negative, in the SLM and in misspecified equations. The second more influential factor is the parameter θ, associated to spatial spillovers. Its impact is beneficial for all the cases though it appears to be more important for the MJ test; the other three criteria are a bit less sensitive. Sample size is also relevant in all the cases and T has a relatively higher impact than n. Finally, as said before, parameter β_1 is not significant in any circumstance, but the SLM case, which means that the *signal-to-noise* ratio should not be a major factor to consider when choosing the best weighting matrix.

Table 7 completes the *response-surface* analysis with the F tests of equality in the coefficients of the *response-surface* estimates. According to the sequence of F tests, the most dissimilar method is the MJ approach, and then *Bayes*. On the contrary, *Entropy* and AIC are almost indistinguishable approaches for the four types of $DGPs$ according to this simple analysis.

Table 7: F test for the equality of coefficients in the response-surface estimates

SDEM case	*Bayes*	*MJ test*	*AIC*
Entropy	55.000 (0.00)	87.331 (0.00)	**1.535 (0.17)**
Bayes	–	34.720 (0.00)	4.558 (0.00)
MJ test	–	–	61.774 (0.00)
SDM case	*Bayes*	*MJ test*	*AIC*
Entropy	4.699 (0.00)	34.886 (0.00)	**0.471 (0.83)**
Bayes	–	14.791 (0.00)	3.300 (0.00)
MJ test	–	–	28.553 (0.00)
SLM case	*Bayes*	*MJ test*	*AIC*
Entropy	61.544 (0.00)	8685.34 (0.00)	**0.500 (0.78)**
Bayes	–	432.170 (0.00)	45.475 (0.00)
MJ test	–	–	7423.01 (0.00)
MIX case	*Bayes*	*MJ test*	*AIC*
Entropy	4.454 (0.00)	118.882 (0.00)	**2.056 (0.06)**
Bayes	–	65.420 (0.00)	5.171 (0.00)
MJ test	–	–	85.234 (0.00)

Note: p-value appear between brackets.

5 Empirical applications

The empirical applications in this section are based on two well-known economic models. The first one is a model of economic growth estimated by Ertur and Koch (2007) using a cross-section sample of 91 countries over the period 1960–1995. The other example is an economic model of productivity estimated by Munnell (1990) using panel data on 48 states in US over the period 1970–1979. The purpose of this section is to check the selection of spatial weight matrices in each case.

5.1 Study case 1: Ertur and Koch (2007)

Ertur and Koch (2007) build a growth equation to model technological interdependence between countries using spatial externalities. The main hypotheses of interaction is that the stock of knowledge in one country produces externalities that cross national borders and spill over into neighboring countries, with an intensity which decreases with distance. The authors use the criterion of pure geographical distance.

The benchmark model assumes an aggregate Cobb-Douglas production function with constant returns to scale in labour and physical capital:

$$Y_i(t) = A_i(t)K_i^{\alpha}(t)L_i^{1-\alpha}(t), \tag{11}$$

where $Y_i(t)$ is output, $K_i(t)$ is the level of reproducible physical capital, $L_i(t)$ is the level of labour, and $A_i(t)$ is the aggregate level of technology specified as:

$$A_i(t) = \Omega(t)k_i^{\phi}(t)\prod_{j\neq i}^{n} A_i^{\delta\omega_{ij}}(t). \tag{12}$$

The aggregate level of technology $A_i(t)$ in a country i depends on three elements. First, a certain proportion of technological progress is exogenous and identical in all countries: $\Omega(t) = \Omega(0)e^{\mu t}$, where μ is a constant rate of technological growth. Second, each country's aggregate level of technology increases with the aggregate level of physical capital per worker $k_i^{\phi}(t) = (K_i(t)/L_i(t))^{\phi}$ with the parameter $\phi \in [0;1]$ capturing the strength of home externalities by physical capital accumulation. Finally, the third term captures the external effects of knowledge embodied in capital located in a different country, whose impact crosses national borders at a diminishing intensity, $\delta \in [0;1]$. The terms ω_{ij} represent the connectivity between country i and its neighbours; this weight is assumed to be exogenous, non-negative and finite.

Following Solow, the authors assume that a constant fraction of output s_i, in every country i, is saved and that labour grows exogenously at the rate n_i. Also, they assume a constant and identical annual rate of depreciation of physical capital for all countries, denoted τ. The evolution of output per worker in country i is governed by the usual fundamental dynamics of the Solow equation which, after some manipulations, lead to the steady-state real income per worker in that country (Ertur and Koch, 2007, p. 1038, eq. 9):

$$y = \Omega + (\alpha + \phi)\, k - \alpha\delta\mathbf{W}k + \delta\mathbf{W}y. \tag{13}$$

This is a spatially augmented Solow model and coincides with the predictor obtained by Solow adding spillover effects. In terms of spatial econometrics, we identify a *Spatial Durbin Model*, *SDM*, in the equation which can be expressed as:

$$y = x\beta + \rho\mathbf{W}y + \mathbf{W}x\theta + \varepsilon. \tag{14}$$

Equation (14) is estimated using information on real income, investment and population growth for a sample of 91 countries over the period $1960-1995$. Regarding the spatial weighting matrix, Ertur and Koch consider two geographical distance functions: the inverse of squared distance (which is the main hypothesis) and the negative exponential of squared distance (to check robustness in the specification). We also consider a third matrix using the inverse of the distance.

Let us call the three weighting matrices as $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ which are row-standardized: $\omega_{hij} = \omega^*_{hij} / \sum_{j=1}^{n} \omega^*_{hij};\ h = 1, 2, 3$ where:

$$\omega^*_{1ij} = \begin{cases} 0 & if\ i = j \\ d_{ij}^{-2} & otherwise \end{cases} ; \quad \omega^*_{2ij} = \begin{cases} 0 & if\ i = j \\ e^{-2d_{ij}} & otherwise \end{cases} ; \quad \omega^*_{3ij} = \begin{cases} 0 & if\ i = j \\ d_{ij}^{-1} & otherwise \end{cases} , \qquad (15)$$

with d_{ij} as the great-distance between the capitals of countries i and j.

The authors analyze several specifications checking for different theoretical restrictions and alternatives spatial equations. We concentrate our revision in the non-restricted equation of Ertur and Koch (in the sense that it includes more coefficients than advised by theory). Table 8 presents the SDM version of this equation using the three alternative weighting matrices specified before (the first two columns coincides with those in Table I, columns 3-4, pp. 1047, of Ertur and Koch, 2007). The the last four rows in the Table shows the value of the selection criteria corresponding to each case.

Table 8: Ertur & Koch case. Unrestricted SDM estimates

Model/Weight matrix	***SDM / W1***	***SDM / W2***	***SDM / W3***
constant	1.178 (0.62)	0.678 (0.36)	5.045 (0.96)
$\log(s)$	0.829 (8.24)	0.795 (7.60)	0.908 (8.50)
$\log(n + 0.05)$	−1.500 (−2.62)	−1.452 (−2.61)	−1.711 (−2.68)
$\mathbf{W} \times \log(s)$	−0.283 (−1.51)	−0.345 (−2.06)	0.468 (1.19)
$\mathbf{W} \times \log(n + 0.05)$	0.528 (0.62)	0.118 (0.15)	2.177 (1.02)
$\mathbf{W} \times \log(y)$	0.716 (9.61)	0.643 (8.43)	0.899 (13.67)
Selection Criteria			
Entropy	28.021	29.631	34.616
Bayesian	0.871	0.127	0.002
MJ test	11.158	9.388	10.208
AIC	95.885	99.100	109.132

Note: t-ratios appear between brackets.

The preferred model by Ertur and Koch corresponds to $SDM/\mathbf{W}_1$ which coincides with the selection corresponding to the criterion of minimum *Entropy,* the *Bayesian* posterior probability and AIC. The selection of the minimum J test is $\mathbf{W}_2$.

Other results in Ertur and Koch concern to the Spatial Error Model version of the steady-state equation of (13), or SEM model. The intention of the authors is to visualize the presence of spatial correlation in the traditional non spatial Solow equations; we use this case as an example of selection of weighting matrices in misspecified models. The main results appear in Table 9 (which can be compared with columns 2-3 of Table II, in Ertur and Koch, 2007, p. 1048).

Table 9: Ertur & Koch case. Unrestricted SEM estimates

Model/Weight matrix	***SEM / W1***	***SEM / W2***	***SEM / W3***
constant	6.457 (4.22)	6.706 (4.62)	5.892 (3.02)
$\log(s_i)$	0.828 (8.36)	0.804 (7.87)	0.992 (8.94)
$\log(n_i + 0.05)$	-1.702 (-3.03)	-1.552 (-2.85)	-2.269 (-3.65)
$\mathbf{W} \times \varepsilon_i$	0.823 (15.69)	0.737 (12.19)	0.937 (22.08)
Selection Criteria			
Entropy	30.973	31.734	42.049
Bayesian	0.655	0.345	0.000
MJ test	0.171e^{-12}	0.043e^{-12}	0.085e^{-12}
AIC	97.870	99.391	120.021

Note: t-ratios appear between brackets.

The selection of most adequate $\mathbf{W}$ matrix does not change. Using the values of *Entropy* criterion we select the model in which intervenes the matrix $\mathbf{W}_1$, the same as with the Bayesian approach and the AIC criterion; MJ continues selecting $\mathbf{W}_2$.

5.2 Study case 2: Munnell (1990)

Munnell et al. (1990) suggests a Cobb-Douglas production function in each state of the US (excluding "islands" Alaska and Hawaii and the district of Columbia, for a total of 48 states) observed for the period 1970 and 1986. The dependent variable, output of the production function, is the gross state product, $\log(gsp)$, and the explanatory variables considered are the endowment of public capital, $\log(pcap)$ (roads, water facilities and other utilities), the private capital, $\log(pc)$, employment, $\log(emp)$, and the unemployment rate, *unemp*, in order to proxy for the effects of the business cycle. The model can be expressed formally as follows:

$$\log(gsp_{it}) = \alpha + \beta_1 \log(pcap_{it}) + \beta_2 \log(pc_{it}) + \beta_3 \log(emp_{it}) + \beta_4 unemp_{it} + \mu_i + \varepsilon_{it}, \tag{16}$$

with $i = 1, \ldots, 48$, and $t = 1970, ..., 1979$. This dataset had been used by Millo et al. (2012) and Álvarez et al. (2017) among others. For this example, we consider subset of the original data including only the observations between 1970 and 1979. This selection is to keep the number of regions and time periods into the panel dimensions used in the Monte Carlo simulation.

We estimate a fixed effect model in the context of $SDEM$ and SDM equations. Following previous studies, the hypothesis of spatial spill-over is based on contiguity criterion ($\mathbf{W}_1$ in our case). This matrix presents an average of connectivity of 4.5 neighbours and the median of neighbours is 4. Using this information, we propose two additional spatial weighting matrices using k-nearest neighbours criterion: $\mathbf{W}_2$ is constructed using 4 nearest neighbors and $\mathbf{W}_3$ with 5 nearest neighbors.

The three weighting matrices are row-standardized: $\omega_{hij} = \omega^*_{hij} / \sum_{j=1}^{n} \omega^*_{hij}$; $h = 1, 2, 3$ where:

$$\omega^*_{1ij} = \begin{cases} 1 & if\ (i,j)\ have\, common\, frontier \\ 0 & otherwise \end{cases};$$

$$\omega^*_{2ij} = \begin{cases} 1 & if\ d_{ij} \leq d_{i(k=4)} \\ 0 & otherwise \end{cases}; \tag{17}$$

$$\omega^*_{3ij} = \begin{cases} 1 & if\ d_{ij} \leq d_{i(k=5)} \\ 0 & otherwise \end{cases},$$

with d_{ij} as the Euclidean-distance between the centroids of counties i and j, and $d_{i(k)}$ indicates the distance of $k-th$ neighbor.

Previously to applied the set of criteria, we check the significance of spatial effects using LM tests obtaining as final specification a $SDEM$ model with two spatial lagged explanatory variables, $\mathbf{W} \times \log(pc)$ and $\mathbf{W} \times \log(emp)$. This is the preferred model and, also, we present the results of SDM model as a misspecified model. All results are presented in the Tables 10 and 11.

Table 10: Munnell case. SDEM estimates

Model/Weight matrix	***SDEM / W1***	***SDEM / W2***	***SDEM / W3***
$\log(pcap)$	0.103 (3.07)	0.111 (3.32)	0.106 (3.15)
$\log(pc)$	0.289 (8.50)	0.289 (8.31)	0.288 (8.30)
$\log(emp)$	0.629 (18.02)	0.637 (17.76)	0.606 (17.37)
$unemp$	−0.003 (−2.40)	−0.003 (−2.34)	−0.002 (−1.79)
$\mathbf{W} \times \log(pc)$	−0.201 (−3.81)	−0.232 (−4.71)	−0.250 (−4.43)
$\mathbf{W} \times \log(emp)$	0.191 (4.13)	0.224 (4.66)	0.283 (5.14)
$\mathbf{W} \times \varepsilon$	0.420 (8.28)	0.357 (6.56)	0.432 (7.81)
Selection Criteria			
Entropy	−1.131	−1.117	−1.128
Bayesian	0.763	0.000	0.237
MJ test	15.890	5.958	0.292
AIC	−2.134	−2.115	−2.132

Note: t-ratios appear between brackets.

The criterion of minimum Entropy, the Bayesian posterior probability and AIC are coincident to select as preferred model the pair $SDEM/\mathbf{W}1$. The selection of the minimum J test is $SDEM/\mathbf{W}3$.

For the misspecified model, the SDM, the selection for each criterion of the weighting matrix is similar to the previous one model. In Table 11, Entropy, the Bayesian posterior probability and AIC select the pair $SDM/\mathbf{W}1$ and minimum J test chooses $SDM/\mathbf{W}3$.

Table 11: Munnell case. SDM estimates

Model/Weight matrix	***SDM / W1***	***SDM / W2***	***SDM / W3***
$\log(pcap)$	0.108 (3.34)	0.126 (3.92)	0.114 (3.58)
$\log(pc)$	0.311 (8.60)	0.306 (8.42)	0.306 (8.45)
$\log(emp)$	0.589 (16.04)	0.614 (16.45)	0.578 (15.85)
$unemp$	−0.003 (−2.72)	−0.003 (−2.71)	−0.002 (−2.21)
$\mathbf{W} \times \log(pc)$	−0.293 (−5.94)	−0.299 (−6.35)	−0.318 (−6.37)
$\mathbf{W} \times \log(emp)$	−0.118 (−2.05)	−0.067 (−1.09)	−0.070 (−1.07)
$\mathbf{W} \times \log(y)$	0.430 (8.77)	0.359 (6.80)	0.430 (7.98)
Selection Criteria			
Entropy	−1.137	−1.121	−1.131
Bayesian	0.961	0.000	0.039
MJ test	8.693	6.246	1.122
AIC	−2.144	−2.121	−2.139

Note: t-ratios appear between brackets.

6 Conclusion

Much of the applied spatial econometrics literature sets an exogenous approximation to the **W** matrix. Implicitly, it is assumed that the user has valuable knowledge with respect to the way that individuals in the sample interact among them, which allows us to build a weighting matrix. This matrix is habitually considered true and it is rarely questioned. In recent years, new literature advocates for a more data driven approach to the **W** issue. We strongly support this tendency, which should be dominant in the future; however, our focus in this paper is on the exogenous approach.

The problem in applied work is rather frequent: the user has a finite collection of weighting matrices, they all are coherent with the case of study, and she/he needs to select one of them. Which is the best **W**? We can answer this question using different proposals currently present in the literature: the *Bayesian* posterior probability, the J tests in its variants or simple model selection criteria, such as AIC or BIC, very common in mainstream econometrics (sure, there are other useful alternatives). We add a fourth one, based on the *Entropy* of the estimated distribution function. Our criterion *h(y)* is a measure of uncertainty which fits pretty well in the **W** decision problem. The *h(y)* statistics depends on the estimated covariance matrix of the corresponding model offering a more complete picture of the suitability of the distribution function (each defined by a particular **W**), to deal with the data at hand.

The conclusions of our Monte Carlo are very enlightening. First, we can assert that it is possible to identify, with certain confidence, the true weighting matrix (if it really exists); in this sense, the selection criteria do a good job. However, the four criteria should not be taken as indifferent, especially in samples of small size (n or T). The ordering is clear: *Entropy* in first place, AIC and *Bayesian* posterior probability slightly worse, and then J in the fourth position. As shown in the paper, the value of the spatial parameter has the greatest impact to guarantee a correct selection, but this information is unobservable to the researcher. However, the user effectively controls the amount of information involved in the exercise, and this is also a key factor. The advice is clear: use as much information as you have because the quality of the decision improves with the information. Once again, the way the information accrues is not neutral: the length of the time series in the panel is more relevant than the number of cross-sectional units in the sample.

Our final recommendation for applied researchers is to check for the adequacy of the weighting matrix and, in case of having various candidates, take a decision using well-defined criteria such as the *Entropy*. The case studies presented in Section 5 illustrates this procedure.

References

Akaike, H. (1973). *2nd International Symposium on Information Theory*, Chapter Information Theory and an Extension of the Maximum Likelihood Principle, pp. 267–281. Akademiai Kiodo.

Álvarez, I. C., J. Barbero, and J. L. Zofío (2017). A panel data toolbox for matlab. *Journal of Statistical Software 76*(1), 1–27.

Anselin, L. (1984). Specification tests on the structure of interaction in spatial econometric models. *Papers of the Regional Science Association 54*, 165–182.

Anselin, L. (1988). *Spatial econometrics: Methods and models*, Volume 4. Kluwer Academic Publishers, Dordrecht.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics 27*(3), 247–267.

Bavaud, F. (1998). Models for spatial weights: a systematic look. *Geographical analysis 30*(2), 153–171.

Beenstock, M. and D. Felsenstein (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis 44*(4), 386–397.

Benjanuvatra, S. and P. Burridge (2015). Qml estimation of the spatial weight matrix in the mr-sar model. *Working Paper 15/24. Department of Economics and Related Studies, University of York.*

Besag, J. and D. Higdon (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*(4), 691–746.

Bhattacharjee, A. and C. Jensen-Butler (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics 43*(4), 617–634.

Burridge, P. (2012). Improving the j test in the sarar model by likelihood-based estimation. *Spatial Economic Analysis 7*(1), 75–107.

Burridge, P. and B. Fingleton (2010). Bootstrap inference in spatial econometrics: the j-test. *Spatial Economic Analysis 5*(1), 93–119.

Debarsy, N. and C. Ertur (2016). Interaction matrix selection in spatial econometrics with an application to growth theory. *Document de Recherche du Laboratoire d'Economie d'Orleans*.

Ertur, C. and W. Koch (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of applied econometrics 22*(6), 1033–1062.

Florax, R. J. and T. De Graaff (2004). The performance of diagnostic tests for spatial dependence in linear regression models: a meta-analysis of simulation studies. In L. Anselin, R. Florax, and S. Rey (Eds.), *Advances in spatial econometrics*, pp. 29–65. Springer.

Florax, R. J. and S. Rey (1995). The impacts of misspecified spatial interaction in linear regression models. In L. Anselin and R. Florax (Eds.), *New directions in spatial econometrics*, pp. 111–135. Springer.

Franzese Jr, R. J. and J. C. Hays (2007). Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis 15*, 140–164.

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

Gibbons, S. and H. G. Overman (2012). Mostly pointless spatial econometrics? *Journal of Regional Science 52*(2), 172–191.

Hagemann, A. (2012). A simple test for regression specification with non-nested alternatives. *Journal of Econometrics 166*(2), 247–254.

Haining, R. P. (2003). *Spatial data analysis: theory and practice.* Cambridge University Press.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory 21*(01), 60–68.

Hays, J. C., A. Kachi, and R. J. Franzese (2010). A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology 7*(3), 406–428.

Hepple, L. W. (1995a). Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds. *Environment and Planning A 27*(3), 447–469.

Hepple, L. W. (1995b). Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms. *Environment and Planning A 27*(4), 615–644.

Hepple, L. W. (2004). Bayesian model choice in spatial econometrics. In J. LeSage and K. Pace (Eds.), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, Volume 18, pp. 101–126. Emerald Group Publishing Limited.

Kelejian, H. H. (2008). A spatial j-test for model specification against a single or a set of non-nested alternatives. *Letters in Spatial and Resource Sciences 1*(1), 3–11.

Kelejian, H. H. and G. Piras (2011). An extension of kelejian's j-test for non-nested spatial models. *Regional Science and Urban Economics 41*(3), 281–292.

Kelejian, H. H. and G. Piras (2016). An extension of the j-test to a spatial panel data framework. *Journal of Applied Econometrics 31*(2), 387–402.

Kooijman, S. (1976). Some remarks on the statistical analysis of grids especially with respect to ecology. *Annals of Systems Research 5*, 113–132.

Lee, L.-f. and J. Yu (2012). Qml estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spatial Economic Analysis 7*(1), 31–74.

Leenders, R. T. A. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social networks 24*(1), 21–47.

LeSage, J. (2007). Spatial econometrics toolbox. *available at: www. spatial-econometrics. com*.

LeSage, J. and R. Pace (2009). *Introduction to spatial econometrics*. Statistics: A Series of Textbooks and Monographs. Chapman and Hall, CRC press.

LeSage, J. P. and R. K. Pace (2014). The biggest myth in spatial econometrics. *Econometrics 2*(4), 217–249.

Maddala, G. (1983). Qualitative and limited dependent variable models in econometrics.

Millo, G., G. Piras, et al. (2012). splm: Spatial panel data models in r. *Journal of Statistical Software 47*(1), 1–38.

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological) 10*(2), 243–251.

Munnell, A. H., L. M. Cook, et al. (1990). How does public infrastructure affect regional economic performance? *New England economic review* (Sep), 11–33.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association 70*(349), 120–126.

Oud, J. H. and H. Folmer (2008). A structural equation approach to models with spatial dependence. *Geographical Analysis 40*(2), 152–166.

Piras, G. and N. Lozano (2012). Spatial j-test: some monte carlo evidence. *Statistics and Computing 22*, 169–183.

Pötscher, B. (1991). Effects of model selection on inference. *Econometric Theory 7*, 163–185.

Qu, X. and L. f. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*, 209–232.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Shannon, C. and C. Weaver (1949). *The Mathematical Theory of Communication.* University of Illinois Press: Urbana.

Snijders, T. A., C. E. Steglich, and M. Schweinberger (2007). *Longitudinal models in the Behavioural ond Related Sciences*, Chapter Modeling the co-evolution of networks and behavior, pp. 41–71. Mahwah: Lawrence Erlbaum.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*, 434–449.