



Evaluating the performance of AIC and BIC for selecting spatial econometric models

Christos Agiakloglou¹ · Apostolos Tsimpanos²

Received: 19 December 2020 / Accepted: 30 September 2022 / Published online: 26 December 2022
© The Author(s) 2022

Abstract

This study investigates using a Monte Carlo analysis the performance of the two most important information criteria, such as the Akaike's Information Criterion and the Bayesian Information Criterion, not only in terms of selecting the true spatial econometric model but also in term of detecting spatial dependence in comparison with the LM tests for the simple two spatial models SLM and SEM. The analysis is also extended by incorporating several other spatial econometric models, such as the SLX, SDM, SARAR and SDEM, along with heteroscedastic and non-normal errors. Simulation results show that under ideal conditions these criteria can assist the analyst to select the true spatial econometric model and detect properly spatial dependence.

Keywords Spatial dependence · Spatial econometric models · LM tests · Information criteria · Monte Carlo analysis

JEL Classification C20 · C21 · C52

1 Introduction

Spatial autocorrelation in errors is a very common problem in linear regression models with spatial data, which should be treated with caution, since it violates the assumption of the random sample, leading the analyst to ambiguous results. This behavior typically arises from observations corresponding to geographically proximate locations that are correlated because of their spatial dependence. As Anselin (1988a) has clarified, this spatial dependence along with spatial heterogeneity define

✉ Christos Agiakloglou
agiaklis@unipi.gr

¹ Department of Economics, University of Piraeus, Karaoli & Dimitriou 80, 185 34 Piraeus, Greece

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

the concept of spatial effects. For this purpose, a set of Lagrange Multiplier (LM) tests, known as spatial dependence tests (see [Burrige 1980](#); [Anselin 1988b](#); [Anselin et al. 1996](#)), have been developed in the literature to assist the analyst in terms of selecting and estimating the most appropriate spatial econometric model that considers the presence of spatial dependence. Moreover, spatially autocorrelated errors can also appear in spatial regression analysis as a symptom of a false indication of spatial dependence due to spurious behavior, as [Fingleton \(1999\)](#), [Mur and Trivez \(2003\)](#) and [Agiakloglou et. al. \(2015\)](#) have indicated.

Nevertheless, these LM tests that have been widely used in many empirical applications contain two important drawbacks. The first one is related to the restrictive alternate model structure imposed by these tests and the second one is associated with their reliability in selecting the right model. Indeed, as it is known, these tests are applied exclusively to the choice between a simple econometric model and a spatial model with a spatial lag structure either in the dependent variable or in the error, whereas in several cases their application often leads to inconsistent conclusions as to the choice of the right spatial econometric model, a problem that has been addressed by [Anselin and Florax \(1995\)](#). In addition, [LeSage and Pace \(2009\)](#) point out that spatial dependence tests were developed and established in the logic that their statistics are calculated solely from the residuals derived from the estimation of the simple econometric model using least squares estimation without requiring estimating the corresponding spatial econometric model. Clearly, these tests for spatial dependence versus the null hypothesis of no dependence do not require maximum likelihood estimation of the spatial model under the alternative hypothesis. For this reason and given the current availability of software, [LeSage and Pace \(2009\)](#) suggested that the selection of a spatial model should be made in the context of comparing the likelihoods of different models, while the analysis should start from a more general model that nests both the spatial lag model and the spatial error model.

Thus, it will be very interesting to investigate whether the use of any information criterion can help the analyst to select the true spatial econometric model, knowing that these criteria are usually applied to any quantitative analysis and that their performance has been limited explored to spatial econometric analysis. For this purpose, a Monte Carlo analysis is conducted to evaluate the performance of the two most frequently used information criteria, such as the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC), using only the three most important spatial econometric models, such as the SIM, SLM and SEM, suitable for the application of the LM tests, not only in terms of detecting spatial dependence but also in terms of selecting the right spatial econometric model as a complementary approach to model selection using the LM tests for these models. Simulation results show that these criteria can assist the analyst to identify the right spatial econometric model and spatial dependence more effectively in some cases than the LM tests. Note that the simulation process is conducted using rook and queen construction matrices along with a real geographical structure, i.e., the spatial structure of Greece, which has a lot of geographical peculiarities resulting in quite asymmetrical spatial weights matrices. The research is also expanded, in term of selecting the right spatial econometric model, using the two aforementioned information criteria, by considering more spatial econometric models, namely the SLX, SDM, SARAR,

and SDEM, as well as two non-ideal situations, such as heteroscedasticity and non-normality, where the results vary considerably. Hence, the objective of this research is concentrated on selecting the best-fitted spatial econometric model given the weights matrix formation, rather than searching for the most appropriate weights matrix formation knowing the spatial econometric model, as discussed in Zhang and Yu (2018).

The remaining of the paper is organized as follows. Section 2 presents the three simple econometric models, namely the SIM, SLM and SEM, along with the SDM, as a special case of the SEM, the information criteria AIC and BIC and analyses the strategies applied for the LM tests. Section 3 describes the design of the simulation process and discusses the results. Section 4 presents the extension of the Monte Carlo analysis with all the elements that have been added including the additional spatial econometric models, such as the SLX, SARAR, and SDEM, along with heteroscedastic and non-normal errors and discusses the results. The concluding remarks are included in Sect. 5.

2 The LM tests for spatial econometric models and the information criteria

Consider the Spatial Independent Model (SIM), also known as the Non-Spatial Econometric Model (NSEM), defined as:

$$y = X\beta + \epsilon$$

where y is a $(n \times 1)$ vector of observations of the dependent variable, X is a $[n \times (k + 1)]$ matrix of observations of k independent variables with values of 1 for its first column to include the presence of the constant term, β is the $[(k + 1) \times 1]$ vector of coefficients of the model and ϵ is the $(n \times 1)$ random vector following the standard assumptions of regression, i.e., $\epsilon \sim N(0, \sigma^2 I)$.

Spatial econometric models have been introduced in the literature as multidirectional extensions of the time series econometric models on the geographical space, defining in that sense dependence for the values of a variable according to the geographical positions of its values and of the values of all independent variables in the model, including the error term and not according to their chronological dependence. The spatial dependence is incorporated into the model by the presence of the spatial weights matrix W , which defines the spatial interactions between the n neighboring regions and it is used in its row-standardized form.¹

The two most important spatial econometric models are the Spatial Lag Model (SLM) defined as:

$$y = \rho W y + X\beta + \epsilon$$

¹ Several criteria have been proposed in the literature for spatial weights formation including boundary contiguity and distance measures, as can be seen in Cliff and Ord (1981) and in Anselin (1988a).

and the Spatial Error Model (SEM) defined as:

$$y = X\beta + \varepsilon$$

with

$$\varepsilon = \lambda W\varepsilon + u$$

where u is a $(n \times 1)$ random vector and Wy and $W\varepsilon$ are the spatially lagged vectors that incorporate spatial dependence consisting of weighted averages of the values of the variables in the n neighboring regions, while ρ and λ are the spatial lag coefficients of the dependent variable and the errors respectively. In addition, the SEM model can be expressed as a spatial econometric model with spatial lags for all variables of the model, known as the Spatial Durbin Model (SDM), defined as:

$$y = \rho Wy + X\beta + WX\theta + \varepsilon$$

where $\rho = \lambda$, $-\lambda\beta = \theta$, a condition that can be tested by performing the test presented by Mur and Angulo (2006), known as a common factor test, to investigate whether the model is actually a spatial error model or a more general spatial model, and the WX is the spatial lag matrix of the independent variables. Hence, when the SDM is estimated, while the true generating spatial model is the SEM, one should expect to get the same results.

It also important to mention that the values of the coefficients ρ and λ are not necessarily restricted strictly to the interval $(-1, +1)$, as in time series analysis. The estimation is implemented provided that the Jacobian matrix is non-singular, an outcome that is related to the eigenvalues of the spatial weights matrices. More specific, row-standardized spatial weights matrices have always the largest eigenvalue equals to unity, something that ensures that the upper limit of the interval will always be $+1$, while the value of the lower limit is unknown, and several times can be smaller than -1 (see LeSage and Pace 2009). Thus, if the coefficient takes values inside the feasible interval, corresponding to the applied weights matrix, the Jacobian determinant will be positive, comforting that its logarithm exists and therefore the log-likelihood function will be well defined.

The selection of the best fitted model for a given set of spatial data is typically made through the LM tests for spatial dependence. In particular, the LM test for a SEM (LM-ERR), introduced by Burridge (1980), and the LM test for a SLM (LM-LAG), presented by Anselin (1988b), are conducted on a SIM under the null hypothesis, against a SEM and a SLM under the alternative hypothesis, respectively, and the data contains no spatial effects if the null hypothesis is accepted by both tests. Issues arise when both tests reject the null hypothesis, a behavior that tends to appear very often in practice, as indicated by Anselin et al. (1996), where the tests cannot clearly identify the type of the spatial effect, unless one test rejects and the other one accepts the null hypothesis. For this reason, Florax et al. (2003) proposed to select the spatial econometric model for which the LM statistic will have the highest value, hereafter strategy I. Moreover, and as an effort to minimize this problematic behavior, comparable robust LM tests have been constructed by Anselin et al. (1996), namely the Robust LM-Error test (LM-EL) and the Robust LM-Lag

test (LM-LE), showing that these tests have more power in locating the correct spatial model than the simple LM tests. Hence, the presence of spatial dependence is inspected through these new tests, hereafter strategy II, and if the null hypothesis is rejected by both robust LM tests, the spatial econometric model is selected according to the highest value of one of the two robust LM statistics. Lastly, another strategy has been proposed by Florax et al. (2003), known as hybrid strategy, which combines the classical and the robust LM tests, but it turns out that this strategy leads to the same results as the classical approach (strategy I), as indicated by Florax et al. (2003) and proved by Mur and Angulo (2009).

The choice also of the best fitted model for a given set of spatial data can also be conducted based on the values of a pre-selected information criterion, as this technique is typically applied in every quantitative analysis that involves model selection. Hence, the values of the two most often used in practice information criteria, such as the Akaike Information Criterion (AIC), presented by Akaike (1973), and the Bayesian Information Criterion (BIC), developed by Schwarz (1978), as an attempt to improve the performance of AIC, defined respectively as:

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

and

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

where $\ln \hat{L}$ is the maximized value of the log-likelihood function, p is the number of parameters estimated from the econometric model and n is the sample size used for the estimation of the model, can be computed right after the estimation of any spatial model by maximum likelihood estimation and the best fitted model is selected according to the minimum value of that criterion, a technique that has been used in practice by Chi and Zhu (2008) on their empirical work for demographic data to select the best fitted spatial econometric model.

For this reason, it will be very interesting to study the performance of these criteria for spatial data in lieu of the performance of the LM tests not only in terms of identifying spatial dependence but also in terms of selecting the best spatial econometric model for these limited alternatively spatial models, knowing that each criterion has its own penalty function and therefore different pattern for model selection. Indeed, as it is known, AIC has the tendency to select models with large number of parameters, whereas BIC typically chooses small models, as a true approximation of their unknown population behavior.² Note that the behavior of these criteria has been investigated for geostatistical models by Hoeting et al. (2006) and Lee and Ghosh (2009) and for spatial processes by Agiakloglou and Tsimpanos (2021), but not for spatial econometric models.

² Further details regarding the properties of AIC and BIC information criteria can be found in Judge et al. (1985), Burnham and Anderson (2002) and in Diebold (2007).

3 Simulation results

The performance of the two previously presented information criteria, i.e., the AIC and the BIC, is investigated in terms of selecting the true spatial econometric model among the three alternative models, namely, the SIM, the SLM and the SDM, using a Monte Carlo analysis, where the SEM is included as an estimated SDM. The simulation process is conducted by considering only one independent variable derived from a uniform $U(0, 10)$ distribution which is simulated only once and then it remained constant for all iterations. Thus, the matrix X has dimensions $(n \times 2)$ consisting of one independent variable and a column of ones to estimate the constant term. The vector of coefficients β with dimension (2×1) is assumed to take values of one for both of its elements. The random error vector ε is derived from a $N(0, \mathbf{I})$ distribution and it is added to the vector $X\beta$ to produce the vector of the dependent variable y for the non-spatial econometric model.

Spatial dependence is introduced into the models by defining a row-standardized spatial weights matrix W with dimensions $(n \times n)$ which is constructed using the rook (four neighbors-common edge) and the queen (eight neighbors-common edge and vertex) contiguity definitions, over a squared regular lattice for dimensions 10×10 and 20×20 providing samples of 100 and 400 observations. Calculation of the eigenvalues for the four spatial matrices shows that the lower value of the spatial coefficient for the rook criterion is -1 for both sample sizes, meaning that the feasible range of values that the parameter can take is $(-1, 1)$, whereas the lower value for the queen criterion is -1.97 for sample size of 100 observations and -1.921 for sample size of 400 observations, leading to feasible ranges of $(-1.97, 1)$ and $(-1.921, 1)$ for sample sizes of 100 and 400 observations respectively (see Bivand et al. 2013 and Agiakloglou and Tsimpanos 2021). In addition, the simulation process is extended by including weight matrices derived from a geographical structure of Greece at the local authority districts of Kallikrates Operational Programme consisting of 325 municipalities.³ The weights matrices are constructed to capture real geographical structure according to the 4-nearest neighbors and the 8-nearest neighbors definitions, based on the geographical coordinates of the centroid for each municipality. Therefore, the formation of the weights matrices will be considered as given for the whole Monte Carlo analysis, although Kelejian and Piras (2011) proposed a J-test for investigating alternative spatial econometric models with different weights matrices under the null hypothesis of a specific spatial econometric model (see also Jin and Lee 2013).

The simulation process is conducted in R using the SPDEP package developed by Bivand (2015) to generate: a) the SIM, b) the SLM, by multiplying the right-hand side of the SIM by the spatial multiplier $(\mathbf{I} - \rho W)^{-1}$, that is:

$$y = (\mathbf{I} - \rho W)^{-1} X\beta + (\mathbf{I} - \rho W)^{-1} \varepsilon$$

³ The Greek geographical structure is obtained from <https://geodata.gov.gr/en/dataset/oria-demon-kallikrates>, excluding thought the Mountain Athos region.

Table 1 Percentage of selections for all three models based on AIC and BIC as well as on LM strategies based on the 5% nominal level when the true generating model is the SIM using 1000 replications

IC	Model	N	100		400		325	
			Rook	Queen	Rook	Queen	K=4	K=8
AIC	SIM		75.70	76.40	79.40	78.50	81.10	77.50
	SLM		15.30	14.50	12.00	12.90	12.00	14.20
	SDM		9.00	9.10	8.60	8.60	6.90	8.30
BIC	SIM		94.90	95.50	98.10	97.90	98.70	97.80
	SLM		4.50	3.60	1.60	1.70	1.20	2.20
	SDM		0.60	0.90	0.30	0.40	0.10	0.00
<i>Strategy</i>								
I	SIM		89.60	91.60	91.10	91.00	92.40	90.20
	SLM		5.80	5.00	4.30	4.50	4.00	5.70
	SEM		4.60	3.40	4.60	4.50	3.60	4.10
II	SIM		89.70	91.20	90.40	91.50	91.20	90.30
	SLM		5.40	4.40	4.80	4.30	5.10	5.60
	SEM		4.90	4.40	4.80	4.20	3.70	4.10

Bold indicates the correct model

c) the SEM as:

$$y = X\beta + (I - \lambda W)^{-1}u$$

where the vector error term u with $(n \times 1)$ dimensions is derived from $N(0, I)$ distribution and the spatial parameters ρ and λ can take values within the feasible range intervals, as previously mentioned for these spatial weights matrices formulation, and d) the SDM as follows:

$$y = (I - \rho W)^{-1}(X\beta + WX\theta + \epsilon)$$

while no restrictions are applied for the values of θ . All models are estimated by maximizing the log-likelihood function so that the values of both information criteria can be calculated. The best fitted econometric model is selected according to the minimum value of the pre-defined criterion based on 1000 replications. Note that the SEM is not estimated directly but only indirectly as a SDM and the information criterion should select the SDM when the true generating model is the SEM.

Table 1 presents the percentage selection rates of both criteria when the true generating model is the SIM. As can be seen from this table, the BIC performs very well in terms of selecting the true model with selection rates close to 95% and 98% for samples of 100 and 400 observations respectively, regardless of the spatial weights matrix formation, including the Greek weight matrices. On the other hand, the selection rate of the true model based on both strategies using the LM tests is smaller than the selection rate of BIC, as can be seen from Table 1, although the empirical levels of all LM tests are close to the nominal level of 5% regardless of the sample size and the spatial weights matrix formation, a result that can also be found in Anselin and Florax (1995) with two independent variables used in the regression analysis. Hence, the BIC, unlike the AIC, will lead the analyst to the right model

selection with confidence slightly larger than any of the LM tests strategy, especially for large sample sizes. The selection rates for all three econometric models based on both information criteria when the true generating model is the SLM are reported on Table 2. As can be seen from this table, the selection rate of the true model by both criteria is not affected by the value and by the sign of the spatial autoregressive parameter ρ , including the extreme case of a negative value smaller than -1 for the queen formation, as well as by the spatial matrix formation. The performance of both criteria is determined mainly by the sample size, i.e., the AIC selects the true model at a rate of 82% and 84% for sample sizes of 100 and 400 observations respectively, while the BIC selects the true model more accurate at rates of 96% and 99%, respectively. Hence, the BIC outperforms the AIC, as in the previous case, in terms of selecting the right spatial econometric model, for every given value of the spatial autoregressive parameter and sample size reaching levels close to certainty. It is also important to indicate that the SIM is not selected at all by both criteria, except for small values of ρ and small sample size at a very low rate. Furthermore, the spatial dependence as well as the right spatial econometric model are also recognized successfully by the LM tests, since one of the two the LM tests is designed for this alternative model structure. In that sense the LM tests using both strategies select the true model, i.e., the SLM, with more confidence than the BIC, reaching very frequently levels of certainty, as can be seen from Table 3, regardless of the matrix formation.⁴

Table 4 reports the selection rates for all three econometric models by both information criteria when the true generating model is the SEM. As can be seen from this table, both criteria select the true model in its equivalent form, i.e., the SDM, more aggressively than the previous case, reaching very frequently levels of 100%. However, the performance of both criteria is strongly affected by the absolute value of the spatial error parameter λ , by the sample size and by the weights matrix formation, unlike the previous case. The true model is selected more frequently by both criteria as the absolute value of λ increases, regardless of the weights matrix formation, reaching the level of certainty for large absolute values of the spatial error parameter. For small absolute values of λ the SDM is not selected with certainty, simply because the spatial effects of a spatial error model are not significant and therefore, they appear in regression analysis as nuisance. This symptom is observed more rigorously for small sample sizes by both criteria and especially by the BIC using the queen formation or the $K=8$ matrix formation where the selected best fitted model is the SIM. Unlike the previous case, the performance of both criteria is strongly affected by the formation of the weights matrices with the rook formation giving better results. In general, the AIC performs better than the BIC for this case, for every given value of the spatial error parameter, sample size and weights matrix formation, except for cases where both criteria select the right model with certainty. In addition, the presence of spatial error dependence is also identified by the LM tests, since one of them is specifically designed for this alternative spatial model

⁴ An abnormal behavior of the LM-EL test for the queen weights matrix formation and for small sample size is observed.

Table 2 Percentage of selections for all three models based on AIC and BIC when the true generating model is the SLM for various values of ρ using 1000 replications

N	IC	Model	ρ										
				-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9	
100	AIC	SIM	<i>Rook</i>	0.00	0.00	0.00	1.70	2.90	0.00	0.00	0.00		
				80.90	82.10	83.00	81.30	80.10	81.40	81.10	81.60		
				19.10	17.90	17.00	17.00	17.00	18.60	18.90	18.40		
	BIC	SIM		0.00	0.00	0.00	8.20	11.00	0.00	0.00	0.00		
				96.70	97.00	96.90	88.80	85.70	95.70	96.10	95.50		
				3.30	3.00	3.10	3.00	3.30	4.30	3.90	4.50		
	AIC	SIM	<i>Queen</i>	0.00	0.00	0.00	0.00	15.00	18.40	0.00	0.00	0.00	
				81.80	84.30	83.50	83.50	68.90	64.70	82.10	82.70	81.80	
				18.20	15.70	16.50	16.50	16.10	16.90	17.90	17.30	18.20	
		BIC	SIM		0.00	0.00	0.00	0.00	39.70	42.00	0.00	0.00	0.00
					96.90	96.10	96.20	96.20	57.80	54.80	95.70	95.80	96.20
					3.10	3.90	3.80	3.80	2.50	3.20	4.30	4.20	3.80
400	AIC	SIM	<i>Rook</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
				86.00	85.50	85.50	84.60	83.00	81.90	83.00	84.00		
				14.00	14.50	14.50	15.40	17.00	18.10	17.00	16.00		
	BIC	SIM		0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00		
				98.20	98.80	98.40	98.00	98.90	98.40	98.40	98.20		
				1.80	1.20	1.60	1.90	1.10	1.60	1.60	1.80		
	AIC	SIM	<i>Queen</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
				82.80	82.70	82.60	82.60	83.70	83.00	84.00	84.50	84.90	
				17.20	17.30	17.40	17.40	16.30	17.00	16.00	15.50	15.10	
		BIC	SIM		0.00	0.00	0.00	0.00	1.80	1.60	0.00	0.00	0.00
					98.10	98.70	98.70	98.70	96.70	97.10	98.80	98.80	98.40
					1.90	1.30	1.30	1.30	1.50	1.30	1.20	1.20	1.60

Table 2 (continued)

N	IC	Model	ρ									
			-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9	
325	AIC	SIM	$K=4$									
		SLM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		SDM	84.30	84.10	84.10	84.80	86.40	85.50	84.70	84.80	84.00	
	BIC	SIM	15.70	15.90	15.90	15.20	13.60	14.50	15.30	15.20	16.00	
		SLM	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	
		SDM	98.00	97.90	98.00	98.20	98.40	98.80	98.30	98.90	98.50	
	AIC	SIM	$K=8$									
			SLM	2.00	2.10	2.00	1.80	1.50	1.20	1.70	1.10	1.50
			SDM	0.00	0.00	0.00	0.00	0.80	0.10	0.00	0.00	0.00
		BIC	SLM	83.60	84.20	84.30	83.60	82.90	84.70	85.00	84.70	84.20
			SDM	16.40	15.80	15.70	16.40	16.30	15.20	15.00	15.30	15.80
			SIM	0.00	0.00	0.00	0.00	7.20	5.20	0.00	0.00	0.00
BIC	SLM	98.40	98.60	98.30	98.30	91.60	93.60	98.90	98.30	98.20		
	SDM	1.60	1.40	1.70	1.70	1.20	1.20	1.10	1.70	1.80		

Bold indicates the correct model

structure. The LM tests will eventually select the true model using both strategies with slightly less certainty than the AIC, also reaching levels of 100%, as can be seen from Table 5.

Table 6 presents the selection rates of both information criteria when the true generating process is the SDM. The objective in this case is to investigate the performance of the information criteria in terms of selecting a more complicated spatial econometric model, which is equivalent to SEM under certain restrictions, an action that cannot be implemented directly by the LM tests, since these tests have not been constructed for this spatial model structure. As can be seen from this table, which reports simulation results for values of ρ and θ equal to 0.2, 0.5 and 0.8, based on 1000 replications, the performance of both information criteria is extremely very good in terms of selecting the true model.⁵ Moreover, none of the two criteria ever selects the SIM, indicating that the spatial dependence is clearly recognized either by the true model or alternatively by the SLM for cases where the values of both parameters are small or just one of them. However, even in the case where the SLM

⁵ Simulation trials have also been conducted for negative values of θ as well as for absolute values of θ greater than 1 using positive values of ρ . The results obtained for negative values of θ are identical to the positive ones, except for small values of θ , where the percentage rates are typically smaller, while the results, obtained for absolute values of θ greater than one, are the same as the largest positive value of θ . Similar results are also obtained in general for negative values of ρ with some minor differences.

Table 3 Percentage of selections for all three models based on LM strategies at the 5% nominal level when the true generating model is the SLM for various values of ρ using 1000 replications

N	Str	Model	ρ										
				-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9	
100	I	SIM	<i>Rook</i>										
				0.00	0.00	0.00	6.10	8.30	0.00	0.00	0.00		
				100.00	100.00	100.00	91.10	89.30	100.00	100.00	100.00		
			SEM	0.00	0.00	0.00	2.80	2.40	0.00	0.00	0.00		
		II	SIM	0.00	0.00	0.00	11.90	14.20	0.00	0.00	0.00		
			SLM	100.00	100.00	100.00	86.60	84.60	100.00	100.00	100.00		
	SEM		0.00	0.00	0.00	1.50	1.20	0.00	0.00	0.00			
	I	SIM	<i>Queen</i>										
				0.00	0.00	0.00	0.00	36.80	30.90	0.00	0.00	0.00	
				100.00	100.00	100.00	100.00	58.70	62.70	99.50	99.90	100.00	
			SEM	0.00	0.00	0.00	0.00	4.50	6.40	0.50	0.10	0.00	
		II	SIM	0.00	0.00	0.00	0.20	47.80	39.80	0.00	0.00	0.00	
SLM			100.00	100.00	100.00	99.80	50.70	56.90	99.50	99.90	100.00		
SEM	0.00		0.00	0.00	0.00	1.50	3.30	0.50	0.10	0.00			
400	I	SIM	<i>Rook</i>										
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
				100.00	100.00	100.00	99.80	100.00	100.00	100.00	100.00		
			SEM	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00		
		II	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			SLM	100.00	100.00	100.00	99.80	100.00	100.00	100.00	100.00		
	SEM		0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00			
	I	SIM	<i>Queen</i>										
				0.00	0.00	0.00	0.00	0.80	0.70	0.00	0.00	0.00	
				100.00	100.00	100.00	100.00	98.60	97.90	99.90	100.00	100.00	
			SEM	0.00	0.00	0.00	0.00	0.60	1.40	0.10	0.00	0.00	
		II	SIM	0.00	0.00	0.00	0.00	2.10	1.20	0.00	0.00	0.00	
SLM			100.00	100.00	100.00	100.00	97.70	97.40	99.90	100.00	100.00		
SEM	0.00		0.00	0.00	0.00	0.20	1.40	0.10	0.00	0.00			

Table 3 (continued)

N	Str	Model	ρ	ρ								
				-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9
325	I	SIM	$K=4$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	100.00	100.00	100.00	100.00	99.90	99.70	100.00	100.00	100.00
			SEM	0.00	0.00	0.00	0.00	0.10	0.30	0.00	0.00	0.00
		II	SIM	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00
			SLM	100.00	100.00	100.00	100.00	99.90	99.70	100.00	100.00	100.00
			SEM	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00
	I	SIM	$K=8$	0.00	0.00	0.00	0.00	4.40	1.20	0.00	0.00	0.00
			SLM	100.00	100.00	100.00	100.00	95.20	95.90	99.50	99.40	100.00
			SEM	0.00	0.00	0.00	0.00	0.40	2.90	0.50	0.60	0.00
		II	SIM	0.00	0.00	0.00	0.00	6.70	3.00	0.00	0.00	0.00
			SLM	100.00	100.00	100.00	100.00	93.20	94.30	99.50	99.40	100.00
			SEM	0.00	0.00	0.00	0.00	0.10	2.70	0.50	0.60	0.00

Bold indicates the correct model

is selected, although it is not the true model, the miss-selection problem is mitigated, since this model contains a spatial lag of the dependent variable. In general, the selection rate of the true model increases as sample size increases as well as the absolute values of both parameters increase reaching levels of certainty very quickly even for small sample sizes regardless of the weights matrix formation. For given sample size and value of θ (ρ) the performance of both criteria increases as the value of ρ (θ) increases. Finally, the AIC has an overall better performance than the BIC, since the true model contains more parameters regardless of the weights matrix formation.

The presence of spatial dependence along with the selection of the true model are also investigated by the LM tests, knowing that these tests are not suitable for this spatial model structure. The effort of this attempt is to examine their behavior for a realistic condition, where the limitation of the two simple spatial econometric models is relaxed, since there is a variety of spatial econometric models that can be considered, and more importantly that the spatial structure will not a priori be given. The good news is that spatial dependence is detected by all LM tests with certainty, with some minor abnormal behaviors, although the results are not reported. However, both strategies select with certainty the SLM, when the values of ρ and θ have the same sign regardless of their magnitude, as can be seen from Table 7, which

Table 4 Percentage of selections for all three models based on AIC and BIC when the true generating model is the SEM for various values of λ using 1000 replications

N	IC	Model	λ										
			-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9		
100	AIC	SIM	<i>Rook</i>	0.00	0.00	1.30	47.40	51.10	3.00	0.00	0.00		
				0.00	0.00	1.50	14.60	13.40	2.40	0.00	0.00		
				100.00	100.00	97.20	38.00	35.50	94.60	100.00	100.00		
		BIC		SIM	0.00	0.00	12.10	83.90	86.10	17.60	0.10	0.00	
				SLM	0.00	0.00	4.50	7.30	6.20	4.20	0.00	0.00	
				SDM	100.00	100.00	83.40	8.80	7.70	78.20	99.90	100.00	
	AIC	SIM	<i>Queen</i>	0.00	0.00	1.00	17.10	60.50	64.70	11.80	0.10	0.00	
				0.00	0.10	0.80	6.70	13.30	15.10	9.70	0.50	0.00	
				100.00	99.90	98.20	76.20	26.20	20.20	78.50	99.40	100.00	
		BIC		SIM	0.00	5.90	12.90	57.90	90.70	90.10	36.90	0.50	0.00
				SLM	0.00	1.00	1.90	4.50	4.60	6.90	12.30	2.10	0.10
				SDM	100.00	93.10	85.20	37.60	4.70	3.00	50.80	97.40	99.90
400	AIC	SIM	<i>Rook</i>	0.00	0.00	0.00	10.80	12.60	0.00	0.00	0.00		
				0.00	0.00	0.00	4.10	7.30	0.00	0.00	0.00		
				100.00	100.00	100.00	85.10	80.10	100.00	100.00	100.00		
		BIC		SIM	0.00	0.00	0.00	59.60	63.40	0.10	0.00	0.00	
				SLM	0.00	0.00	0.00	6.10	6.10	0.00	0.00	0.00	
				SDM	100.00	100.00	100.00	34.30	30.50	99.90	100.00	100.00	
	AIC	SIM	<i>Queen</i>	0.00	0.00	0.00	0.20	32.90	30.50	0.00	0.00	0.00	
				0.00	0.00	0.00	0.20	9.80	13.10	0.00	0.00	0.00	
				100.00	100.00	100.00	99.60	57.30	56.40	100.00	100.00	100.00	
		BIC		SIM	0.00	0.00	0.00	6.30	87.30	80.20	0.50	0.00	0.00
				SLM	0.00	0.00	0.00	0.90	3.50	8.10	0.20	0.00	0.00
				SDM	100.00	100.00	100.00	92.80	9.20	11.70	99.30	100.00	100.00

Table 4 (continued)

N	IC	Model	λ								
			-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9
325	AIC	SIM	$K=4$								
		SLM	0.00	0.00	0.00	0.00	25.40	25.00	0.00	0.00	0.00
		SDM	100.00	100.00	100.00	100.00	67.60	66.30	100.00	100.00	100.00
	BIC	SIM	0.00	0.00	0.00	1.00	80.80	76.20	0.00	0.00	0.00
		SLM	0.00	0.00	0.00	0.00	3.20	5.20	0.00	0.00	0.00
		SDM	100.00	100.00	100.00	99.00	16.00	18.60	100.00	100.00	100.00
	AIC	SIM	$K=8$								
		SLM	0.00	0.00	0.00	1.90	45.20	45.20	0.30	0.00	0.00
		SDM	100.00	100.00	100.00	97.30	44.30	40.70	99.50	100.00	100.00
	BIC	SIM	0.00	0.20	1.00	29.20	92.20	87.60	3.40	0.00	0.00
		SLM	0.00	0.00	0.00	1.90	3.10	5.50	1.70	0.00	0.00
		SDM	100.00	99.80	99.00	68.90	4.70	6.90	94.90	100.00	100.00

Bold indicates the correct model

reports simulation results only for both positive values.⁶ For alternate sign of the two parameters the strategies tend to select the SEM instead of the SLM, especially when the absolute values of ρ and θ are equal (or close) to each other, perhaps due to the common factor, as in this case the SDM is indeed the SEM. Hence, the LM tests should be used with caution simply because they cannot recognize any other type of a spatial econometric model other than the SLM or the SEM, even in the case of the SDM which approximates the SEM, as opposed to the information criteria that can easily identify a more complicated spatial econometric structure. In fact, our results confirm the findings of Elhorst and Halleck Vega (2017) that the classical LM tests or the Robust LM tests are not suitable for the SDM case. In general both the information criteria and the strategies of the LM tests can help the analyst to detect successfully the true simple spatial econometric model.

⁶ Simulation trials have also been conducted for positive values of ρ and absolute values of θ greater than one, as well as for negative values of ρ and for all other values of θ including values greater than one in absolute terms. The results obtained for negative values of ρ are identical to the positive ones, except for small values of θ , where the percentage rates are typically smaller, whereas the results, obtained for absolute values of θ greater than one, are the same as those for the positive value of θ for both positive and negative values of ρ .

Table 5 Percentage of selections for all three models based on LM strategies at the 5% nominal level when the true generating model is the SEM for various values of λ using 1000 replications

N	Str	Model	λ										
				-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9	
100	I	SIM	<i>Rook</i>										
				0.00	0.00	3.00	60.90	66.00	4.90	0.00	0.00		
				0.30	0.70	4.40	7.10	6.50	3.30	0.30	0.00		
	II	SIM		0.00	0.00	5.70	66.20	71.10	9.40	0.00	0.00		
				0.30	0.70	3.50	5.00	5.30	2.30	0.30	0.00		
				99.70	99.30	92.60	32.00	27.50	91.80	99.70	100.00		
	I	SIM	<i>Queen</i>										
				0.00	1.40	3.20	33.10	80.60	73.40	14.50	0.10	0.00	
				0.50	1.80	2.70	4.90	4.60	7.20	7.50	1.00	0.40	
	II	SIM		0.00	6.40	12.80	49.70	86.10	76.10	19.60	0.30	0.00	
			0.50	1.10	1.10	1.70	3.30	5.60	5.30	1.00	0.40		
			99.50	92.50	86.10	48.60	10.60	18.30	75.10	98.70	99.60		
400	I	SIM	<i>Rook</i>										
				0.00	0.00	0.00	13.60	17.40	0.00	0.00	0.00		
				0.00	0.00	0.00	4.40	5.30	0.00	0.00	0.00		
	II	SIM		0.00	0.00	0.00	22.10	26.50	0.00	0.00	0.00		
				0.00	0.00	0.00	2.30	2.50	0.00	0.00	0.00		
				100.00	100.00	100.00	75.60	71.00	100.00	100.00	100.00		
	I	SIM	<i>Queen</i>										
				0.00	0.00	0.00	0.30	46.80	38.10	0.10	0.00	0.00	
				0.00	0.00	0.00	0.90	6.00	6.60	0.10	0.00	0.00	
	II	SIM		0.00	0.00	0.00	1.00	58.60	47.30	0.00	0.00	0.00	
			0.00	0.00	0.00	0.80	3.00	3.50	0.10	0.00	0.00		
			100.00	100.00	100.00	98.20	38.40	49.20	99.90	100.00	100.00		

Table 5 (continued)

N	Str	Model	λ										
				-1.4	-0.9	-0.8	-0.5	-0.2	0.2	0.5	0.8	0.9	
325	I	SIM	$K=4$										
			0.00	0.00	0.00	0.00	35.50	27.70	0.00	0.00	0.00		
			SLM	0.00	0.00	0.00	0.10	4.50	5.90	0.10	0.00	0.00	
			SEM	100.00	100.00	100.00	99.90	60.00	66.40	99.90	100.00	100.00	
	II	SIM	0.00	0.00	0.00	0.10	44.10	36.30	0.00	0.00	0.00		
			SLM	0.00	0.00	0.00	0.10	2.40	3.10	0.10	0.00	0.00	
			SEM	100.00	100.00	100.00	99.80	53.50	60.60	99.90	100.00	100.00	
	I	SIM	$K=8$										
			0.00	0.00	0.00	5.90	65.00	53.80	0.20	0.00	0.00		
			SLM	0.00	0.10	0.30	2.60	5.50	6.40	0.20	0.00	0.00	
			SEM	100.00	99.90	99.70	91.50	29.50	39.80	99.60	100.00	100.00	
	II	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
SLM			0.00	0.00	0.00	0.10	2.40	3.10	0.10	0.00	0.00		
SEM			100.00	100.00	100.00	99.80	53.50	60.60	99.90	100.00	100.00		

Bold indicates the correct model

4 Further simulation results

The Monte Carlo analysis is extended by considering several other spatial econometric models as an effort to eliminate the restrictions imposed by the LM tests and to evaluate the performance of the information criteria as far as selecting the right model among a larger variety of spatial econometric models. The models are:

- (a) the Spatial AutoRegressive AutoRegressive (SARAR) model, also known as Spatial Autoregressive Combined Model (SACM):

$$y = \rho W y + X \beta + \epsilon$$

with

$$\epsilon = \lambda W \epsilon + u$$

- (b) the Spatial Lag X (SLX) model:

$$y = X \beta + W X \theta + \epsilon$$

and c) the Spatial Durbin Error Model (SDEM):

$$y = X \beta + W X \theta + \epsilon$$

Table 6 Percentage of selections for all three models based on AIC and BIC when the true generating model is the SDM for various values of ρ and θ using 1000 replications

N	IC	ρ	0.2			0.5			0.8			
			Model / θ			0.2	0.5	0.8	0.2	0.5	0.8	
100	AIC	SIM	<i>Rook</i>									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			41.20	0.10	0.00	29.30	0.00	0.00	15.60	0.00	0.00	
			SDM	58.80	99.90	100	70.70	100	100	84.40	100	100
	BIC	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			67.10	1.30	0.00	55.40	0.30	0.00	36.10	0.00	0.00	
			32.90	98.70	100	44.60	99.70	100	63.90	100	100	
	AIC	SIM	<i>Queen</i>									
			0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			57.10	10.70	0.00	53.40	5.10	0.00	41.70	0.90	0.00	
			SDM	42.70	89.30	100	46.60	94.90	100	58.30	99.10	100
	BIC	SIM	1.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
81.30			30.60	1.70	79.50	20.20	0.40	71.50	5.80	0.00		
17.00			69.40	98.30	20.50	79.80	99.60	28.50	94.20	100		
400	AIC	SIM	<i>Rook</i>									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			6.30	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	
			SDM	93.70	100	100	100	100	100	99.90	100	100
	BIC	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			30.80	0.00	0.00	0.00	0.00	0.00	2.80	0.00	0.00	
			69.20	100	100	100	100	100	97.20	100	100	
	AIC	SIM	<i>Queen</i>									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			25.80	0.10	0.00	17.00	0.10	0.00	6.00	0.00	0.00	
			SDM	74.20	99.90	100	83.00	99.90	100	94.00	100	100
	BIC	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
67.10			0.80	0.00	56.90	0.20	0.00	33.00	0.00	0.00		
32.90			99.20	100	43.10	99.80	100	67.00	100	100		

Table 6 (continued)

N	IC	ρ	0.2			0.5			0.8			
			Model / θ			0.2	0.5	0.8	0.2	0.5	0.8	
325	AIC	SIM	$K = 4$									
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	16.20	0.00	0.00	8.20	0.00	0.00	2.00	0.00	0.00
		SDM	83.80	100	100	91.80	100	100	98.00	100	100	
	BIC	SIM		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			SLM	51.30	0.00	0.00	33.20	0.00	0.00	10.40	0.00	0.00
			SDM	48.70	100	100	66.80	100	100	89.60	100	100
	AIC	SIM	$K = 8$									
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	36.70	0.30	0.00	28.00	0.00	0.00	12.20	0.00	0.00
			SDM	63.30	99.70	100	72.00	100	100	87.80	100	100
		BIC	SIM		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SLM				77.40	4.10	0.00	70.50	0.70	0.00	48.60	0.00	0.00
SDM	22.60			95.90	100	29.50	99.30	100	51.40	100	100	

Bold indicates the correct model

with

$$\epsilon = \lambda W\epsilon + u$$

and they have been generated as:

(a) the SARAR model as:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} (I - \lambda W)^{-1} u$$

(b) the SLX model as:

$$y = X\beta + WX\theta + \epsilon$$

(c) the SDEM as:

$$y = X\beta + WX\theta + (I - \lambda W)^{-1} u$$

using the same simulation process.

Table 8 presents the percentage selection rates of both criteria when the true generating model is the SIM based on several other spatial econometric models. As can be seen from this table, the performance of both criteria remained almost unchanged as that obtained with fewer spatial econometric models and reported

Table 7 Percentage of selections for all three models based on LM strategies at the 5% nominal level when the true generating model is the SDM for various values of ρ and θ using 1000 replications

N	Str	ρ	0.2			0.5			0.8			
			Model / θ	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
100	I	SIM	<i>Rook</i>									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			100	100	100	100	100	100	100	100	100	
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	II	SIM	0.00									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			100	100	100	100	100	100	100	100	100	
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	400	I	SIM	<i>Queen</i>								
				1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				98.60	100	100	100	100	100	100	100	100
				SEM	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
II		SIM	0.90									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			99.00	100	100	100	100	100	100	100	100	
			SEM	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
400		I	SIM	<i>Rook</i>								
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				100	100	100	100	100	100	100	100	100
				SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	II	SIM	0.00									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
			100	100	100	100	100	100	100	100	100	
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	400	I	SIM	<i>Queen</i>								
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				100	100	100	100	100	100	100	100	100
				SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
II		SIM	0.00									
			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	100		100	100	100	100	100	100	100	100		
		SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

Table 7 (continued)

N	Str	ρ	0.2			0.5			0.8						
			Model / θ			0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	
325	I	SIM	$K = 4$			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	100	100	100	100	100	100	100	100	100	100	100	100
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		II	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	100	100	100	100	100	100	100	100	100	100	100	100
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	I	SIM	$K = 8$			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	100	100	100	100	100	100	100	100	100	100	100	100
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		II	SIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	100	100	100	100	100	100	100	100	100	100	100	100
			SEM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 8 Percentage of selections for all six models based on AIC and BIC when the true generating model is the SIM using 1000 replications

IC	Model	N	100		400		325	
			Rook	Queen	Rook	Queen	K=4	K=8
AIC	SIM		72.00	72.30	75.70	75.30	77.80	74.60
	SLM		9.90	11.00	9.30	9.00	7.10	10.20
	SDM		3.00	3.30	2.30	2.00	1.60	2.70
	SARAR		2.40	2.00	2.20	2.50	2.50	1.90
	SLX		10.30	8.30	8.70	9.00	9.40	8.80
	SDEM		2.40	3.10	1.80	2.20	1.60	1.80
BIC	SIM		94.10	94.00	97.50	97.10	97.80	96.90
	SLM		3.10	3.30	1.10	1.30	0.50	1.40
	SDM		0.10	0.30	0.00	0.10	0.00	0.00
	SARAR		0.10	0.00	0.20	0.10	0.00	0.00
	SLX		2.50	2.00	1.10	1.30	1.60	1.70
	SDEM		0.10	0.40	0.10	0.10	0.10	0.00

Bold indicates the correct model

on Table 1. However, the selection rates are slightly smaller for both criteria since more models are considered with the BIC outperforming the AIC, but still having very high rates for selecting the true model.

Table 9 Percentage of selections for all six models based on AIC and BIC when the true generating model is the SLM for various values of ρ using 1000 replications

N	IC	Model	ρ								
			0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	
100	AIC	SIM	Rook					Queen			
				2.80	0.00	0.00	0.00	15.10	0.00	0.00	0.00
			SLM	59.90	69.10	69.20	68.50	45.40	68.50	70.40	69.00
			SDM	3.40	11.50	16.20	16.60	2.10	6.40	13.70	15.60
			SARAR	4.30	10.80	14.50	14.90	3.60	11.00	14.50	15.20
			SLX	24.70	0.10	0.00	0.00	30.30	4.40	0.00	0.00
		SDEM	4.90	8.50	0.10	0.00	3.50	9.70	1.40	0.20	
	BIC	SIM		9.50	0.00	0.00	0.00	37.40	0.00	0.00	0.00
			SLM	65.00	92.10	93.20	92.50	38.40	86.70	92.50	92.80
			SDM	0.60	2.20	3.70	4.40	0.10	0.50	3.40	3.40
			SARAR	0.30	2.10	3.10	3.10	0.40	2.50	3.50	3.70
			SLX	24.00	0.90	0.00	0.00	23.40	8.10	0.00	0.00
		SDEM	0.60	2.70	0.00	0.00	0.30	2.20	0.60	0.10	
400	AIC	SIM	Rook					Queen			
				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			SLM	74.40	75.30	72.80	72.00	73.00	74.10	74.40	73.50
			SDM	4.90	13.20	14.10	13.60	3.50	10.50	12.30	13.20
			SARAR	6.10	10.30	13.10	14.40	3.50	9.90	13.30	13.30
			SLX	4.00	0.00	0.00	0.00	15.20	0.00	0.00	0.00
		SDEM	10.60	1.20	0.00	0.00	4.80	5.50	0.00	0.00	
	BIC	SIM		0.00	0.00	0.00	0.00	1.50	0.00	0.00	0.00
			SLM	91.80	96.70	96.90	96.90	81.20	96.90	97.50	97.20
			SDM	0.20	1.20	1.50	1.70	0.70	1.00	0.90	1.50
			SARAR	0.50	1.40	1.60	1.40	0.20	1.00	1.60	1.30
			SLX	7.10	0.00	0.00	0.00	16.30	0.00	0.00	0.00
		SDEM	0.40	0.70	0.00	0.00	0.10	1.10	0.00	0.00	
325	AIC	SIM	K=4					K=8			
				0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
			SLM	75.00	77.20	74.40	72.80	68.30	73.90	73.00	71.60
			SDM	4.40	11.40	12.60	14.00	2.50	10.10	13.20	13.90
			SARAR	5.20	10.50	13.00	13.20	3.10	12.00	13.80	14.50
			SLX	8.40	0.00	0.00	0.00	22.40	0.00	0.00	0.00
		SDEM	7.00	0.90	0.00	0.00	3.60	4.00	0.00	0.00	
	BIC	SIM		0.00	0.00	0.00	0.00	3.90	0.00	0.00	0.00
			SLM	88.80	97.20	97.90	97.70	72.40	97.30	96.80	96.50
			SDM	0.40	1.20	1.00	1.40	0.20	0.70	1.50	1.80
			SARAR	0.00	1.20	1.10	0.90	0.20	0.90	1.70	1.70
			SLX	10.40	0.00	0.00	0.00	23.00	0.30	0.00	0.00
		SDEM	0.40	0.40	0.00	0.00	0.30	0.80	0.00	0.00	

Bold indicates the correct model

Things changed when the true model is the SLM especially for the AIC, as can be seen from Table 9, relatively to results reported on Table 2, where now the presence of the SARAR model forces the AIC to select less frequently the true model by a reduction rate approximately 10%.⁷ The BIC, on the other hand, kept the same selection rate of the true model, slightly smaller than before, due to the presence of more spatial econometric models, but still very high at the 95% area, regardless of the sample size and the weights matrix formation. Contrary, the results became very complicated when the true model is the SEM, as can be seen from Tables 4 and 10. Both criteria behave very similarly by selecting around 40% the true model, instead of 100% that was before, 32% the SARAR model and 28% the SDEM for moderate and large values of the parameter, whereas for small values both criteria select the SIM failing to detect the presence of spatial dependence. Hence, the existence of more spatial econometric models damaged the good image of both criteria obtained with fewer models when the true spatial econometric model is the SEM. Lastly, Table 11 reports the selection rates of both criteria for the SLX model among all six spatial econometric models, whereas Table 12 reports the section rates only for the generated models SDM, SARAR and SDEM that have two parameters. For the SLX case the selection rates of the true model increase as the value of the parameter increases and/or the sample size with the BIC outperforming AIC and reporting very high rates. Similarly, for the other models the selection rates increase as the values of both parameters increase and/or as the sample size increases regardless of the weights matrix formation reaching levels of certainty.⁸

The simulation process is extended by considering two non-ideal situations, such as the case of a non-constant variance of the error term and non-normality, as in Mur and Angulo (2009), although these conditions are not taken into account in the calculation of both criteria and therefore they are not expected to improve their behavior. The heteroscedasticity is incorporated into the model by assuming that the error term is defined as:

$$\varepsilon \sim N(0, \sigma^2 d_r \mathbf{I})$$

where d_r is the distance of the centroid of each cell r from the centroid of the upper left cell of the grid. Of course, there are several other ways of modeling heteroscedasticity, as for example considering d_r to be the distance of the centroid of each cell r from the centroid of the central or the lower left cell of the grid (see for example Mur and Angulo, 2009), including the possibility of employing Bayesian analysis even for the case of selecting spatial econometric models under different weights matrix formations (see for example Lesage and Parent, 2007; Crespo Cuaresma and Feldkircher, 2013, Debarsy and Lesage, 2022 and Fernandez et al., 2001). However, the purpose of this research is focused on the influence of heteroscedasticity in the

⁷ Note that only results for positive values of the parameters are reported since the negative values gave the same results.

⁸ The simulation results for the real geographical structure of Greece are not reported for spacing issues and simply because they are very similar with the results of Table 12 obtained by the other two weights matrices formation.

Table 10 Percentage of selections for all six models based on AIC and BIC when the true generating model is the SEM for various values of λ using 1000 replications

N	IC	Model	λ								
			0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	
100	AIC	SIM	Rook					Queen			
			48.70	2.80	0.00	0.00	60.90	11.20	0.10	0.00	
		SLM	10.90	2.10	0.00	0.00	12.00	8.70	0.50	0.00	
		SDM	14.20	43.50	45.10	45.60	8.80	35.70	46.00	45.70	
		SARAR	10.00	29.20	31.60	32.10	4.50	23.10	29.00	30.70	
		SLX	6.20	0.10	0.00	0.00	8.20	2.70	0.00	0.00	
	SDEM	10.00	22.30	23.30	22.30	5.60	18.60	24.40	23.60		
	BIC	SIM	84.40	16.90	0.10	0.00	88.20	35.10	0.50	0.00	
		SLM	5.40	4.20	0.00	0.00	5.80	11.40	1.90	0.10	
		SDM	3.10	34.80	45.00	45.60	1.90	22.30	44.60	45.60	
		SARAR	1.90	24.10	31.60	32.10	0.60	16.30	28.70	30.70	
		SLX	2.40	0.20	0.00	0.00	2.80	2.40	0.00	0.00	
SDEM		2.80	19.80	23.30	22.30	0.70	12.50	24.30	23.60		
400	AIC	SIM	Rook					Queen			
			12.30	0.00	0.00	0.00	28.80	0.00	0.00	0.00	
		SLM	6.50	0.00	0.00	0.00	11.70	0.00	0.00	0.00	
		SDM	34.20	42.30	40.30	39.80	23.00	41.00	40.00	39.90	
		SARAR	25.70	32.10	32.40	32.00	19.10	33.10	32.50	32.00	
		SLX	0.60	0.00	0.00	0.00	3.20	0.00	0.00	0.00	
	SDEM	20.70	25.60	27.30	28.20	14.20	25.90	27.50	28.10		
	BIC	SIM	62.90	0.10	0.00	0.00	79.40	0.50	0.00	0.00	
		SLM	5.50	0.00	0.00	0.00	7.80	0.20	0.00	0.00	
		SDM	13.00	42.20	40.30	39.80	4.30	40.70	40.00	39.90	
		SARAR	10.20	32.10	32.40	32.00	4.20	33.00	32.50	32.00	
		SLX	0.30	0.00	0.00	0.00	0.90	0.00	0.00	0.00	
SDEM		8.10	25.60	27.30	28.20	3.40	25.60	27.50	28.10		
325	AIC	SIM	K=4					K=8			
			23.20	0.00	0.00	0.00	42.90	0.30	0.00	0.00	
		SLM	7.60	0.00	0.00	0.00	11.70	0.10	0.00	0.00	
		SDM	25.40	40.80	39.60	39.10	16.10	39.80	40.10	40.90	
		SARAR	23.40	33.80	33.70	33.50	11.30	30.80	31.10	31.10	
		SLX	2.30	0.00	0.00	0.00	6.10	0.00	0.00	0.00	
	SDEM	18.10	25.40	26.70	27.40	11.90	29.00	28.80	28.00		
	BIC	SIM	75.50	0.00	0.00	0.00	86.60	3.20	0.00	0.00	
		SLM	5.20	0.00	0.00	0.00	5.30	1.60	0.00	0.00	
		SDM	5.90	40.80	39.60	39.10	3.00	37.90	40.10	40.90	
		SARAR	7.00	33.80	33.70	33.50	2.00	29.30	31.10	31.10	
		SLX	0.30	0.00	0.00	0.00	1.30	0.00	0.00	0.00	
SDEM		6.10	25.40	26.70	27.40	1.80	28.00	28.80	28.00		

Bold indicates the correct model

Table 11 Percentage of selections for all six models based on AIC and BIC when the true generating model is the SLX for various values of θ using 1000 replications

N	IC	Model	θ										
			0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9			
100	AIC	SIM	Rook					Queen					
			3.90	0.00	0.00	0.00	22.10	0.00	0.00	0.00			
		SLM	22.30	0.50	0.00	0.00	18.70	5.90	0.10	0.00			
		SDM	5.50	10.40	13.30	13.90	5.60	7.50	10.40	11.60			
		SARAR	3.90	3.50	0.00	0.00	3.60	7.50	4.50	2.60			
		SLX	60.10	74.20	74.00	73.40	44.30	69.50	70.80	70.90			
		SDEM	4.30	11.40	12.70	12.70	5.70	9.60	14.20	14.90			
	BIC	SIM	15.50	0.00	0.00	0.00	52.30	0.10	0.00	0.00			
		SLM	19.20	1.60	0.00	0.00	12.00	7.80	1.00	0.20			
		SDM	1.00	2.30	3.10	3.30	0.90	2.40	1.80	2.20			
		SARAR	0.90	1.10	0.00	0.00	0.80	1.70	1.60	1.30			
		SLX	62.70	92.60	94.20	94.00	32.90	85.60	91.90	92.40			
		SDEM	0.70	2.40	2.70	2.70	1.10	2.40	3.70	3.90			
		400	AIC	SIM	Rook					Queen			
0.00	0.00				0.00	0.00	0.30	0.00	0.00	0.00			
SLM	6.00			0.00	0.00	0.00	12.90	0.20	0.00	0.00			
SDM	5.10			11.00	13.20	13.40	5.60	9.50	12.10	11.70			
SARAR	8.90			1.00	0.00	0.00	5.70	5.70	0.10	0.00			
SLX	74.10			77.10	74.80	74.50	72.10	75.00	76.20	76.20			
SDEM	5.90			10.90	12.00	12.10	3.40	9.60	11.60	12.10			
BIC	SIM		0.00	0.00	0.00	0.00	5.70	0.00	0.00	0.00			
	SLM		7.50	0.00	0.00	0.00	12.70	0.50	0.00	0.00			
	SDM		0.20	0.90	1.50	1.70	0.20	0.40	0.90	0.80			
	SARAR		0.30	0.20	0.00	0.00	0.30	0.80	0.10	0.00			
	SLX		91.80	98.00	97.60	97.40	80.80	97.70	98.10	98.20			
	SDEM		0.20	0.90	0.90	0.90	0.30	0.60	0.90	1.00			
	325		AIC	SIM	K=4					K=8			
0.00		0.00			0.00	0.00	1.20	0.00	0.00	0.00			
SLM		9.20		0.00	0.00	0.00	16.40	0.70	0.00	0.00			
SDM		5.20		9.30	10.50	11.00	5.30	8.40	10.30	11.40			
SARAR		7.50		2.80	0.00	0.00	6.90	7.70	0.60	0.20			
SLX		75.30		78.80	78.60	77.90	67.70	73.50	77.50	76.50			
SDEM		2.80		9.10	10.90	11.10	2.50	9.70	11.60	11.90			
BIC		SIM	0.30	0.00	0.00	0.00	10.60	0.00	0.00	0.00			
		SLM	9.80	0.10	0.00	0.00	14.60	1.80	0.00	0.00			
		SDM	0.80	1.30	1.30	1.20	0.70	0.70	1.00	0.90			
		SARAR	0.60	0.30	0.00	0.00	0.50	1.30	0.30	0.10			
		SLX	88.40	97.60	97.70	97.80	73.40	95.20	97.30	97.50			
		SDEM	0.10	0.70	1.00	1.00	0.20	1.00	1.40	1.50			

Bold indicates the correct model

Table 12 Percentage of selections only for the generated models SDM, SARAR and SDEM based on AIC and BIC for various values of ρ , λ and θ using 1000 replications

N	IC	True Mo del	ρ, λ														
			0.5			0.8			0.9								
			λ, θ														
			0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	0.9		
100	AIC	SDM	<i>Rook</i>														
			6.90	49.70	66.00	71.90	58.00	98.20	99.80	99.90	80.50	100	100	88.30	100	100	
		SARAR	18.50	54.50	69.80	75.60	29.80	77.60	93.20	96.20	41.40	93.70	99.80	99.90	43.20	100	100
		SDEM	7.10	34.60	38.50	39.20	37.70	75.90	89.50	90.90	41.20	82.90	98.80	98.90	42.60	84.10	99.20
	BIC	SDM	0.00	28.10	48.60	53.70	37.10	97.90	99.80	99.90	62.60	100	100	100	72.40	100	100
		SARAR	9.10	49.90	69.80	75.60	14.50	69.60	93.00	96.20	20.20	84.20	99.70	99.90	22.00	86.70	99.90
		SDEM	0.70	15.90	21.50	21.90	25.40	55.80	85.10	86.40	41.00	78.00	98.70	98.90	42.60	83.10	99.20
	AIC	SDM	<i>Queen</i>														
			0.90	15.10	32.20	35.70	27.40	79.90	95.00	96.90	51.40	97.40	100	100	61.40	98.70	100
		SARAR	12.10	39.80	57.70	63.80	17.40	55.60	82.30	90.10	25.30	75.20	97.80	99.60	28.70	82.00	99.50
		SDEM	4.00	10.70	23.40	24.40	21.10	32.70	61.90	68.40	28.60	52.40	75.60	83.10	27.90	53.30	76.50
	BIC	SDM	0.00	1.10	14.00	17.90	7.80	64.40	93.80	96.60	26.60	93.20	100	100	35.00	97.10	100
		SARAR	4.60	31.50	56.90	63.70	6.90	43.30	80.60	89.70	10.20	58.60	96.00	99.20	12.30	65.00	98.10
		SDEM	0.20	0.90	6.80	8.60	10.60	8.20	36.20	49.70	28.20	45.30	61.20	69.40	27.90	52.10	78.30

Table 12 (continued)

N	IC	True Mo del	ρ, λ				0.5				0.8				0.9					
			λ, θ	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	
400	AIC	SDM	<i>Rook</i>																	
			50.10	87.70	96.20	97.60	92.00	100	100	100	99.20	100	100	100	100	100	100	100	100	100
			33.40	64.80	82.40	88.10	72.60	97.00	99.90	99.90	84.50	100	100	100	100	100	87.00	100	100	100
			25.20	72.70	81.00	82.50	58.60	96.40	99.60	99.70	66.30	98.60	100	100	100	73.30	99.50	100	100	100
			26.20	79.40	93.70	96.00	79.90	100	100	100	96.80	100	100	100	100	99.30	100	100	100	100
	BIC	SDM	22.00	64.80	82.40	88.10	47.00	97.00	99.90	99.90	54.70	99.90	100	100	58.10	100	100	100	100	
			1.20	53.50	59.80	60.70	57.80	92.40	99.60	99.70	66.30	98.60	100	100	73.30	99.50	100	100	100	
			<i>Queen</i>																	
			20.10	61.20	77.90	82.60	67.50	99.30	100	100	90.70	100	100	100	96.00	100	100	100	100	100
			23.70	52.50	71.30	77.80	45.20	84.90	96.50	99.00	60.40	99.10	100	100	65.40	99.70	100	100	100	100
	BIC	SDM	7.50	48.20	55.60	56.80	43.90	72.90	93.00	94.00	44.10	79.70	98.10	99.30	48.10	84.00	98.10	99.40	99.40	
			0.60	38.20	56.20	62.60	37.40	99.20	100	100	66.10	100	100	100	79.00	100	100	100	100	
			11.00	52.40	71.30	77.80	19.80	84.90	96.50	99.00	27.10	99.00	100	100	30.80	99.40	100	100	100	
			0.40	18.80	30.30	31.20	38.40	38.50	92.30	93.90	44.10	79.70	97.20	98.80	48.10	84.00	98.10	99.20	99.20	

Table 13 Percentage of selections only for the generated models SIM, SLM, SEM and SLX under heteroskedasticity based on AIC and BIC for various values of ρ , λ and θ using 1000 replications

N	IC	True model	ρ, λ, θ					
			0.0	0.2	0.5	0.8	0.9	
100	AIC	SIM	<i>Rook</i>					
			61.10					
			SLM	37.00	70.40	59.70	55.20	
		SEM (SDM)	5.00	11.30	14.60	15.40		
		SLX	11.70	15.50	21.80	24.60		
		BIC	SIM	88.00				
				SLM	26.00	86.40	90.70	87.70
				SEM (SDM)	0.50	3.60	6.00	7.30
		SLX	3.20 4.80 7.60 9.30					
	AIC		SIM	<i>Queen</i>				
				57.80				
		SLM		27.50	63.10	60.90	55.40	
		SEM (SDM)	4.20	11.10	16.30	16.90		
		SLX	12.50	13.00	15.30	16.30		
		BIC	SIM	92.80				
				SLM	16.90	69.10	90.40	88.00
				SEM (SDM)	0.30	1.30	5.40	6.50
		SLX	3.20 3.90 5.60 6.40					
400	AIC		SIM	<i>Rook</i>				
				58.50				
		SLM		64.70	66.00	57.90	56.20	
		SEM (SDM)	8.10	10.90	12.60	13.00		
		SLX	13.30	15.00	20.00	22.20		
		BIC	SIM	88.70				
				SLM	61.50	96.00	94.10	93.00
				SEM (SDM)	0.60	1.90	3.20	3.50
		SLX	2.00 3.20 5.40 6.10					
	AIC		SIM	<i>Queen</i>				
				60.10				
		SLM		53.40	73.10	61.20	59.10	
		SEM (SDM)	5.50	9.80	11.50	12.90		
		SLX	10.30	12.10	14.60	15.80		
		BIC	SIM	92.70				
				SLM	37.30	95.20	94.40	93.00
				SEM (SDM)	0.10	1.40	1.80	2.10
		SLX	1.50 2.20 2.80 3.20					

model selection process by AIC and BIC and not on the form of its appearance and that is why only one case is considered.

Indeed, as can be seen from Table 13, which reports the selection rates only for the generated models with one parameter, the behavior of both criteria has changed,

either partially or significantly, under heteroscedasticity.⁹ For the SIM and the SLM cases, the selection rate has decreased by a small percentage using the BIC and by a large percentage, close to 15%, using the AIC, while for the SEM and the SLX cases both criteria fail to detect the true model. In fact, the best-fitted model for the SEM case is the SLM having the highest selection rate, although it is not reported, whereas, respectively, for the SLM model is the SIM, where in this case both criteria fail to detect spatial dependence under heteroscedasticity. The selection failure of both criteria appears also for the SDM and SDEM under heteroscedasticity, unlike the SARAR model which is selected even with certainty in some cases, as can be seen from the Table 14, which reports the selection rates only for the generated model. Although it is not reported, both criteria prefer to select either the SLM or the SARAR model when the true generated model is either the SDM or SDEM.

Normality, on the other hand, seems not to be a very important issue since the selection behavior of both criteria did not change drastically, as can be seen from Tables 15 and 16 that report results for models with one and two parameters respectively, assuming that the errors are generated from a log-normal distribution. Indeed, the selection rates are slightly smaller than those obtained under normality and identical under student t distribution, although these results are not reported.

5 Concluding remarks

The identification of a model that most spatially expresses the spatial dependence on a given data set is a matter of paramount importance for every researcher in any spatial econometric analysis. The models that are typically considered to express spatial dependence in empirical applications are usually the SLM and the SEM, due to the fact that these two models can be identified by the existing spatial dependence tests, while other more general or more complicated spatial models that could better express the form of spatial dependence are less frequently considered. For this reason, this study examines the behavior of the two most important information criteria, i.e., the AIC and the BIC, in terms of detecting the true spatial econometric model not only in lieu with the existing LM tests for the two simple spatial models but also for more spatial econometric models that their existence cannot be identified by these tests. Simulation results show that these criteria contribute significantly and more effectively than the existing LM tests to the selection process of the true spatial econometric model for the simple two models the SLM and the SEM.

For the case of more spatial econometric models such as the SLX, SDM, SARAR and SDEM that have been considered additionally in this simulation process the results are mixed. For most of the spatial econometric models the BIC had incredible selection behavior with the AIC having typically, but not always, smaller selection rates of the true model. Exception of this statement is the SEM where both

⁹ Note that the simulation results for the real geographical structure of Greece are not included due to many instability problems that arose during the simulation process, while in general, the behavior of both criteria was similar as in the cases of the other two weights matrices formation.

Table 14 Percentage of selections only for the generated models SDM, SARAR, and SDEM under heteroskedasticity based on AIC and BIC for various values of ρ , λ and θ using 1000 replications

N	IC	True Model	ρ, λ						0.8										
			0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9					
100	AIC	SDM	<i>Rook</i>																
			4.40	5.00	6.80	7.90	10.20	13.60	19.10	20.90	12.20	16.70	23.20	25.80	12.20	17.70	24.00	26.90	
			8.50	19.60	44.40	52.70	16.60	49.00	84.30	92.20	42.10	83.40	99.10	100	48.90	91.70	100	100	
	BIC	SDM	<i>Rook</i>																
			3.90	4.00	5.70	6.30	6.10	7.40	9.90	11.10	8.10	7.60	10.30	11.50	9.40	8.80	10.80	12.40	
			0.50	0.50	1.00	1.40	2.20	3.60	5.80	7.40	3.30	5.40	8.40	10.00	3.50	5.30	9.30	10.50	
	AIC	SDM	<i>Queen</i>																
			1.90	2.50	18.00	30.20	1.90	17.20	62.00	82.10	14.80	60.70	97.20	99.60	26.40	80.70	99.60	100	
			0.60	0.50	0.80	1.00	1.80	1.30	1.90	2.80	2.60	1.30	2.20	2.50	2.30	1.70	2.10	2.70	
	100	AIC	SDM	<i>Rook</i>															
				3.20	2.90	3.20	3.30	9.10	9.10	11.10	11.50	13.10	15.00	18.40	19.30	14.00	16.70	19.20	20.40
				5.70	8.10	28.90	39.70	8.20	25.30	60.30	76.80	26.80	59.70	94.20	98.60	37.40	75.50	98.60	99.50
BIC		SDM	<i>Rook</i>																
			3.10	3.80	4.20	4.40	8.30	8.60	9.00	9.30	10.40	10.30	11.20	11.60	11.50	11.10	11.80	11.50	
			0.10	0.20	0.30	0.30	1.20	1.60	2.40	2.60	3.30	4.80	6.50	6.90	4.00	5.60	7.30	7.90	
BIC	SDM	<i>Queen</i>																	
		1.60	1.20	7.00	15.60	1.50	4.60	37.10	58.00	6.60	36.40	86.30	97.40	14.30	57.30	97.40	99.40		
		0.40	0.50	0.80	0.90	1.90	1.50	1.50	1.70	3.30	2.90	2.90	2.90	3.80	3.10	3.00	3.10		

Table 14 (continued)

N	IC	True Model	ρ, λ						0.8						0.9					
			0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9		
400	AIC	SDM	<i>Rook</i>																	
			7.50	9.10	12.20	13.00	10.50	13.10	16.90	18.30	11.40	14.40	18.40	20.60	11.70	15.00	19.80	21.30		
		SARAR	7.90	35.40	62.70	77.30	34.40	82.30	99.60	100	60.20	99.40	100	100	74.40	100	100	100		
		SDEM	7.10	8.40	11.10	12.60	8.00	8.80	11.20	12.90	8.90	9.90	12.40	12.90	10.20	9.50	13.30	13.80		
	BIC	SDM	0.80	1.10	1.70	2.60	1.70	2.70	3.90	5.60	1.80	3.20	5.10	6.10	2.00	3.40	5.30	6.30		
		SARAR	1.00	2.90	30.20	45.50	2.40	48.70	96.30	99.70	28.10	96.00	100	100	42.10	99.70	100	100		
		SDEM	0.60	1.30	2.00	2.20	0.80	1.00	1.90	2.20	1.40	0.90	1.60	2.10	1.40	0.90	1.40	1.90		
	AIC	SDM	<i>Queen</i>																	
			5.50	5.90	7.40	8.20	10.50	11.70	14.10	14.40	11.00	12.80	15.50	17.00	11.10	13.00	15.90	17.80		
		SARAR	9.40	18.40	45.30	57.20	17.90	54.80	92.50	98.50	43.90	92.00	100	100	57.30	98.40	100	100		
		SDEM	3.50	4.10	5.10	5.60	6.50	6.10	6.90	7.00	7.70	7.40	6.90	7.00	7.20	7.40	7.50	7.60		
	BIC	SDM	0.10	0.40	0.80	0.90	1.10	1.50	2.30	2.90	1.50	2.10	3.00	3.40	1.70	1.90	3.20	3.50		
		SARAR	1.50	0.90	13.20	25.20	0.90	15.00	74.60	92.20	12.90	74.00	99.90	100	23.60	91.80	100	100		
		SDEM	0.50	0.50	0.30	0.40	1.10	1.10	1.30	1.10	1.50	1.10	1.20	1.30	1.40	1.20	1.10	1.10		

Table 15 Percentage of selections only for the generated models SIM, SLM, SEM and SLX under Log-normal errors based on AIC and BIC for various values of ρ, λ and θ using 1000 replications

N	IC	True Model	ρ, λ, θ					
			0.0	0.2	0.5	0.8	0.9	
100	AIC	SIM	<i>Rook</i>					
			71.70					
			SLM	50.10	68.80	68.40	68.70	
		SEM (SDM)	8.80	34.30	39.30	38.40		
		SLX	42.70	73.10	77.30	77.80		
		BIC	SIM	93.60				
				SLM	37.80	92.20	91.50	91.70
				SEM (SDM)	2.00	25.40	38.50	38.30
		AIC	SIM	<i>Queen</i>				
	74.40							
	SLM			34.40	74.60	71.70	70.30	
	SEM (SDM)		4.90	26.90	37.20	38.00		
	SLX		25.80	62.10	72.70	73.50		
	BIC		SIM	93.70				
				SLM	22.50	87.00	93.50	93.00
				SEM (SDM)	1.10	14.20	34.50	36.20
	SLX		12.40	59.10	86.80	89.00		
	400	AIC	SIM	<i>Rook</i>				
73.60								
SLM				77.20	70.60	71.90	71.70	
SEM (SDM)			22.80	22.80	36.40	36.60		
SLX			69.20	78.70	81.20	80.70		
BIC			SIM	94.20				
				SLM	88.20	95.90	96.30	96.20
				SEM (SDM)	6.10	6.10	36.40	36.60
SLX			59.10	96.00	97.50	97.40		
AIC		SIM	<i>Queen</i>					
			72.00					
			SLM	72.90	72.30	73.40	73.10	
		SEM (SDM)	13.80	35.30	35.90	36.60		
		SLX	52.00	73.90	78.30	78.80		
		BIC	SIM	93.90				
				SLM	61.50	96.90	96.40	96.50
				SEM (SDM)	2.90	31.80	35.90	36.60
		SLX	31.40	93.40	97.10	97.40		

criteria select this model at a low rate, close to 40%, picking very frequently its relative models the SARAR and the SDEM.

The simulation process is extended under heteroscedasticity and non-normality, two non-ideal conditions that produced smaller selection rates of the true model by both

Table 16 Percentage of selections only for the generated models SDM, SARAR and SDEM under Log-normal errors based on AIC and BIC for various values of ρ , λ and θ using 1000 replications

N	IC	True Model	ρ, λ			0.5			0.8			0.9					
			0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2			
100	AIC	SDM	<i>Rook</i>														
			3.60	25.70	41.20	44.90	31.60	79.30	91.80	92.90	52.10	93.80	98.80	99.00	57.90	95.50	99.50
			17.70	52.90	68.60	71.40	25.80	71.60	93.20	97.30	36.90	89.10	99.90	100	42.30	94.80	100
	BIC	SDM	<i>Rook</i>														
			4.30	20.80	29.70	31.20	21.90	44.60	75.00	78.40	31.10	62.10	85.20	90.60	34.10	64.20	86.90
			0.30	10.40	22.50	24.40	15.30	70.50	90.20	92.30	28.00	86.10	97.70	98.40	32.70	89.60	98.90
	AIC	SDM	<i>Queen</i>														
			7.40	42.10	67.20	71.30	11.40	58.00	90.60	96.80	17.10	75.00	99.70	100	20.60	85.20	100
			0.80	5.60	14.20	15.60	10.90	19.20	65.30	72.50	29.60	48.50	74.70	84.40	33.60	59.00	86.60
	BIC	SDM	<i>Queen</i>														
			0.70	5.80	14.70	16.80	14.90	45.50	71.20	76.20	30.60	75.10	92.80	94.50	36.70	83.00	97.30
			9.00	33.40	52.90	59.00	13.60	49.50	82.10	90.40	22.70	71.00	98.60	99.80	29.80	81.60	99.70
BIC	SDM	<i>Queen</i>															
		1.20	5.20	13.60	15.50	15.20	15.00	37.20	44.50	24.80	36.40	47.70	52.70	25.00	38.30	57.10	
		0.10	0.40	3.50	5.40	2.80	24.50	59.20	67.70	11.30	54.80	87.50	91.70	14.00	64.40	94.60	
BIC	SDM	<i>Queen</i>															
		2.50	18.50	48.50	56.80	4.50	30.80	76.10	87.20	7.00	50.60	95.20	99.30	9.70	63.10	99.90	
		0.10	0.20	2.80	4.40	6.10	2.80	10.10	17.50	21.40	25.00	26.50	29.10	23.90	32.30	39.80	

Table 16 (continued)

N	IC	True Model	ρ, λ		0.5				0.8				0.9					
			λ, θ	ρ, λ	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9	0.2	0.5	0.8	0.9		
400	AIC	SDM	<i>Rook</i>															
			29.90	65.10	77.60	79.50	61.90	96.20	99.50	99.60	80.70	99.90	100	100	85.20	100	100	100
			33.30	62.90	80.40	86.30	49.40	90.60	99.60	100	74.60	99.50	100	100	82.80	100	100	100
	BIC	SDEM	12.90	60.70	68.20	70.10	42.20	72.50	91.40	94.10	51.80	85.80	97.80	99.30	57.10	91.30	99.00	99.60
			8.20	50.10	64.80	68.00	36.20	95.40	99.50	99.60	50.20	99.50	100	100	55.70	99.60	100	100
			15.70	62.50	80.40	86.30	23.30	89.90	99.60	100	36.50	99.50	100	100	46.00	100	100	100
	AIC	SDM	<i>Queen</i>															
			12.00	37.00	54.80	58.00	38.90	81.50	93.90	95.50	58.50	97.20	99.70	99.80	64.40	98.90	100	100
			22.70	55.50	68.10	74.90	30.90	77.90	97.30	99.30	47.70	97.40	100	100	56.90	99.60	100	100
	BIC	SDEM	4.80	32.80	45.10	46.20	31.40	41.30	72.60	79.60	36.00	61.80	80.10	85.50	39.30	67.70	85.60	89.30
			0.00	15.00	29.70	32.00	15.10	75.60	93.70	95.50	25.30	93.30	99.60	99.70	29.80	95.10	99.80	100
			7.20	50.50	68.10	74.90	10.10	70.00	97.20	99.30	18.00	94.70	100	100	23.60	99.00	100	100
SDEM	SDEM	0.60	3.90	20.80	22.30	18.70	6.50	48.60	66.60	36.00	57.00	61.60	67.60	39.30	67.20	78.80	81.50	

criteria, with AIC having a larger decrease. However, heteroscedasticity for the generated models SEM, SLX, SDM and SDEM strongly influence the selection process leading to similar or relative models but not to the true and in some cases failing even to recognize spatial dependence. On the other hand, things did not change drastically under the relaxation of normality. Indeed, smaller selection rates of the true model are obtained under log-normal distribution, without observing any abnormal behaviors, whereas under student t distribution the selection rates were almost the same.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43071-022-00030-x>.

Acknowledgements We are indebted to the editor Giuseppe Arbia and to the two anonymous referees for their useful comments and suggestions that improved the overall research and the presentation of this manuscript.

Funding Open access funding provided by HEAL-Link Greece.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agiakloglou C, Tsimpanos A (2021) Evaluating information criteria for selecting spatial processes. *Ann Reg Sci* 66:677–697
- Agiakloglou C, Tsimbos C, Tsimpanos A (2015) Is spurious behaviour an issue for two independent stationary spatial autoregressive SAR(1) processes? *Appl Econ Lett* 22:1372–1377
- 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 (1988a) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht
- Anselin L (1988b) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geogr Anal* 20:1–17
- Anselin L, Florax RJ (1995) Small sample properties of tests for spatial dependence in regression models: Some further results. In: Anselin L, Florax RJ (eds) New directions in spatial econometrics. Springer-Verlag, Berlin, pp 21–74
- Anselin L, Bera A, Florax R, Yoon M (1996) Simple diagnostic tests for spatial dependence. *Reg Sci Urban Econ* 26:77–104
- 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
- Bivand R (2015) Spdep: spatial dependence: weighting schemes. Statistics and models. R package version 0.5-82. <http://CRAN.R-project.org/package=spdep>
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. New York, Springer
- Burridge P (1980) On the cliff-ord test for spatial correlation. *J R Stat Soc B* 42:107–108
- Chi G, Zhu J (2008) Spatial regression models for demographic analysis. *Popul Res Policy Rev* 27:17–42
- Cliff AD, Ord JK (1981) Spatial processes: models and applications. Pion Ltd, London

- Crespo Cuaresma J, Feldkircher M (2013) Spatial Filtering, model uncertainty and the speed of income convergence in Europe. *J Appl Economet* 28(4):720–741
- Debarys N, LeSage J (2022) Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices. *J Bus Econ Stat* 40:547–558
- Diebold FX (2007) *Elements of forecasting*, 4th edn. Thomson South-Western, Mason
- Elhorst JP, Halleck Vega S (2017) The SLX model: extensions and the sensitivity of spatial spillovers to W. *Pap De Econ Esp* 152:34–50
- Fernandez C, Ley E, Steel M (2001) Benchmark priors for Bayesian model averaging. *J Econom* 16(5):563–576
- Fingleton B (1999) Spurious spatial regression: some Monte Carlo results with a spatial unit root and spatial cointegration. *J Reg Sci* 39:1–19
- Florax RJGM, Folmer H, Rey SJ (2003) Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Reg Sci Urban Econ* 33(5):557–579
- Hoeting JA, Davis RA, Merton AA, Thompson SE (2006) Model selection for geostatistical models. *Ecol Appl* 16(1):87–98
- Jin F, Lee L-F (2013) Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Reg Sci Urban Econ* 43:590–616
- Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee T-C (1985) *The theory and practice of econometrics*, 2nd edn. Wiley, New York
- Kelejian HH, Piras G (2011) An extension of Kelejian's J-test for non-nested spatial model. *Reg Sci Urban Econ* 41:281–292
- 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
- LeSage J, Parent O (2007) Bayesian model averaging for spatial econometric models. *Geogr Anal* 39:241–267
- Mur J, Angulo AM (2006) The spatial durbin model and the common factor tests. *Spat Econ Anal* 1:207–226
- Mur J, Angulo AM (2009) Model selection strategies in a spatial setting: some additional results. *Reg Sci Urban Econ* 39:200–213
- Mur J, Trivez FJ (2003) Unit roots and deterministic trend in spatial econometrics models. *Int Reg Sci Rev* 26:289–312
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Zhang X, Yu J (2018) Spatial weights matrix selection and model averaging for spatial autoregressive models. *J Econom* 203:1–18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.