ORIGINAL PAPER

Evaluating information criteria for selecting spatial processes

Christos Agiakloglou[1] · **Apostolos Tsimpanos**[2]

Abstract

Information criteria have been widely used in many quantitative applications as an effort to select the most appropriate model that describes well enough the unknown population behavior for a given dataset. Studies have shown that their performance depends on several elements and the selection of the best fitted model is not always the same for all criteria. For this purpose, this research evaluates the performance of the three most often used information criteria, such as the Akaike information criterion, the Bayesian information criterion and Hannan and Quinn information criterion, for selecting spatial processes, taking into account that the sample in spatial analysis is regarded as a realization of a spatial process that incorporates the spatial dependence between the observations. Using a Monte Carlo analysis for the three most frequently applied in practice spatial processes, such as the first-order spatial autoregressive process, SAR(1), the first-order spatial moving average process, SMA(1), and the mixed spatial autoregressive moving average process, SARMA(1, 1), this study finds that these information criteria can assist the analyst to select the true process, but their behavior depends on sample size as well as on the magnitude of the spatial parameters, leading occasionally to alternative competitive processes.

JEL Classification C20 · C21 · C52 · C53

1 Introduction

An important aspect of analyzing data collected from different geographical regions, known as spatial data, is the emergence of spatial autocorrelation, a situation where the values of a variable are correlated according to their geographical positions,

---

✉ Christos Agiakloglou
agiaklis@unipi.gr

1   Department of Economics, University of Piraeus, Karaoli & Dimitriou 80, 18534 Piraeus, Greece

2   Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Samos, Greece

creating clusters of observations, and, as Anselin (1988) reports, spatial autocorrelation is attributed to spatial dependence, which along with spatial heterogeneity, is outcome of spatial effects inherent in this type of data. The presence of spatially nonindependent observations causes serious problems in quantitative analysis, since the sample contains less information than a counterpart with independent spatial elements, and, moreover, the concept of the random sample is violated, as Schabenberger and Gotway (2005) have indicated. Therefore, any conventional statistical inference will produce unreliable results, unless spatial dependence is incorporated to the model, as it happens in time-series data with time autocorrelation. For this reason, a spatial sample should be considered as a realization of a spatial process and not as a random sample, as in time-series analysis.

The traditional Box and Jenkins (1976) methodology for time-series analysis has been extended to spatial analysis as an effort to model spatial dependence between observations of the same variable. Likewise, spatial correlograms and spatial partial correlograms are constructed, using the Moran's I spatial autocorrelation coefficient or some other measures that have been proposed as presented by Cliff and Ord (1981), to identify the most adequate spatial generation mechanism of an observed dataset. However, diagrams may often be unable to identify correctly the underlying mechanism producing in that sense confusing results, as in time-series analysis, where this issue of selecting the best fitted model is addressed by several information criteria.

Hence, it will be very interesting to examine and evaluate the performance of the three most often used in practice information criteria in spatial analysis, such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan and Quinn information criterion (HQC), in terms of selecting pure spatial processes. Indeed, the behavior of these criteria has been investigated thoroughly in the literature for time-series processes and regression models but not for spatial models, if you exclude the studies of Hoeting et al. (2006) and Lee and Ghosh (2008) in which they considered geostatistical models, i.e., models used for point-referenced geostatistical data that incorporate the dependence applied to a covariance function that determines the relationship for observations at different distance locations. Using a Monte Carlo analysis, this study finds that these information criteria can successfully contribute to spatial modeling, although their overall behavior depends not only on the sample size but also on the magnitude of the spatial parameters of the true generating processes.

The remaining of the paper is organized as follows. Section 2 depicts the most important spatial processes that will be considered for the simulation analysis and presents the three aforementioned information criteria. Section 3 describes the design of the simulation analysis and discusses the results. Finally, the concluding remarks are presented in Sect. 4.

2 Spatial processes and information criteria

Spatial processes can be regarded as multidirectional extensions of the well-known time-series processes on the geographical space, meaning that the dependence among values of a variable is expressed according to their geographical positions and not

according to their chronological order. For example, for a sample of n cross-sectional observations collected from n different geographical units, the spatial dependence is incorporated into the process by the definition of a spatial (n × n) weights matrix $\mathbf{W}$ that captures the spatial interaction between the neighboring locations. The matrix $\mathbf{W}$ is usually used in its row-standardized form such that its *Wij* element is different from zero if the locations *i* and *j* are neighbors, otherwise it is zero, where the determination of the neighbors for each spatial unit is clearly the most important issue for constructing such a matrix in spatial analysis. Indeed, a variety of criteria have been proposed in the literature for spatial weights formation including boundary contiguity and distance measures, as, for example, can be seen in Cliff and Ord (1981) and in Anselin (1988). Contiguity criteria consider as neighbors the spatial units which share common borders, so they are contiguous, while other criteria which are based on distance measures define the neighborhood according to the distance between two regions. A brief presentation of the three most commonly used spatial processes that express spatial dependence, namely the spatial autoregressive process of order 1, the spatial moving average process of order 1 and the mixed spatial autoregressive moving average process of orders 1 and 1, is given below.

The spatial autoregressive process of order 1, i.e., SAR(1), was initially introduced by Whittle (1954) and by Besag (1974) as an extension of the autoregressive process of order 1, i.e., AR(1), in time-series analysis to geographical context. Utilizing matrix notation and considering zero mean value for the examined variable, the SAR(1) process, as presented by LeSage and Pace (2009), is defined as:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is an $(n \times 1)$ vector of observations of the process collected from n geographical points, $\mathbf{W}$ is the (n × n) spatial weights matrix, ρ is the spatial autoregressive parameter and $\boldsymbol{\varepsilon}$ is an $(n \times 1)$ white noise random vector. The vector $\mathbf{W}\mathbf{y}$ is called spatial lag, and each element, for a row-standardized $\mathbf{W}$, is a weighting average of $\mathbf{y}$ values in neighboring units for every region.

The log-likelihood function of a SAR(1) process, assuming that $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ with $\mathbf{I}$ being the identity matrix and σ^2 a constant variance, is obtained as:

$$\ln L\left(\rho, \sigma^2\right) = -\frac{n}{2} \ln (2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{(\mathbf{y} - \rho \mathbf{W} \mathbf{y})'(\mathbf{y} - \rho \mathbf{W} \mathbf{y})}{2\sigma^2} + \ln |\mathbf{I} - \rho \mathbf{W}|$$

where $|\mathbf{I} - \rho \mathbf{W}|$ is the Jacobian determinant for the transformation of the random vector $\boldsymbol{\varepsilon}$ into the vector $\mathbf{y}$. Substituting the maximum likelihood estimator for the variance of the process, i.e., $\sigma^2_{ML} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/n$, the log-likelihood function becomes:

$$\ln L(\rho) = -(n/2)\ln (2\pi) - (n/2) - (n/2)\ln \left((1/n)(\mathbf{y} - \rho \mathbf{W} \mathbf{y})'(\mathbf{y} - \rho \mathbf{W} \mathbf{y})\right) + \ln |\mathbf{I} - \rho \mathbf{W}|$$

which is clearly only a function of the parameter ρ.

Next, the spatial moving average process of order 1, i.e., SMA(1), is defined as:

$$\mathbf{y} = \boldsymbol{\varepsilon} - \lambda \mathbf{W} \boldsymbol{\varepsilon}$$

where again $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \sigma^2 \mathbf{I})$ and λ is the spatial moving average coefficient, as presented by Haining (1978). The log-likelihood function of a SMA(1) process is:

$$\ln L(\lambda, \sigma^2) = -(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 - (1/2\sigma^2)\left[(\mathrm{I} - \lambda\mathbf{W})^{-1}\mathbf{y}\right]'(\mathrm{I} - \lambda\mathbf{W})^{-1}\mathbf{y} - \ln|\mathrm{I} - \lambda\mathbf{W}|$$

which can be written as:

$$\ln L(\lambda) = -(n/2)\ln(2\pi) - (n/2) - (n/2)\ln\left((1/n)\left[(\mathrm{I} - \lambda\mathbf{W})^{-1}\mathbf{y}\right]'(\mathrm{I} - \lambda\mathbf{W})^{-1}\mathbf{y}\right) - \ln|\mathrm{I} - \lambda\mathbf{W}|$$

for $\sigma^2_{ML} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/n$ and $|\mathrm{I} - \lambda\mathbf{W}|$ being the Jacobian determinant. Note, that as in the case of SAR(1), the log likelihood of SMA(1) is a function solely of the parameter λ.

Lastly, as in time-series analysis mixed models are also defined in spatial analysis, known as spatial autoregressive moving average processes, i.e., SARMA, as introduced by Huang (1984). The simplest mixed spatial autoregressive moving average process is the SARMA(1, 1) process defined as:

$$\mathbf{y} = \rho\mathbf{W}_1\mathbf{y} + \boldsymbol{\varepsilon} - \lambda\mathbf{W}_2\boldsymbol{\varepsilon}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ denote (n × n) spatial weights matrices for the autoregressive and the moving average term, respectively. It should be pointed out that the SARMA(1, 1) process is properly defined if and only if different weight matrices are used for the two components of the process, as can be seen in Mur and Angulo (2007). Unlike time-series analysis where for a proper definition of an ARMA(1, 1) process, the autoregressive parameter must not be equal to the moving average parameter, the SARMA(1, 1) process is well defined even for equal values of ρ and λ, provided that the weights matrices are different, i.e., $\mathbf{W}_1 \neq \mathbf{W}_2$. Actually, a SARMA(1, 1) process is a process that combines global and local effects, since the SAR(1) process expresses the global spatial dependence, i.e., influences from one geographical point that spread and affect the whole study region, whereas the SMA(1) process defines local spatial dependence with effects covering only the neighborhood regions.

The log-likelihood function of a SARMA (1, 1) process is obtained as:

$$\ln L(\rho, \lambda, \sigma^2) = -(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 - (1/2\sigma^2)\left[\mathbf{B}^{-1}\mathbf{A}\mathbf{y}\right]'\mathbf{B}^{-1}\mathbf{A}\mathbf{y} + \ln|\mathbf{A}| - \ln|\mathbf{B}|$$

where $\mathbf{A} = \mathrm{I} - \rho\mathbf{W}_1$ and $\mathbf{B} = \mathrm{I} - \lambda\mathbf{W}_2$, and takes the following form:

$$\ln L(\rho, \lambda) = -(n/2)\ln(2\pi) - (n/2) - (n/2)\ln\left((1/n)\left[\mathbf{B}^{-1}\mathbf{A}\mathbf{y}\right]'\mathbf{B}^{-1}\mathbf{A}\mathbf{y}\right) + \ln|\mathbf{A}| - \ln|\mathbf{B}|$$

which is a function only of the spatial parameters.

The log-likelihood functions of the three presented spatial processes are maximized by applying a numerical method for an observed dataset in order to estimate the spatial coefficients ρ and λ, emphasizing the fact that the most important element of this maximization process is the log-determinant of the Jacobian matrix. For this purpose, Ord (1975) has proposed a convenient method for dealing with this issue by using the eigenvalues of $\mathbf{W}$. For example, the log-determinant term for the spatial parameter ρ can be decomposed as:

$$\ln |I - \rho\mathbf{W}| = \sum_{j=1}^{n} \ln\left(1 - \rho\,\omega_j\right)$$

where ω_j are the eigenvalues of the weights matrix $\mathbf{W}$, provided that these values are real numbers. If, on the other hand, the eigenvalues are complex, Bivand et al. (2013) have suggested an alternative method of computing the log-determinant of the Jacobian matrix as follows:

$$\ln |I - \rho\mathbf{W}| = L_1 + L_2$$

where

$$L_1 = \sum_{j=1}^{k} \ln\left[\left(1 - \rho\,\omega_j\right)\left(1 - \rho\,\overline{\omega}_j\right)\right] = \sum_{j=1}^{k} \ln\left[\left(1 - \rho\alpha_j\right)^2 + \left(\rho b_j\right)^2\right]$$

and

$$L_2 = \sum_{j=k+1}^{n-k} \ln\left(1 - \rho\,\zeta_j\right)$$

while $\omega_j = \alpha_j + ib_j$ and $\overline{\omega}_j = \alpha_j - ib_j$ denoting the $2\,k$ complex eigenvalues, i is the imaginary unit and ζ_j represents a real eigenvalue.

However, it should be mentioned that in spatial analysis, unlike time-series analysis, the values of the coefficients ρ and λ are not necessarily restricted strictly to the interval $(-1, +1)$, but the estimation process can be implemented provided that the Jacobian matrix is nonsingular, an outcome that is related to the eigenvalues of the spatial weights matrices. Row-standardized spatial weights matrices have always the largest eigenvalue equal to unity, something which ensures that the upper limit of the interval will be always $+1$, while the value for the lower limit is unknown and several times smaller than -1. Ord (1975) has demonstrated that the spatial parameters for symmetric matrices before standardization could take values within the interval $\left(1/\omega_{\min}, 1/\omega_{\max}\right)$, where $\omega_{\min}$ and $\omega_{\max}$ are the smallest and largest real eigenvalues of $\mathbf{W}$. On the other hand, in the case of asymmetric row-standardized weights matrices with complex eigenvalues, LeSage and Pace (2009) suggested that the Jacobian is nonsingular when the spatial parameters have values in the interval $\left(1/r_s, 1\right)$, where r_s is the most negative purely real eigenvalue of $\mathbf{W}$. Lastly, if the parameters take values inside the feasible interval, corresponding to the applied weights matrix, the Jacobian determinant will be positive, meaning that its logarithm will exist, and the likelihood function of a process will be well defined.

Hence, if the log-likelihood functions are maximized, the best fitted model is selected according to the minimum value of any of the widely used information criteria. The first criterion most often appeared in practice is the Akaike information criterion (AIC) suggested by Akaike (1973) which was developed on the Kullback–Leibler divergence measure for evaluating the discrepancy between a true model and a candidate model. The AIC is computed as follows:

$$\text{AIC} = -2 \ln \hat{L} + 2p$$

where $\ln \hat{L}$ is the maximized value of the log-likelihood function and p is the number of parameters of the process. The other two well often applied criteria are the Bayesian information criterion (BIC), suggested by Schwarz (1978) as an attempt to improve AIC performance, and the Hannan and Quinn information criterion (HQC), suggested by Hannan and Quinn (1978), defined, respectively, as follows:

$$\text{BIC} = -2 \ln \hat{L} + p \ln n$$

and

$$\text{HQC} = -2 \ln \hat{L} + 2p \ln \ln n$$

where n is the sample size used for the estimation. Obviously, the number of existing information criteria is not limited to those three previously presented, but these criteria are typically the most often used in practice not only because it is fairly easy to compute their values but also because they are reported by almost every statistical package. However, since it is known in the literature that the AIC criterion is strongly negatively biased in small samples, as shown by Sugiura (1978) and Hurvich and Tsai (1989), bias-corrected version of AIC proposed by Hurvich and Tsai (1989) and denoted as AICc, i.e.,

$$\text{AICc} = \text{AIC} + \frac{2p(p + 1)}{n - p - 1}$$

is also considered.

The philosophy of an information criterion is based on the quantification of the goodness of fit of an estimated model including a penalty for the number of estimated parameters and for the sample size. The selection process of a model consists of estimating several different models, for which the value of the designated criterion is computed, and the best fitted model is the one at which the value of the criterion is minimized, emphasizing that alternative information criteria do not always select the same best model. Actually, AIC is known as an asymptotically efficient criterion, meaning that, if the true model is not among the candidate models, the criterion chooses the model with the minimum one-step expected quadratic forecasting error as the sample increases, i.e., AIC chooses the model which is the best approximation of the unknown and, perhaps, of an infinite-dimensional model. On the other hand, BIC and HQC are considered as consistent criteria in the sense that if the true model is among the candidate models, these criteria select the true model with probability approaching to 1 as sample increases (see details regarding their properties in Diebold 2007, Judge et al. 1985 and Burnham and Anderson 2002).

Overall, AIC has the tendency to select a more elaborate model, as Jones (1975) and Shibata (1976) have demonstrated that for autoregressive time-series processes, AIC overestimates the true order of the process, while BIC chooses simpler models according to the parsimonious concept. In this lieu, Hurvich and Tsai (1989) showed that AIC is a biased estimator of the Kullback–Leibler information causing in that

sense overfitting of the model. Therefore, it will be very exciting to evaluate the performance of these information criteria for spatial analysis, i.e., for observation in a lattice environment appropriate for economic and regional data, knowing also that their behavior has been examined for time-series and regression models.

3 Simulation results

The performance of the three previously presented information criteria, i.e., AIC, BIC and HQC, is investigated in this section for spatial data using a Monte Carlo analysis, in terms of selecting the right spatial process among the three candidate processes, i.e., SAR(1), SMA(1) and SARMA(1, 1). Spatial dependence is introduced into the processes by defining two separate row-standardized spatial weights matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ for the autoregressive and moving average terms, respectively. More precisely, matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are constructed using the rook (four neighbors-common edge) and the queen (eight neighbors-common edge and vertex) contiguity definitions, respectively, over a squared regular lattice for dimensions 10×10 and 20×20 providing samples of 100 and 400 observations. Using this formation, the matrices are symmetric originally, but they become asymmetric after row standardization, although it must be said that their structure behaves relatively like a symmetric matrix. Moreover, the research is extended into a situation with extremely asymmetric weights matrices in order to further investigate the performance of these information criteria under a realistic environment. For this purpose, the geographical structure of Greece at the local authority districts of Kallikrates Operational Programme consisting of 325 municipalities is considered to construct matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ according to the four-nearest-neighbor and the eight-nearest-neighbor definitions, respectively, based on the geographical coordinates of the centroid for each municipality.[1] Note that in this case the matrices are asymmetric before and after row standardization, since the nearest neighbors criterion defines spatial relations asymmetrically.

Table 1 presents the feasible ranges of the spatial parameters ρ and λ according to the spatial structures as specified above. As can be seen from this table, the lower bounds of the spatial coefficients are significantly less than -1 for all matrices except for those created with rook contiguity. Moreover, the row-standardized weights matrices that have been constructed with rook definition over the regular lattices and with the nearest neighbors over the Greek geographical structure yield complex eigenvalues.

Having described the formation of $\mathbf{W}_1$ and $\mathbf{W}_2$ matrices, the spatial processes are generated as follows: the SAR(1) process as $\mathbf{y} = \left(\mathbf{I} - \rho\mathbf{W}_1\right)^{-1}\boldsymbol{\varepsilon}$, the SMA(1) process as $\mathbf{y} = \left(\mathbf{I} - \lambda\mathbf{W}_2\right)\boldsymbol{\varepsilon}$ and the SARMA(1, 1) process as $\mathbf{y} = \left(\mathbf{I} - \rho\mathbf{W}_1\right)^{-1}\left(\mathbf{I} - \lambda\mathbf{W}_2\right)\boldsymbol{\varepsilon}$, where the spatial parameters ρ and λ are allowed to take values within the feasible range intervals defined in Table 1 and the vector of random errors ε is considered

---

[1] The Greek geographical structure is obtained from https://geodata.gov.gr/en/dataset/oria-demon-kalli krates, excluding thought the Mountain Athos region.

Table 1 Characteristics of the weights matrices along with the lower and upper limits of the spatial parameters ρ and λ

n	Weights (row-standardized)	Complex eigenvalues	Most negative purely real eigenvalue	Largest eigenvalue	Lower limit ρ, λ	Upper limit ρ, λ
100	Rook	Yes	−1	1	−1	1
100	Queen	No	−0.508	1	−1.970	1
400	Rook	Yes	−1	1	−1	1
400	Queen	No	−0.520	1	−1.921	1
325	4-Nearest Neighbors	Yes	−0.676	1	−1.478	1
325	8-Nearest Neighbors	Yes	−0.368	1	−2.720	1

as $N(0, I)$. The whole simulation process is conducted in R using the SPDEP package developed by Bivand (2015) for the spatial weights matrices manipulation as well as for the generation of the processes. Each spatial generation process is then estimated for all three model specifications, i.e., for the SAR(1) with $\mathbf{W}_1$, for the SMA(1) with $\mathbf{W}_2$ and for the SARMA(1, 1) with $\mathbf{W}_1$ and $\mathbf{W}_2$ as the weights matrices, respectively, by maximizing the log-likelihood function, using the nlminb function in R, so that the values of all three information criteria can be calculated. The best fitted model is selected according to the minimum value of any criterion based on 1000 replications.

Table 2 presents the percentages of model selection by any of the three criteria when the true generating process is the SAR(1) model. As can be seen from this table, the selection behavior of all three information criteria is pretty much similar, independent of the construction of the weights matrix, and it depends mainly on the magnitude of the parameter ρ, meaning that the true SAR(1) process is selected more often by all criteria as the absolute value of ρ increases for given sample, reaching almost at the 99% level based on the BIC criterion for $n = 400$. It turns out though that the BIC criterion has the best behavior, among the other two criteria, in terms of most frequently selecting the true model regardless of sample size and the value of ρ, a result that confirms the overall well-known behavior of BIC in terms of selecting parsimonious models over AIC. The HQC criterion behaves closely to BIC criterion but with lower probability of selecting the true model. For small absolute values of ρ, the second best fitted model is the SMA(1) process selected by all criteria, whereas for large absolute values of ρ, the second best fitted model selected mainly by AIC is the SARMA(1, 1) process. Finally, as sample size increases, the true model is selected more often even for small values of ρ, i.e., for $\rho = 0.2$, the SAR(1) process is selected by BIC 64.9% and 84.4% for $n = 100$ and 400, respectively. It should be noted that the corrected AIC criterion, although it is not reported, behaves very similarly, if not identically in most cases, as the AIC criterion. The AICc selects slightly more frequently the correct spatial process than the AIC for small sample sizes, whereas for large sample sizes the selection rate is almost the same. For example, for $n = 100$ and for values of $\rho = -0.9$, -0.5, 0.2 and 0.8, the AICc selects the SAR(1) process 82.9%, 81.3%, 59.4% and 84.5%, respectively, as

Table 2 Percentage of selections for all three spatial processes based on AIC, BIC and HQC when the true data-generating model is the SAR(1) process for various values of ρ using 1000 replications

N	IC	Model	ρ									
			− 1.3	− 1.1	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)			80.7	81.1	80.4	57.4	58.8	79.6	83.2	83.8
		SMA(1)			0	0	4.8	34.1	31.3	8.4	0	0
		SARMA(1, 1)			19.3	18.9	14.8	8.5	9.9	12	16.8	16.2
	BIC	SAR(1)			96.8	96.5	92.1	62.4	64.9	87.6	96.7	96.9
		SMA(1)			0	0	5.9	36.5	32.8	9.3	0.4	0
		SARMA(1, 1)			3.2	3.5	2	1.1	2.3	3.1	2.9	3.1
	HQC	SAR(1)			89.3	89.3	87.4	60.7	63.5	85.2	92.6	91.7
		SMA(1)			0	0	5.1	35.9	32.2	9.3	0.1	0
		SARMA(1, 1)			10.7	10.7	7.5	3.4	4.3	5.5	7.3	8.3
400	AIC	SAR(1)			83.1	83.2	84.7	76.8	75.8	82.8	82.8	82.1
		SMA(1)			0	0	0	13.1	14.5	0	0	0
		SARMA(1, 1)			16.9	16.8	15.3	10.1	9.7	17.2	17.2	17.9
	BIC	SAR(1)			98.3	98.3	98.5	85.5	84.4	98.5	98.7	98.6
		SMA(1)			0	0	0.1	13.9	15.2	0.4	0	0
		SARMA(1, 1)			1.7	1.7	1.4	0.6	0.4	1.1	1.3	1.4
	HQC	SAR(1)			94.1	94.7	94	83.4	82.4	94.4	93.8	93
		SMA(1)			0	0	0	13.8	15.2	0.1	0	0
		SARMA(1, 1)			5.9	5.3	6	2.8	2.4	5.5	6.2	7
325	AIC	SAR(1)	84.3	84.1	83.8	84.2	84.3	69.5	69.4	83.9	84.3	83.9
		SMA(1)	0	0	0	0	0.6	22.2	21.3	0.4	0	0
		SARMA(1, 1)	15.7	15.9	16.2	15.8	15.1	8.3	9.3	15.7	15.7	16.1
	BIC	SAR(1)	98.2	98.2	98.1	98	97.1	76.3	77.3	97.3	98.1	97.8
		SMA(1)	0	0	0	0	1	22.6	21.6	1.3	0	0
		SARMA(1, 1)	1.8	1.8	1.9	2	1.9	1.1	1.1	1.4	1.9	2.2
	HQC	SAR(1)	94.7	94.8	94.7	94.6	94.2	74	75.3	93.1	94.6	93.8
		SMA(1)	0	0	0	0	0.9	22.5	21.4	1	0	0
		SARMA(1, 1)	5.3	5.2	5.3	5.4	4.9	3.5	3.3	5.9	5.4	6.2

opposed to the relevant AIC selection rates of 80.7%, 80.4%, 58.8% and 83.2% presented in Table 2.

Similar conclusions can be made for the special case of using a real geographical structure to generate the spatial processes. The main interesting feature of this case is that the SAR(1) process is well defined even for values of ρ smaller than −1, as can be seen from Table 1. For all these values, the behavior of all information criteria is similar, if not identical, to the largest absolute value of the autoregressive parameter within the range of (−1, 1), as can be seen from Table 2 for suggestive

values of ρ equal to -1.3 and -1.1.[2] Hence, the presence of these values does not provide any additional information of relevance to the Monte Carlo analysis, but these values do exist mathematically to ensure the consistency of the Jacobian matrix so that the likelihood function can be computed. In practice, however, it is very unlikely to obtain such values as estimates of the spatial autoregressive parameter, although the feasible range of all values of ρ obtained from the construction of the weights matrix, based on a specific geographical structure, is included in the maximization process.

The selection rates for all three spatial processes based on all information criteria when the true generating process is the SMA(1) model are reported in Table 3. It is evident that all three information criteria behave quite similarly, as in the previous case, regardless of the construction of the weights matrix. The SMA(1) process is selected more often as the absolute value of the moving average parameter λ increases for all samples, reaching almost at the 99% level based on the BIC criterion for $n = 400$. Among all criteria, the SMA(1) process is selected most frequently for any given value of λ and sample size by BIC followed by HQC and AIC. For small absolute values of λ the second best fitted model is the SAR(1) process selected by all criteria, as equivalently the SMA(1) process was the second best fitted model selected by all criteria in the case of the true generating SAR(1) process for small absolute values of the autoregressive parameter, whereas for large absolute values of λ the selection of the SARMA(1, 1) process appears to be as the second choice for the AIC criterion. It is also worth mentioning that the selection process of the true model is improving as sample increases by all criteria and for all values of the moving average parameter, i.e., for $\lambda = 0.2$, the SMA(1) process is selected by AIC 58.9% and 72.4% for $n = 100$ and 400, respectively. As in the case of the SAR(1) process, the AICc criterion selects the correct spatial moving average process slightly more frequently than the AIC only for small sample sizes.

Moreover, the construction of the $\mathbf{W}_2$ matrix to generate SMA(1) processes, either as a queen definition or as a special case of a real geographical structure with eight neighbors contiguity definitions, resulted to values of the moving average parameter smaller than -1 for both cases, as Table 1 reports. Like the SAR(1) case, the use of these values does not provide any additional information to the overall analysis, since the behavior of all information criteria in terms of selecting the best fitted model remains the same as their behavior for the largest absolute value of the moving average parameter within the range of $(-1, 1)$, as can be seen from Table 3 for selected values of λ equal to -1.4 for the queen formation and -1.4 and -2.4 for the real geographical structure, as presented in Table 3.[3]

Tables 4 and 5 report the selection percentages of each candidate model based on all three information criteria obtained through this simulation analysis when the true generating model is the SARMA(1, 1) process for specific values of ρ and λ, since

---

[2] Note that similar results can be obtained for values of ρ equal to -1.4 and -1.2.

[3] Note that similar results can be obtained for values of λ equal to -1.1, -1.3, -1.5 and -1.8 for the queen formation and for values of λ equal to -1.1, -1.3, -1.5, -1.8, -2.0 and -2.6 for the special geographical case with eight nearest neighbors.

Table 3 Percentage of selections of all three spatial processes based on AIC, BIC and HQC information criteria when the true data generating model is the SMA(1) process for various values of λ using 1000 replications

N	IC	Model	λ									
			− 2.4	− 1.4	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)		0	4.5	8.2	22.1	38.8	32.2	9	0.1	0
		SMA(1)		83.7	83.4	81.9	70.9	53.9	58.9	80.3	84.1	83
		SARMA(1, 1)		16.3	12.1	9.9	7	7.3	8.9	10.7	15.8	17
	BIC	SAR(1)		0.4	6.8	9.7	22.8	40.9	35	9.5	0.2	0
		SMA(1)		96.5	92.3	89.4	76.4	57.8	63	87.3	96.2	93.7
		SARMA(1, 1)		3.1	0.9	0.9	0.8	1.3	2	3.2	3.6	6.3
	HQC	SAR(1)		0.3	6	9.5	22.6	40	33.8	9.5	0.1	0
		SMA(1)		91.4	89.5	87.3	74.3	56.5	61.3	85.1	92.3	89.8
		SARMA(1, 1)		8.3	4.5	3.2	3.1	3.5	4.9	5.4	7.6	10.2
400	AIC	SAR(1)		0	0	0	0	2	25.4	18.9	0.4	0
		SMA(1)		84	82.6	82.7	83.5	66.1	72.4	83.6	83.8	83.5
		SARMA(1, 1)		16	17.4	17.3	14.5	8.5	8.7	16	16.2	16.5
	BIC	SAR(1)		0	0	0.1	4.2	26.1	20	0.5	0	0
		SMA(1)		98.5	98.7	98.7	95.3	73.5	79.3	98.2	98.5	98.7
		SARMA(1, 1)		1.5	1.3	1.2	0.5	0.4	0.7	1.3	1.5	1.3
	HQC	SAR(1)		0	0	0	3.6	25.8	19.3	0.4	0	0
		SMA(1)		94	94.5	94.3	92.7	71.3	77.4	94	94.4	93.7
		SARMA(1, 1)		6	5.5	5.7	3.7	2.9	3.3	5.6	5.6	6.3
325	AIC	SAR(1)	0	0	0.7	1.5	10.9	32.7	26.8	1.9	0	0
		SMA(1)	83.8	82.2	83.3	83.2	79.3	59.1	63.8	83.2	84.6	83.5
		SARMA(1, 1)	16.2	17.8	16	15.3	9.8	8.2	9.4	14.9	15.4	16.5
	BIC	SAR(1)	0	0	1.7	3.5	11.5	34.1	28.1	3	0	0
		SMA(1)	98.2	98.8	97.7	96.1	88	65.4	71.2	96.2	97.8	98
		SARMA(1, 1)	1.8	1.2	0.6	0.4	0.5	0.5	0.7	0.8	2.2	2
	HQC	SAR(1)	0	0	1.2	2.9	11.5	33.8	28	2.9	0	0
		SMA(1)	93.5	93.5	93.4	92.2	85.8	63.9	69.7	92.4	94.1	94.3
		SARMA(1, 1)	6.5	6.5	5.4	4.9	2.7	2.3	2.3	4.7	5.9	5.7

it was not technically feasible to include all results in one table for all possible combinations of ρ and λ. For this purpose, Tables 4 and 5 report only simulation results for small, moderate and large positive values of ρ and λ, respectively, against several values of the other parameter.[4] As can be seen from these tables, the main findings can be summarized into four points, meaning that the percentage of selecting

---

[4] Simulations have been conducted for small, moderate and large negative values of ρ and λ, against the reported values of Tables 4 and 5 of the other parameters, including cases with the extreme negative values.

Table 4 Percentage of selections for all three spatial processes based on AIC, BIC and HQC information criteria when the true data generating model is the SARMA(1, 1) process for $\rho = 0.2$, 0.5 and 0.9 and for various values of λ using 1000 replications

N	IC	ρ	0.2							
		Model/λ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	10.5	16.8	39.9	57.6	43.9	6.4	0.1	0
		SMA(1)	58.3	58.2	52.1	39.4	34.1	60.8	62.1	56.1
		SARMA(1, 1)	31.2	25	8	3	22	32.8	37.8	43.9
	BIC	SAR(1)	20.3	25.8	43.3	59	52.8	7.7	0.2	0
		SMA(1)	75.5	72.7	56.6	40	41.6	80.7	84.9	81
		SARMA(1, 1)	4.2	1.5	0.1	1	5.6	11.6	14.9	19
	HQC	SAR(1)	16	22	42.7	58.3	48.3	7.3	0.2	0
		SMA(1)	68	67.5	56.2	39.8	38.9	71.5	73	69
		SARMA(1, 1)	16	10.5	1.1	1.9	12.8	21.2	26.8	31
400	AIC	SAR(1)	0	0	7	58	37.4	0.3	0	0
		SMA(1)	17.5	18.7	22.9	24.2	15.5	31.9	26.2	19.8
		SARMA(1, 1)	82.5	81.3	70.1	17.8	47.1	67.8	73.8	80.2
	BIC	SAR(1)	0.1	0.9	26.7	68.2	62.1	0.4	0	0
		SMA(1)	55	57.3	60.7	31.8	25.2	72.4	67.6	59.1
		SARMA(1, 1)	44.9	41.8	12.6	0	12.7	27.2	32.4	40.9
	HQC	SAR(1)	0	0.1	15	66.2	53.4	0.4	0	0
		SMA(1)	32.5	34	40.1	30.5	21	51.7	44	36.6
		SARMA(1, 1)	67.5	65.9	44.9	3.3	25.6	47.9	56	63.4
325	AIC	SAR(1)	1.1	2.6	22	58.5	43.1	0.8	0	0
		SMA(1)	37.1	39	42.5	36.3	22.1	41.5	21.8	6.1
		SARMA(1, 1)	61.8	58.4	35.5	5.2	34.8	57.7	78.2	93.9
	BIC	SAR(1)	7.8	14.6	38.9	62.1	63.5	0.9	0	0
		SMA(1)	74.1	73.2	60.4	37.9	28.6	75.6	56.9	27.7
		SARMA(1, 1)	18.1	12.2	0.7	0	7.9	23.5	43.1	72.3
	HQC	SAR(1)	2.5	7.4	34.9	61.6	55.4	0.8	0	0
		SMA(1)	53.8	55.7	55.9	37.7	25.6	56.7	38	14.1
		SARMA(1, 1)	43.7	36.9	9.2	0.7	19	42.5	62	85.9
N	IC	ρ	0.5							
		Model/λ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	11.1	20.8	53	77.6	69.2	33.4	0.6	0.1
		SMA(1)	4.3	5.1	7.7	8.9	5.5	7.9	16.3	10.7
		SARMA(1, 1)	84.6	74.1	39.3	13.5	25.3	58.7	83.1	89.2
	BIC	SAR(1)	32.9	46	76	85.6	85.9	51.9	1.7	0.1
		SMA(1)	16.8	17.3	17.2	13.1	5.8	13.8	40.5	29.6
		SARMA(1, 1)	50.3	36.7	6.8	1.3	8.3	34.3	57.8	70.3
	HQC	SAR(1)	21	31.3	66	82.8	78.3	42.8	1	0.1
		SMA(1)	8.4	9.8	12.4	11.7	5.7	10.6	24.9	17.8
		SARMA(1, 1)	70.6	58.9	21.6	5.5	16	46.6	74.1	82.1

Table 4 (continued)

N	IC	ρ	0.5							
		Model/λ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
400	AIC	SAR(1)	0	0	11.3	65.9	59.2	3.8	0	0
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	100	100	88.7	34.1	40.8	96.2	100	100
	BIC	SAR(1)	0.4	2.6	47.5	91.6	89.4	21.2	0	0
		SMA(1)	0	0	0.2	0.4	0.1	0.3	2	0.3
		SARMA(1, 1)	99.6	97.4	52.3	8	10.5	78.5	98	99.7
	HQC	SAR(1)	0	0.4	24.3	80.7	76.8	10	0	0
		SMA(1)	0	0	0	0	0	0.2	0.1	0
		SARMA(1, 1)	100	99.6	75.7	19.3	23.2	89.8	99.9	100
325	AIC	SAR(1)	1.2	3.1	28.4	70.1	65.8	7.9	0	0
		SMA(1)	0	0	0	0.2	0.7	0	0.2	0
		SARMA(1, 1)	98.8	96.9	71.6	29.7	33.5	92.1	99.8	100
	BIC	SAR(1)	10.7	20.8	63.9	93.9	90.8	33.8	0	0
		SMA(1)	0.3	0.7	1.4	1.9	0.7	0.3	0.8	0.1
		SARMA(1, 1)	89	78.5	34.7	4.2	8.5	65.9	99.2	99.9
	HQC	SAR(1)	3.4	8.8	44.8	83.6	80.7	16.5	0	0
		SMA(1)	0	0	0.3	0.9	0.7	0.1	0.3	0
		SARMA(1, 1)	96.6	91.2	54.9	15.5	18.6	83.4	99.7	100

N	IC	ρ	0.9							
		Model/λ	− 0.9	− 0.8	−0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	2.5	5.1	32	72.1	79.4	56	24.3	9.7
		SMA(1)	0	0	0	0	0	0	0.1	0.1
		SARMA(1, 1)	97.5	94.9	68	27.9	20.6	44	75.6	90.2
	BIC	SAR(1)	10	19.4	60.1	91.2	94.9	81.1	48.1	22.6
		SMA(1)	0	0	0	0	0	0.1	0.2	0.4
		SARMA(1, 1)	90	80.6	39.9	8.8	5.1	18.8	51.7	77
	HQC	SAR(1)	4.6	10.5	45.3	83.8	89.4	68.2	33.2	14.3
		SMA(1)	0	0	0	0	0	0.1	0.2	0.3
		SARMA(1, 1)	95.4	89.5	54.7	16.2	10.6	37.7	66.6	85.4
400	AIC	SAR(1)	0	0	3	53.5	63.1	17.4	0.6	0.1
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	100	100	97	46.5	36.9	82.6	99.4	99.9
	BIC	SAR(1)	0	0	17.5	86.8	91.3	51.2	7.8	0.6
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	100	100	82.5	13.2	8.7	48.8	92.2	99.4
	HQC	SAR(1)	0	0	7.7	71.4	79.4	31.1	2.1	0.3
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	100	100	92.3	28.6	20.6	68.9	97.9	99.7

Table 4 (continued)

N	IC	ρ	0.9							
		Model/λ	− 0.9	− 0.8	−0.5	− 0.2	0.2	0.5	0.8	0.9
325	AIC	SAR(1)	0	0.2	11.7	63.5	66.3	10.4	0.3	0
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	100	99.8	88.3	36.5	33.7	89.6	99.7	100
	BIC	SAR(1)	1	3.7	40.8	89.7	92.5	41	1	0.1
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	99	96.3	59.2	10.3	7.5	59	99	99.9
	HQC	SAR(1)	0.1	0.9	21.8	79.4	82.2	23	0.5	0
		SMA(1)	0	0	0	0	0	0	0	0
		SARMA(1, 1)	99.9	99.1	78.2	20.6	17.8	77	99.5	100

the right model by all information criteria increases, regardless of the construction of the weights matrices: a) as the absolute values of ρ, λ and n increase, b) as the absolute value of λ increases for given values of ρ and n, c) as the absolute value of ρ increases for given values of λ and n and d) as sample size increases for given values of ρ and λ.[5] Unlike the previous two cases, AIC has now the dominant role of selecting the right model most frequently, followed by HQC, especially for small sample sizes, whereas the BIC criterion in this case is the least reliable criterion for selecting the true model, a finding that was expected, since, as discussed, the AIC criterion has the tendency to select large models.

For small values of the moving average parameter as well as for the autoregressive parameter, all criteria select almost exclusively the SAR(1) process and/or the SMA(1) process and not the true SARMA(1, 1) process, especially for small sample size. For example, for $\lambda=0.2$ and $\rho=0.9$, the SAR(1) process is selected 79.4% by AIC, 94.9% by BIC and 89.4% by HQC, as Table 4 reports for $n=100$, whereas for $\rho=0.2$ and for $\lambda=0.9$ the SMA(1) process is selected 56.1% by AIC, 81% by BIC and 69% by HQC, as Table 5 reports for $n=100$. Clearly, when the moving average (autoregressive) parameter is very low, the autoregressive (moving average) term prevails and that is why all information criteria prefer model SAR(1) (SMA(1)) instead of the true SARMA(1, 1) model. However, when the values of both parameters are large, all criteria select more frequently the true SARMA (1, 1) process, i.e., for values of λ and ρ equal to (0.8, 0.9) and (0.9, 0.8), the SARMA(1, 1) process is selected 75.6% and 95.9%, respectively, by AIC criterion, even for $n=100$. Unlike

---

[5] The term absolute value is used to include all cases. Similar results are also obtained for the negative values of both parameters, i.e., for values of −0.2, −0.5 and −0.9 against all other values of the other parameter, although are not reported here, with minor exceptions regarding the magnitude of the selection rate of the true model for small sample size.

Table 5 Percentage of selections for all three spatial processes based on AIC, BIC and HQC information criteria when the true data generating model is the SARMA(1, 1) process for of $\lambda = 0.2$, 0.5 and 0.9 and for various values of ρ using 1000 replications

N	IC	λ	0.2							
		Model/ρ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	59.2	59.6	61.3	48.2	43.9	69.2	79.9	79.4
		SMA(1)	0	0	5.8	45.4	34.1	5.5	0.1	0
		SARMA(1, 1)	40.8	40.4	32.9	6.4	22	25.3	20	20.6
	BIC	SAR(1)	84	84.3	80	50.8	52.8	85.9	94.4	94.9
		SMA(1)	0	0.1	11.6	48.9	41.6	5.8	0.2	0
		SARMA(1, 1)	16	15.6	8.4	0.3	5.6	8.3	5.4	5.1
	HQC	SAR(1)	70.7	70.7	71.5	50.3	48.3	78.3	88	89.4
		SMA(1)	0	0	8.4	48.1	38.9	5.7	0.2	0
		SARMA(1, 1)	29.3	29.3	20.1	1.6	12.8	16	11.8	10.6
400	AIC	SAR(1)	22	23.8	29.4	33.9	37.4	59.2	66.9	63.1
		SMA(1)	0	0	0	23.4	15.5	0	0	0
		SARMA(1, 1)	78	76.2	70.6	42.7	47.1	40.8	33.1	36.9
	BIC	SAR(1)	62.3	64.6	70.8	56.9	62.1	89.4	92.8	91.3
		SMA(1)	0	0	0.1	40.3	25.2	0.1	0	0
		SARMA(1, 1)	37.7	35.4	29.1	2.8	12.7	10.5	7.2	8.7
	HQC	SAR(1)	39.5	42.1	47.5	47.9	53.4	76.8	82	79.4
		SMA(1)	0	0	0	34.4	21	0	0	0
		SARMA(1, 1)	60.5	57.9	52.5	17.7	25.6	23.2	18	20.6
325	AIC	SAR(1)	44.3	45.4	49	47	43.1	65.8	69.9	66.3
		SMA(1)	0	0	0.6	39.1	22.1	0.7	0	0
		SARMA(1, 1)	55.7	54.6	50.4	13.9	34.8	33.5	30.1	33.7
	BIC	SAR(1)	81.4	82.5	84.6	55.5	63.5	90.8	93.2	92.5
		SMA(1)	0	0	3	44.2	28.6	0.7	0	0
		SARMA(1, 1)	18.6	17.5	12.4	0.3	7.9	8.5	6.8	7.5
	HQC	SAR(1)	63.2	65	68	54.2	55.4	80.7	84.8	82.2
		SMA(1)	0	0	1.2	43.5	25.6	0.7	0	0
		SARMA(1, 1)	36.8	35	30.8	2.3	19	18.6	15.2	17.8
N	IC	λ	0.5							
		Model/ρ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	6.4	7.3	9.8	12	6.4	33.4	54.3	56
		SMA(1)	0	0	5.7	63	60.8	7.9	0.3	0
		SARMA(1, 1)	93.6	92.7	84.5	25	32.8	58.7	45.4	44
	BIC	SAR(1)	19.7	20.4	27.8	19.1	7.7	51.9	77.7	81.1
		SMA(1)	0	0	16.4	77.6	80.7	13.8	0.3	0.1
		SARMA(1, 1)	80.3	79.6	55.8	3.3	11.6	34.3	22	18.8
	HQC	SAR(1)	10.5	11.8	15.6	15.4	7.3	42.8	64.6	68.2
		SMA(1)	0	0	9.8	71.5	71.5	10.6	0.3	0.1
		SARMA(1, 1)	89.5	88.2	74.6	13.1	21.2	46.6	35.1	31.7

Table 5 (continued)

N	IC	λ	0.5							
		Model/ρ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
400	AIC	SAR(1)	0	0	0	0.3	0.3	3.8	17.9	17.4
		SMA(1)	0	0	0	24.7	31.9	0	0	0
		SARMA(1, 1)	100	100	100	75	67.8	96.2	82.1	82.6
	BIC	SAR(1)	0	0.1	0.5	1.2	0.4	21.2	50.9	51.2
		SMA(1)	0	0	0	61.1	72.4	0.3	0	0
		SARMA(1, 1)	100	99.9	99.5	37.7	27.2	78.5	49.1	48.8
	HQC	SAR(1)	0	0	0.2	0.4	0.4	10	31.1	31.1
		SMA(1)	0	0	0	39.6	51.7	0.2	0	0
		SARMA(1, 1)	100	100	99.8	60	47.9	89.8	68.9	68.9
325	AIC	SAR(1)	0.2	0.3	1	1.9	0.8	7.9	14.4	10.4
		SMA(1)	0	0	0	37.7	41.5	0	0	0
		SARMA(1, 1)	99.8	99.7	99	60.4	57.7	92.1	85.6	89.6
	BIC	SAR(1)	3	3.6	6.9	8.2	0.9	33.8	45.3	41
		SMA(1)	0	0	1.7	72.7	75.6	0.3	0	0
		SARMA(1, 1)	97	96.4	91.4	19.1	23.5	65.9	54.7	59
	HQC	SAR(1)	1.3	1.3	2.1	3.7	0.8	16.5	28.5	23
		SMA(1)	0	0	0.5	53.6	56.7	0.1	0	0
		SARMA(1, 1)	98.7	98.7	97.4	42.7	42.5	83.4	71.5	77
N	IC	λ	0.9							
		Model/ρ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
100	AIC	SAR(1)	0	0	0	0	0	0.1	3	9.7
		SMA(1)	0	0	2.7	56.3	56.1	10.7	1.1	0.1
		SARMA(1, 1)	100	100	97.3	43.7	43.9	89.2	95.9	90.2
	BIC	SAR(1)	0	0	0	0	0	0.1	7.9	22.6
		SMA(1)	0	0	8.6	80.6	81	29.6	4.4	0.4
		SARMA(1, 1)	100	100	91.4	19.4	19	70.3	87.7	77
	HQC	SAR(1)	0	0	0	0	0	0.1	4.4	14.3
		SMA(1)	0	0	5	68.7	69	17.8	1.9	0.3
		SARMA(1, 1)	100	100	95	31.3	31	82.1	93.7	85.4
400	AIC	SAR(1)	0	0	0	0	0	0	0	0.1
		SMA(1)	0	0	0	15.9	19.8	0	0	0
		SARMA(1, 1)	100	100	100	84.1	80.2	100	100	99.9
	BIC	SAR(1)	0	0	0	0	0	0	0.1	0.6
		SMA(1)	0	0	0	48.9	59.1	0.3	0	0
		SARMA(1, 1)	100	100	100	51.1	40.9	99.7	99.9	99.4
	HQC	SAR(1)	0	0	0	0	0	0	0	0.3
		SMA(1)	0	0	0	28.2	36.6	0	0	0
		SARMA(1, 1)	100	100	100	71.8	63.4	100	100	99.7

Table 5 (continued)

N	IC	λ	0.9							
		Model/ρ	− 0.9	− 0.8	− 0.5	− 0.2	0.2	0.5	0.8	0.9
325	AIC	SAR(1)	0	0	0	0	0	0	0	0
		SMA(1)	0	0	0	2.6	6.1	0	0	0
		SARMA(1, 1)	100	100	100	97.4	93.9	100	100	100
	BIC	SAR(1)	0	0	0	0	0	0	0	0.1
		SMA(1)	0	0	0	17.8	27.7	0.1	0	0
		SARMA(1, 1)	100	100	100	82.2	72.3	99.9	100	99.9
	HQC	SAR(1)	0	0	0	0	0	0	0	0
		SMA(1)	0	0	0	8	14.1	0	0	0
		SARMA(1, 1)	100	100	100	92	85.9	100	100	100

the previous two cases, the corrected AIC criterion selects less frequently the true SARMA(1, 1) process than the AIC criterion only for small sample sizes; i.e., for $n = 100$ and for values of ρ and λ equal to (0.2, 0.8) and (0.9, −0.5), the SARMA(1, 1) process is selected by AICc 36.2% and 66.2%, respectively, as opposed to the 37.8% and 68% selection rates by AIC reported in Table 4, and similar results can be found for all other values of λ and ρ. For large sample sizes, both criteria select the true model at the same rate.

It should be stated though that the whole simulation process is conducted even for identical values of both parameters ρ and λ due to the fact that the weights matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are different, excluding in that sense the presence of a common spatial root. If both weights matrices were exactly the same, the SARMA(1, 1) process would have been simply a spatial white noise process for all identical values of ρ and λ, causing serious issues in the whole Monte Carlo analysis. Lastly, similar conclusions can be found for the special geographical structure that was used to construct the weights matrices, indicating the behavior of the selection of all information criteria for values smaller than −1 for both parameters; it was similar to the extreme absolute values of both parameters and did not deliver any additional information to the overall analysis, as already has been recognized.

4 Concluding remarks

The objective of this study was to investigate the behavior of the three most frequently used information criteria, AIC, BIC and HQC, for model selection among competitive models for spatial data, using a Monte Carlo analysis. For this purpose, three spatial processes, the SAR(1), the SMA(1) and the SARMA(1, 1) processes, are generated either by using a hypothetical geographical structure based on regular grids or by using a real geographical structure based on a map of the administrative

units of Greece. For the autoregressive part, the weight matrix is constructed with a rook definition for the theoretical case of both grids and with the four closest neighbor definitions for the real case, whereas for the moving average part the weight matrix is constructed with a queen definition for the theoretical case of both grids and with the eight closest neighbor definitions for the real case. The three spatial models are estimated next, based on the predefined weights matrices, using the nlminb function in R that maximizes the log-likelihood function to obtain the corresponding values of all three information criteria. The best fitted model is selected by the minimum value of any of the three used information criteria.

Simulation results showed that the behavior of these criteria is not the same. BIC performs better for selecting models with small number of parameters, whereas AIC works better for large models, findings that coincide with the overall knowledge of the performance of these criteria. On the other hand, the HQC criterion always comes second in order among the three criteria either right after BIC for small models or right after AIC for large models. However, the selection mechanism of the true generating process contains a large volume of ambiguity, especially when the sample size is small and/or when the values of both parameters are also small, indicating that all three criteria have difficulty in successfully recognizing the true generating process. Things are even more complicated for spatial analysis.

First, one may argue that the results are sensitive to the construction of the weights matrices, although two different types of structures are employed for the theoretical part in this simulation process. The truth is that in spatial analysis, spatial dependence is defined exogenously for every model with a variety of neighborhood criteria, indicating that these matrices are not uniquely constructed. Moreover, the application of the same spatial neighborhood definition to different spatial structures can lead to completely dissimilar spatial weights matrices. For this purpose, this study considered additionally a real geographical structure, i.e., the spatial structure of Greece, which has a lot of geographical peculiarities resulting in quite asymmetric spatial weights matrices, either as an effort to alternatively support the simulated results or as an effort to minimize the validity of such an argument. It turns out that the behavior of all information criteria is not affected by the construction of these matrices since the same selection rate is obtained regardless of the geographical structures that were used either on the theoretical basis of regular grids or on the practical basis of a real geographical structure, as it is expresses through the patterns of a map.

The concept of properly defining the SARMA(1, 1) process is another important element in spatial analysis, which is highly related to the construction of the weights matrices $\mathbf{W}_1$ and $\mathbf{W}_2$. Typically, the SARMA(1, 1) process must be defined with different spatial weights matrices for its two components, so that issues concerning common spatial roots will not theoretically appear in the analysis, as it was carefully imposed in this applied simulation process. However, the estimation of this process can be executed even with identical weights matrices, but not with the same or near the same values of the autoregressive and moving average parameters. The question though remains unclear as how these weights matrices are viewed as different, since dissimilarity is a subjective and ambiguous concept. The truth is that spatial weights matrices are sparse matrices that only a small number of their elements are different

from zero, regardless of how the dependence is defined. The construction of the second weight matrix, i.e., the matrix that corresponds to the moving average part of the SARMA(1, 1) process, by an alternative method reassures that the two weights matrices are typically different, but not practically. The $\mathbf{W}_2$ matrix is constructed in such a way that leads to a more dense matrix than the $\mathbf{W}_1$ matrix, since it has some more elements different than zero. Therefore, it is hard to tell whether these two matrices are different, knowing that both matrices are sparse matrices and hence it may be difficult to evaluate the performance of these information criteria under these circumstances.

Another way to distinguish processes is to calculate the Frobenius distance between the covariance matrices of two processes. The Frobenius distance measures the distance between two matrices, and it is defined as:

$$F(A, B) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_{ij} - b_{ij} \right)^2}$$

where $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are square matrices having the same dimensions. Clearly, when the value of $F(A, B)$ is close to zero, the matrices A and B are very similar and identical in the case where $F(A, B) = 0$. Hence, the larger the value of the Frobenius distance, the more different the matrices are. The covariance matrices for the SAR(1), SMA(1) and SARMA(1, 1) processes are defined, respectively, as:

$$\mathrm{E}\left[\mathbf{y}\mathbf{y}' \right] = \sigma^2 \left[\left(\mathbf{I} - \rho \mathbf{W}'_1 \right) \left(\mathbf{I} - \rho \mathbf{W}_1 \right) \right]^{-1}$$

$$\mathrm{E}\left[\mathbf{y}\mathbf{y}' \right] = \sigma^2 \left(\mathbf{I} - \lambda \mathbf{W}_2 \right) \left(\mathbf{I} - \lambda \mathbf{W}'_2 \right)$$

$$\mathrm{E}\left[\mathbf{y}\mathbf{y}' \right] = \sigma^2 \left(\mathbf{I} - \rho \mathbf{W}_1 \right)^{-1} \left(\mathbf{I} - \lambda \mathbf{W}_2 \right) \left(\mathbf{I} - \lambda \mathbf{W}_2 \right)' \left[\left(\mathbf{I} - \rho \mathbf{W}_1 \right)^{-1} \right]'$$

and since σ^2 is the same for all processes, it can be ignored. Table 6 presents the Frobenius distance for cases that needed extra attention, i.e., for cases of common values for the two parameters, where the distinction between spatial processes is even more difficult. As can be seen from Table 6, it is easier to distinguish a SAR(1) process from a SMA(1) or a SARMA(1, 1) process than a SMA(1) from a SARMA(1, 1), where the value of the Frobenius distance increases as sample size increases and/or as the absolute value of both parameters increases.

Lastly and more importantly, the estimation procedure is another issue that needs special attention. It is unknown whether the estimated values of the spatial coefficients have been influenced by the very small lower limits of their feasible range intervals, resulting from the queen and the nearest neighbors spatial dependence definitions and/or by the maximization algorithm that the R statistical package uses through the nlminb function. Perhaps, an alternative estimation procedure may produce different results concerning the selection rates of all information criteria, especially for small sample sizes.

Table 6 Frobenius distance between covariance matrices of the following processes, *A*: SAR(1), *B*: SMA(1) and *C*: SARMA(1, 1)

	ρ	λ	$F(A, B)$	$F(A, C)$	$F(B, C)$
100	0.2	0.2	3.629	3.633	0.171
	0.5	0.5	11.889	11.897	0.374
	0.9	0.9	151.882	151.839	1.081
	0.2	− 0.2	1.299	1.319	0.213
	0.5	− 0.5	5.307	5.379	0.705
	0.9	− 0.9	142.808	143.196	3.677
	− 0.2	0.2	1.799	1.807	0.174
	− 0.5	0.5	8.061	8.076	0.415
	− 0.9	0.9	148.388	148.417	1.121
400	0.2	0.2	6.945	6.948	0.215
	0.5	0.5	22.582	22.588	0.458
	0.9	0.9	260.553	260.507	1.381
	0.2	− 0.2	2.593	2.609	0.274
	0.5	− 0.5	10.158	10.221	0.912
	0.9	− 0.9	243.283	243.643	4.682
	− 0.2	0.2	3.546	3.552	0.218
	− 0.5	0.5	15.474	15.485	0.504
	− 0.9	0.9	254.066	254.076	1.275
325	0.2	0.2	5.961	6.140	1.369
	0.5	0.5	21.413	22.080	4.017
	0.9	0.9	459.219	464.737	21.609
	0.2	− 0.2	2.340	2.736	1.382
	0.5	− 0.5	11.477	12.751	4.488
	0.9	− 0.9	445.544	457.854	28.885
	− 0.2	0.2	2.229	2.602	1.319
	− 0.5	0.5	6.823	7.872	3.628
	− 0.9	0.9	20.554	22.968	8.312

Information criteria are by far the best statistical tools for selecting and/or for evaluating models for any type of quantitative analysis. Their role is to guide the analyst in discovering the true underlying generating mechanism of any phenomenon based on a given dataset. However, it turns out that these criteria occasionally fail to select the true model, especially when the sample does not convey strong qualitative and quantitative evidence of the true population behavior of any variable. In this case, all criteria most probably will select the next closest behavior, but not the right one, as it was presented and analyzed in this study. For small sample sizes and for small absolute values of the autoregressive and/or the moving average parameters, the selection rate of the true model was not large enough, whereas in the contrary case the selection rate was so high that the true model was selected in several cases even with certainty.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds), Proceedings of the second international symposium of information theory, Academia Kiado, Budapest, pp 267–281

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc B 36:192–236

Bivand R (2015) Spdep: spatial dependence: weighting schemes. Statistics and models. R package version 0.5–82. http://CRAN.R-project.org/package=spdep

Bivand R, Hauke J, Kossowski T (2013) Computing the jacobian in gaussian spatial autoregressive models: an illustrated comparison of available methods. Geogr Anal 45:150–179

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden Day, San Francisco

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. New York, Springer

Cliff AD, Ord JK (1981) Spatial processes: models and applications. Pion Ltd, London

Diebold FX (2007) Elements of forecasting, 4th edn. Thomson, South- Western

Haining R (1978) The moving average process for spatial interaction. Trans Inst Br Geogr 3:202–225

Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. J R Stat Soc B 41:190–195

Hoeting JA, Davis RA, Merton AA, Thompson SE (2006) Model selection for geostatistical models. Ecol Appl 16(1):87–98

Huang JS (1984) The autoregressive moving average model for spatial analysis. Aust J Stat 26:169–178

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. Biometrika 76:297–307

Jones RH (1975) Fitting autoregressions. J Am Stat Assoc 70:590–592

Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee T-C (1985) The theory and practice of econometrics, 2nd edn. Wiley, NY

Lee H, Ghosh SK (2009) Performance of information criteria for spatial models. J Stat Simul Comput 79(1):93–106

LeSage J, Pace R (2009) Introduction to spatial econometrics. Chapman & Hall/CRC, Boca Raton

Mur J, Angulo AM (2007) Clues for discriminating between moving average and autoregressive models in spatial processes. Span Econ Rev 9:273–298

Ord JK (1975) Estimation methods for models of spatial interaction. J Am Stat Assoc 70:120–126

Schabenberger O, Gotway CA (2005) Statistical methods for spatial data analysis. Chapman & Hall/CRC, Boca Raton

Schwarz G (1978) Estimating the dimension of a model. Annal Stat 6:461–464

Shibata R (1976) Selection of the order of an autoregressive model by akaike's information criterion. Biometrika 63:117–126

Sugiura N (1978) Further analysis of the data by akaike's information criterion and the finite corrections. Commun Stat Theory Methods 7(1):13–26

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