Regional Science and Urban Economics 39 (2009) 200–213

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

# Model selection strategies in a spatial setting: Some additional results

Jesús Mur *, Ana Angulo

*University of Zaragoza, Spain*

ARTICLE INFO

*Article history:*
Received 15 July 2005
Received in revised form 12 October 2007
Accepted 24 May 2008
Available online 15 October 2008

*JEL classification:*
C12
C13
C15
C52
R15

*Keywords:*
Spatial econometrics
Specification strategy
Model selection techniques

ABSTRACT

This paper continues from the discussion of Florax et al. (Florax, R., H. Folmer and S. Rey, 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. Regional Science and Urban Economics, 33, 557–579.), regarding the properties of various specification strategies for spatial econometric models. Habitual practise has popularised a technique based on the well-known Lagrange Multipliers, characterized as a *Specific-to-General* approach, and which seems to give good results. In our work, we contemplate other alternatives, some of which may be seen as slight variations of this proposal, including the selection tests of Vuong (Vuong, Q., 1989. Likelihood ratio-tests for model selection and non-nested hypotheses. Econometrica, 57, 307–333.) and of Clarke (Clarke, K., 2003. Nonparametric model discrimination in international relations. Journal of Conflict Resolutions, 47, 72–93.). We also examine an approach of the *General-to-Specific* type, as clearly opposite to the others. The comparison of the two strategies is carried out through a Monte Carlo experiment, the results of which are quite diffuse, in the sense that we do not find conclusive evidence in favour of either of these two approaches. However, it should be recognized that the *General-to-Specific* strategy seems to be more robust to the existence of anomalies in the Data Generating Process.



## 1. Introduction

Loosely speaking, we could say that there are hundreds of different ways to specify an econometric model. This is not a problem. Difficulties arise with the confidence in the final results of this arbitrariness. In general, the *Axiom of correct specification* (Leamer, 1978) does not hold, implying that there exists some interdependence between the specification path and the final model. Obviously, this situation is very undesirable and its resolution occupies a central position in the literature on econometric methodology (Morgan, 1990).

Our feeling is that this discussion has not ripened sufficiently within the field of spatial econometrics, even though Anselin (1988) already established the basis for its development (in Part III of his textbook). In fact, the list of papers dedicated specifically to comparing specification strategies in a spatial context seems very small. These papers include the seminal works of Blommestein (1983) and Bivand (1984), the simulation experiment of Anselin and Florax (1995), and the more recent paper of Florax et al. (2003). It is, therefore, no surprise that Anselin et al. (2004, p. 25) state that, "much more is needed in terms of comparative studies of competing paradigms and modeling 'philosophies'" in this field.

Our intention is to take up the discussion of Florax et al. (2003) (FFR from now on), where the authors review the strategies of specifying econometric models in a spatial context. According to FFR, at present, a classical, forward stepwise approach dominates, starting from a simple model. The success of this procedure depends, mainly, on the efficacy of the misspecification tests carried out during the process. However, as FFR indicate, we could adopt exactly the opposite approach in a 'Hendry-like specification strategy.'

Through a simulation experiment, FFR find that the first approach tends to outperform the second. This is an interesting result that contradicts part of the literature on econometric model selection (Danilov and Magnus, 2004; Hendry and Krolzig, 2005; Campos et al., 2005; see also Lütkepohl, 2007), and which caused the reaction of Hendry (2006) and the reply of FFR (2006). This contradiction is one of the motivations of our paper. Specifically, we wonder why FFR reach this conclusion and to what extent it is robust to the assumptions introduced into their analysis. In fact, our results indicate that the preference for the forward stepwise approach is not so evident.

In Section 2, we summarise a series of results, well established in mainstream econometrics, regarding the problem of how to compare models. Section 3 describes the search strategies that seem to dominate in a spatial context, as described by FFR, and makes some remarks on their position. In Section 4, we solve a simulation experiment, in order to compare the performance of the different approaches previously discussed. Section 5 presents some conclusions and prospects for future research.

## 2. Econometric model selection: a brief overview

There is a well-known aphorism of Box (1980) that, although a little simplistic, reflects a widespread position about how to specify an

* Corresponding author. Department of Economic Analysis, Gran Vía, 2-4 (50005) Zaragoza, Spain. Tel.: +34 976 761815.
*E-mail addresses:* jmur@unizar.es (J. Mur), aangulo@unizar.es (A. Angulo).

0166-0462/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2008.05.018

econometric model:[1] 'all models are false but some models are useful.' Popper (1979, p. 13) partly shares this scepticism: "But when he (the researcher) has fully digested what we can never justify empirically (…) the claim that a scientific theory is true, (…), then he may consider from the point of view of a seeker for true theories, the question: What principles of preference should we adopt? Are some theories 'better' than others?"

There is no obvious answer to either of these two questions. It would be easier to reach a high consensus on the second, by using the traditional positivist argument: "the adequacy of a theory can be evaluated by how successful it is in coming to grips with the reality it purports to explain" (Spanos, 1986, p. 662), but the agreement will be smaller with respect to the 'principles of preference.' In fact, the catalogue of such principles is very heterogeneous. Among them, we could cite the criteria of robustness, goodness of fit, predictive performance, theoretical consistency, stochastic equilibrium, parsimony or encompassment. Each of these has some specific advantages, but none of them seems to be clearly preferable to the others, as a general guide for specifying econometric models (Hendry and Richard, 1982).

In this paper, we will pay particular attention to three aspects:

(i) Model Selection or Hypothesis Testing;
(ii) *General-to-Specific* (*Gets*) or *Specific-to-General* (*Stge*) Modelling; and,
(iii) Nested or Non-Nested Models.

Dastoor and McAleer (1989) distinguish two broad approaches to the problem of comparing different econometric models. In the first approach, *Model Selection*, we look for the best model among various alternatives, whereas in the second approach, *Hypothesis Testing*, we test whether one model, in particular, is adequate.[2] The difference between the two approaches depends on the loss function adopted (Aznar, 1989). In the first case, the loss function is very diffuse between the alternatives, while in the second it is totally concentrated in the model of the null hypothesis. Unless strong and well-defined preferences exist in favour of one specification, the first strategy seems preferable (see, for example, Leamer, 1978; Sawa, 1978; Zellner, 1978); however, in applied spatial econometrics, strategies of the second type clearly dominate.

Should we start from a restrictive model and then enlarge it, if necessary? Or, is it preferable to begin with the largest specification possible and try to simplify it afterwards? This debate, outlined in the London School of Economics some decades ago, is still alive. The conventional approach, *Specific-to-General* (*Stge) modelling*, insists on a complete theoretical explanation prior to the empirical data analysis. Then, the model is *criticised*, in the sense of Cox and Wermuth (1996), and, if necessary, adjusted or enlarged until an adequate specification emerges. Empirical evidence is confined to little more than quantifying the different versions of the model. In the *General-to-Specific* approach (*Gets*), we begin from a firmly theoretically-rooted model, which must be congruent with the data, and then apply a simplification algorithm, in order to arrive at a simpler, more parsimonious equation which encompasses the other candidates. There are many pros and cons for each approach and the evidence seems to favour *Gets*, although not to the extent suggested by Hoover and Perez (2004, p. 778) when they state that the *Gets* algorithm "usually finds the truth nearly as well as one would if God had whispered the true specification in one's ear." However, in applied econometrics, an approach that is, implicitly or explicitly, of the *Stge* type dominates.

The third question is a bit more technical and refers to the relationship between the models, which can be nested, non-nested or partially nested.[3] The comparison of non-nested models, using a standard testing sequence, is not straightforward, because problems may appear regarding lack of identification, singularity in the information matrix, incorrect test size, etc. Chow (1983) advocated the use of specialised techniques, among which would include the Cox test, the model selection tests, the model selection criteria and the Bayesian approach (Clarke, 2004).

We will not discuss the latter, because it does not fit in with our approach, although we recognize the appeal of the Bayesian methodology (examples may be found in Hepple, 1995a,b; Spiegelhalter et al., 2002; LeSage, 1997, 2004). The Cox (1961, 1962) test compares two models, one in the null hypothesis and the other in the alternative. This framework has become popular through the $J$ and $J_A$ variants (Davidson and MacKinnon, 1981; Fisher and McAleer, 1979; respectively). Nevertheless, by 1987, King and McAleer had already recognised that the superiority of these new variants is, at least, doubtful (King and McAleer, 1987). Later, Gourieroux and Monfort (1989) demonstrated that the $J$ and $J_A$ tests are particular cases of a more general principle: *variance encompassment*. Lastly, Anselin (1986), in a Monte Carlo experiment,[4] concluded that, in a spatial context, the $J$ and $J_A$ tests "should be interpreted with caution, and used in addition to but not instead of other criteria to assess the validity of models" (p. 282).

The most extended technique to compare models in a non-nested framework uses one, or various, model selection criteria. Possibly the most popular is the Akaike Information Criteria (AIC) (Akaike, 1973) or its Bayesian version, the Schwarz Bayesian Information Criteria (SBIC) (Schwarz, 1978). The advantages of an information criteria approach are to be found in the simplicity and clarity of its conclusions: each selection criterion will always choose one model as the *best* model, according to the design of each criterion. Granger, King and White (1995, p. 180) were confident in this line of study: "The use of an information criterion…is most worthy of consideration." Thus, the information criteria approach now occupies a significant role in applied econometrics.

Finally, the novelty of model selection tests, with respect to criteria, is that they explicitly contemplate a situation of indifference between the given alternatives. This is because, on many occasions, the difference between the models is hardly appreciable, and the user should be aware of this. In our opinion, the model selection tests approach combines a solid statistical treatment of the decision problem, characteristic of the Cox test, with the simplicity of the selection criteria. Among the negative aspects, we should mention that (i) the use of selection tests is limited to a comparison of two, rival models, and (ii) these models must fulfil the hypothesis of no correlation. The tests of Vuong (1989) and of Clarke (2003) are two outstanding examples of this approach (see Appendix A for the details).

## 3. Econometric model selection in a spatial setting

Our position is that, although there are peculiarities when using spatial data, common principles should prevail when conducting applied spatial econometrics. The same spirit seems to prevail in the above-mentioned work of FFR. The peculiarities come from the nature of the data which, very often, results in problems of heterogeneity and of spatial dependence. Primarily, attention has been focused on the second topic, which has been considered key to obtaining a good specification. The implicit assumption is that a spatial econometric model will improve when it is capable of better capturing the cross-sectional dependence relationships that underlie the current specification. The issue of heterogeneity is receiving increasing attention

[1] For a review of the question, see, for example, Hausman (1992), Ripley (1996), or Burnham and Anderson (2002).

[2] The latter is similar to the approach that Box (1980) calls *model criticism*.

[3] There are several definitions of the concept of nesting, but most of them evolve around the original ideas of Cox (1962): two, or more, families of distribution functions are non-nested if none can be obtained as a limited case of any of the others.

[4] The aim of this paper was to analyze the performance of the $J$ and $J_A$ tests to find the weighting matrix that has intervened in the generation of the data. See, also, Baltagi and Li (2001), for another use of these tests.

(see, for example, Fotheringham et al., 1999; Pace and LeSage, 2004; Mur et al., 2008), although it still does not seem to be perceived as a strategic aspect in modelling.

### 3.1. The FFR study

The discussion on modelling strategies in FFR (2003) focuses on how to best identify the structure of spatial dependencies that underlies a given set of data. FFR present four different possibilities; three of them belong to the *Stge* framework, while the fourth is of the *Gets* type.

The proposals of the first group are based on a simple (usually static) initial model in which the omission of elements of spatial dependency is tested for, either in the main equation or in the errors of the model. In the so-called *classical* approach, the LMERR (Lagrange Multiplier for error dependence) and the LMLAG (Lagrange Multiplier for spatially lagged dependent variable) tests are used (see Appendix B). The problem is that, although they are designed to test for specification errors in only one direction, they also react to errors in others, which creates uncertainty. The solution proposed by FFR (2003, p. 561) seems *ad hoc*: "If both tests are significant, estimate the specification pointed to by the more significant of the two tests. For example, if LMLAG>LMERR then estimate the spatial lag model (…). If LMLAG<LMERR then estimate the spatial error model." If both are not significant, or if one of them is but the other is not, the model for the next stage is well identified. The second strategy is exactly the same as the previous one, but uses the *robust* tests, LMEL and LMLE, whose advantage is that they are robust to *local* misspecification errors (Anselin et al., 1996). However, this sensitivity does not disappear completely, and some specification uncertainty remains. FFR propose a combination of both strategies to form a new one, called *hybrid*. However, in Appendix C, we show that *hybrid* produces exactly the same results as the *classical* strategy.

The fourth strategy develops a *Gets* approach. The ample model that nests the other specifications is the 'autoregressive distributed lag model of the first order' (ADL), as defined by Bivand (1984, p. 27):

$$\begin{aligned} &y = \rho Wy + X\beta + WX\eta + \varepsilon \\ &\varepsilon \sim N(0; \sigma_\varepsilon^2 I) \end{aligned} \tag{1}$$

The common factor hypothesis reads as: $H_0$: $\rho\beta = -\eta$, and results in the *spatial error model* (SEM):

$$\left.\begin{aligned} &y = X\beta + u \\ &u = \theta Wu + \varepsilon \\ &\varepsilon \sim N(0; \sigma_\varepsilon^2 I) \end{aligned}\right\} \tag{2}$$

FFR (2003, p. 563) state that the "rejection of the common factor hypothesis is taken as evidence against the spatial error specification, and in favor of the (spatial) lag." The latter is the *spatial lag model* (SLM):

$$\begin{aligned} &y = \rho Wy + X\beta + \varepsilon \\ &\varepsilon \sim N(0; \sigma_\varepsilon^2 I) \end{aligned} \tag{3}$$

The resolution of the test is simple, using the likelihood ratio of common factors (LRCOM), which compares the log-likelihoods of the models of Eqs. (1) and (2). The first is the nesting model, and the second is the nested model (Burridge, 1981).

### 3.2. Additional remarks on the FFR (2003) study

The FFR (2003) study is very interesting and contains rather valuable results. However, there are several aspects that may be reconsidered, namely:

i. The model criticism in a *Stge* approach should be thorough.
ii. If we are confronted with non-nested models in a *Stge* approach, it may be worth using techniques for non-nested models.
iii. The general unrestricted model in a *Gets* approach must be congruent with the data.
iv. The simplification search in a *Gets* approach must guarantee that we do not suffer information loss during the process.

The first question refers to a very simple point: we stop the specification search once we arrive at a model that fulfils all the requisites, but we must be sure of this. That is, we will specify an SLM when we find symptoms of having omitted some spatial lag in the initial static model. However, we should also make sure that the error term of the final SLM model is, effectively, a white noise. The same occurs with the SEM: we will model the autocorrelation of the random term, but we must guarantee that the main equation is not misspecified. In the case of the SLM, we may use the $RS_{\lambda|\rho}$ Lagrange Multiplier to test for spatial dependence in the errors, whereas in the SEM case, one option is the $RS_{\rho|\lambda}$ test (see Appendix B). If we reject the null hypothesis in either of these two cases, the final model should be a spatially dynamic model with a spatial (autoregressive) structure in the random term (or SARAR model):

$$\left.\begin{aligned} &y = \rho Wy + X\beta + u \\ &u = \theta Wu + \varepsilon \\ &\varepsilon \sim N(0; \sigma_\varepsilon^2 I) \end{aligned}\right\} \tag{4}$$

The second question refers to the nesting structure of the models. It is clear that the initial static model:

$$\left.\begin{aligned} &y = X\beta + \varepsilon \\ &\varepsilon \sim N(0; \sigma_\varepsilon^2 I) \end{aligned}\right\} \tag{5}$$

is nested, both in the SEM of Eq. (2) and in the SLM of Eq. (3), but the SLM and the SEM models do not have any nesting relationship between them. If the model of Eq. (5) is misspecified, but the LMERR and LMLAG (or LMLE and LMEL) tests do not clearly identify the adequate alternative, there is a decision problem. We can solve this situation with the *ad hoc* proposal of FFR, mentioned previously, or by using techniques designed to compare non-nested models. The selection tests adapt well to this case, (provided that the weighting matrix, *W*, is symmetrical), as do the more general selection tests (which are not restricted by the symmetry clause).

In a *Gets* approach, congruency means that there are no signs of misspecification, which implies that we should be able to confirm that there is a theory-consistent identification of the parameters, that the formulation is admissible by the data, and that there is encompassment of the rival models (Hendry and Richard, 1989). The idea is to start from a congruent model to ensure that the selection process is reliable, using few, but efficient, and well calibrated, diagnostic tests. In our case, we may begin the discussion from the model of Eq. (1), ensuring that this equation is not misspecified by testing, for example, that the error term is a white noise. If there were still some autocorrelation in the error term, the inference would be misleading and we should re-specify the model accordingly (similar to Eq. (4)).

Finally, the simplification algorithm should be well designed, in order to avoid loss of information and incorrect routes. For example, autoregressive errors imply common factor dynamics in the parameters of the equation, which means, as Davidson (2000, p. 168) states, that the SEM model is testable through the LRCOM test. However, the rejection of the common factor dynamics hypothesis does not necessarily mean that we accept the SLM of Eq. (3). First, we have to test for possible externalities in the exogenous variables (the significance of the parameters included in $\eta$). Once again, at each step from Eqs. (1) and (2) or Eq. (3), we must take into account the key diagnostic tests for misspecification.

In Figs. 1 and 2 we summarize the functioning of both strategies of model selection, applied to the spatial case.

In the *Gets* approach, the initial model will be the ADL of Eq. (1), in which we will verify that there are no specification errors, by using the $RS_{\lambda/\rho}$ test (indicated as level 1 in the figure). If the null hypothesis is rejected, this model cannot be used as a starting point and it will be necessary to propose broader models, such as the extended SARAR of

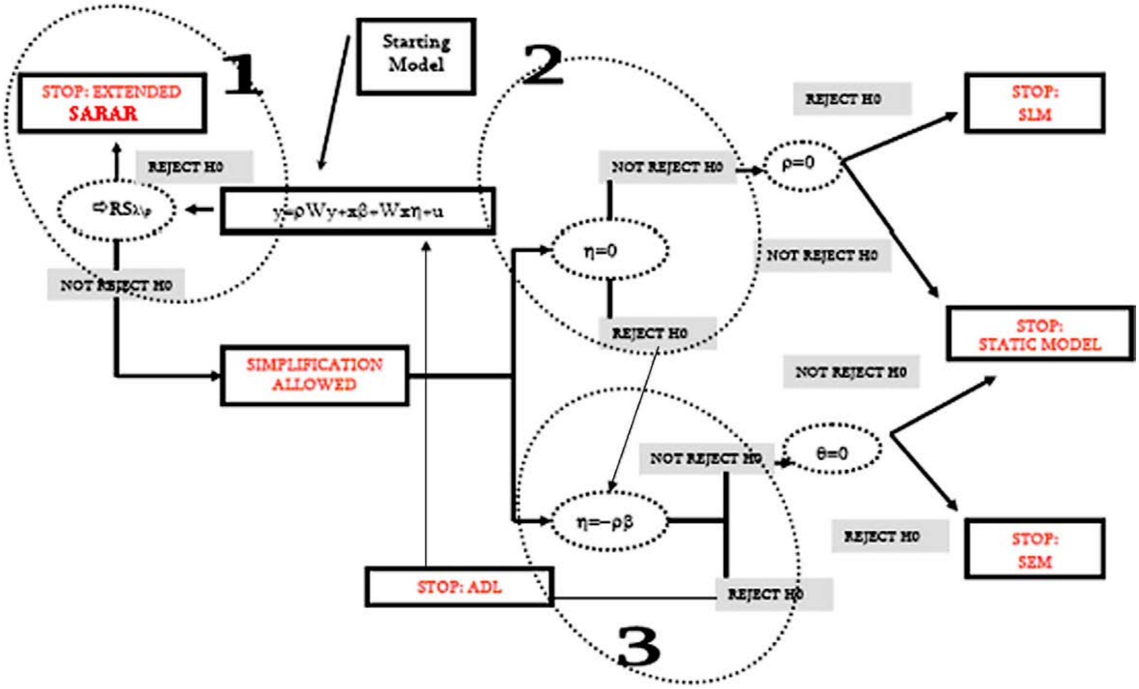


**Fig. 1.** A flow-chart for the *Gets* approach.

Eq. (4). Only when the null hypothesis of the $RS_{\lambda/\rho}$ test is accepted, will it make any sense to begin the simplification process. Secondly, it must be tested (level 2) that the spatial lag of the exogenous variables is relevant in the specification. If we cannot reject the null hypothesis, the evidence points to an SLM or to a static model, such as Eq. (5), depending on what happens with the significance test for the parameter $\rho$. If the null hypothesis is rejected at level 2, we can discard purely SLM structures, and leave three possibilities open: the initial ADL model, an SEM model like Eq. (2), or an equation without spatial dynamics (again, that of Eq. (5)). In this phase of the discussion, level 3, the LRCOM test of common factors plays a major role.

The sequence of the *Stge* approach is simpler. The discussion begins with the static model of Eq. (5), in which symptoms of misspecification should be looked for, using the appropriate tests

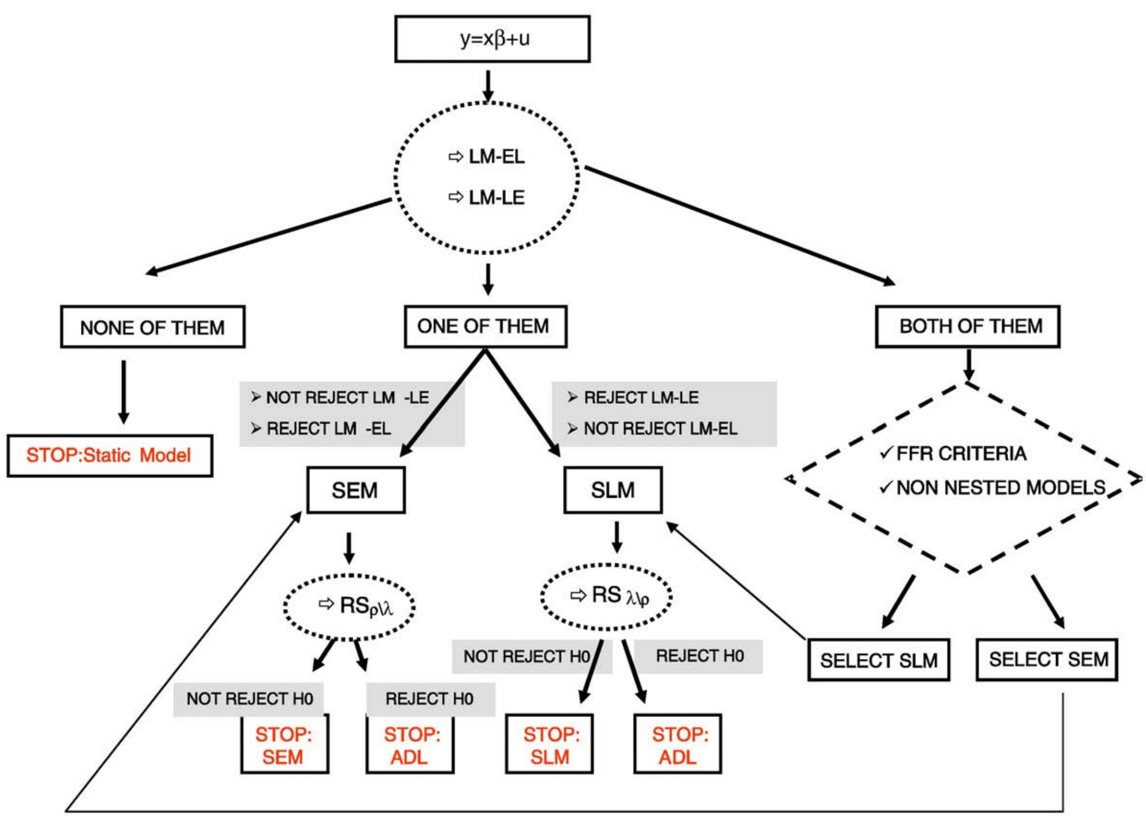


**Fig. 2.** A flow-chart for the *Stge* approach.

(in our case, the robust Multipliers). Initially, three situations may arise. If we cannot reject the null hypothesis of the two robust Multipliers, the discussion ends at this point with the selection of the static model of Eq. (5). If we reject the null hypothesis of one of them and accept the other, the evidence is against the static model and favors a specification with residual autocorrelation (SEM model), or substantive autocorrelation (SLM model). Before reaching a decision in this case, we should make sure that no errors remain in the corresponding specification, through the adequate tests, $RS_{\lambda/\rho}$ or $RS_{\rho/\lambda}$. The third situation, rejecting the null hypothesis of both robust Multipliers, is more problematic and presents us with a formal decision problem, which can be resolved in different ways. One possibility is to use the FFR rule, and another consists of using whatever selection criterion between non-nested models that seems most appropriate.

## 4. Monte Carlo experiments: design and empirical results

In this section, we evaluate the performance of the *Gets* and *Stge* specification strategies, by means of a series of Monte Carlo experiments. We use several spatial layouts, as well as different Data Generating Processes (DGP). Section 4.1 describes the characteristics of the experiments and Section 4.2 focuses on the results.

### 4.1. Design of the Monte Carlo experiments

In this exercise, we use data obtained from four different DGP: a SEM, a SLM, a SARAR and a spatially independent model, or SIM. The problem is to determine which process has intervened in the generation of the data, using only the data. To do so, we will employ a strategy of the *Gets* type or the *Stge* type.

Each of these four DGP has been simulated under different conditions. To be precise, the cases that we have contemplated are as follows:

i. Ideal conditions. This situation corresponds to Eqs. (2)–(5), respectively, in which all the hypotheses are fulfilled.
ii. Heteroskedasticity. The error terms are obtained from a normal distribution with non-constant variance: $\varepsilon_r \sim N(0;\sigma_\varepsilon^2 h_r)$, where $h_r$ reflects the pattern of heteroskedasticity in which we are interested. In this case, we used two spatial heteroskedasticity patterns, denoted as H1 and H2, and a non-spatial scheme, H3. The skedastic function for the first two cases is: $h_r = d(A,r)$, where $d(-)$ is the distance between the centroids of cells $A$ and $r$. In the H1 case, A is a cell situated in the upper left-hand corner of the lattice, whereas, in H2, this cell is located in the centre of the lattice. The skedastic function in case H3 is $h_r = |x_r|$, a non-spatial pattern, which depends on the realization of the regressor, $x_r$.
iii. Non-normal distribution of the error terms. Two distributions are used: a log-normal distribution, and a *t*-student distribution with degrees of freedom equal to 2% of the sample size. The first allows us to measure the consequences of the asymmetry, while the second provides information about the impact of outliers.
iv. Finally, we will explore whether the existence of endogeneity in the data, omitted in the equations, affects the behaviour of the two approaches. In order to do this, we simply introduce a positive correlation between the error term and the regressor.

The remaining characteristics of the exercise are as follows:

a. Only one regressor has been used in the model. The coefficient associated with it takes the value of 10, and 2 in the case of the intercept. Both magnitudes guarantee that, in the absence of spatial effects in the model, the $R^2$ will be close to 0.8.
b. The observations of the regressor and, where necessary, of the random terms $\varepsilon$ and u, have been generated using univariate normal distributions with zero mean and unit variance. That is, $\sigma_\varepsilon^2$ is equal to one, in all expressions.
c. We have used three different sample sizes, with 49, 100 and 225 observations, respectively.

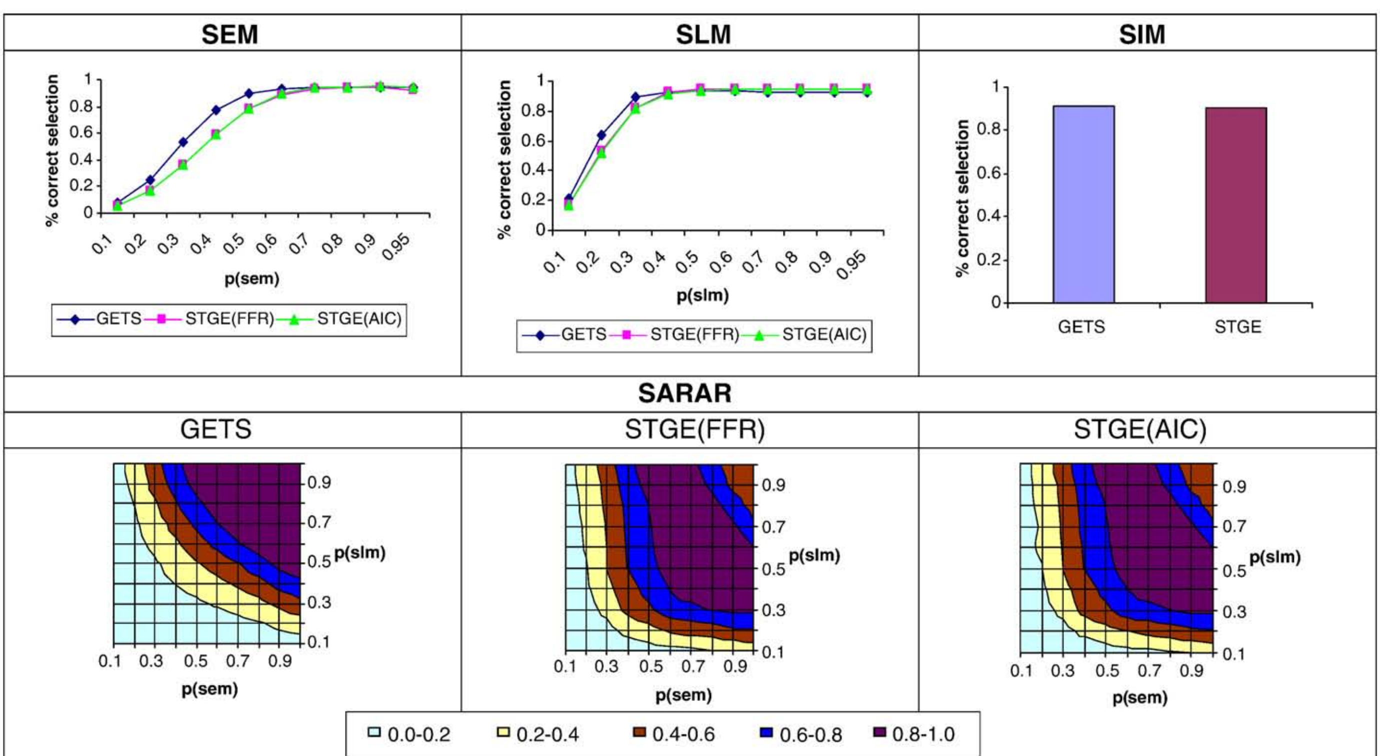


**Fig. 3.** Probability of a correct selection for the different DGP under ideal conditions, ($R = 100$, row-standardised weighting matrix).

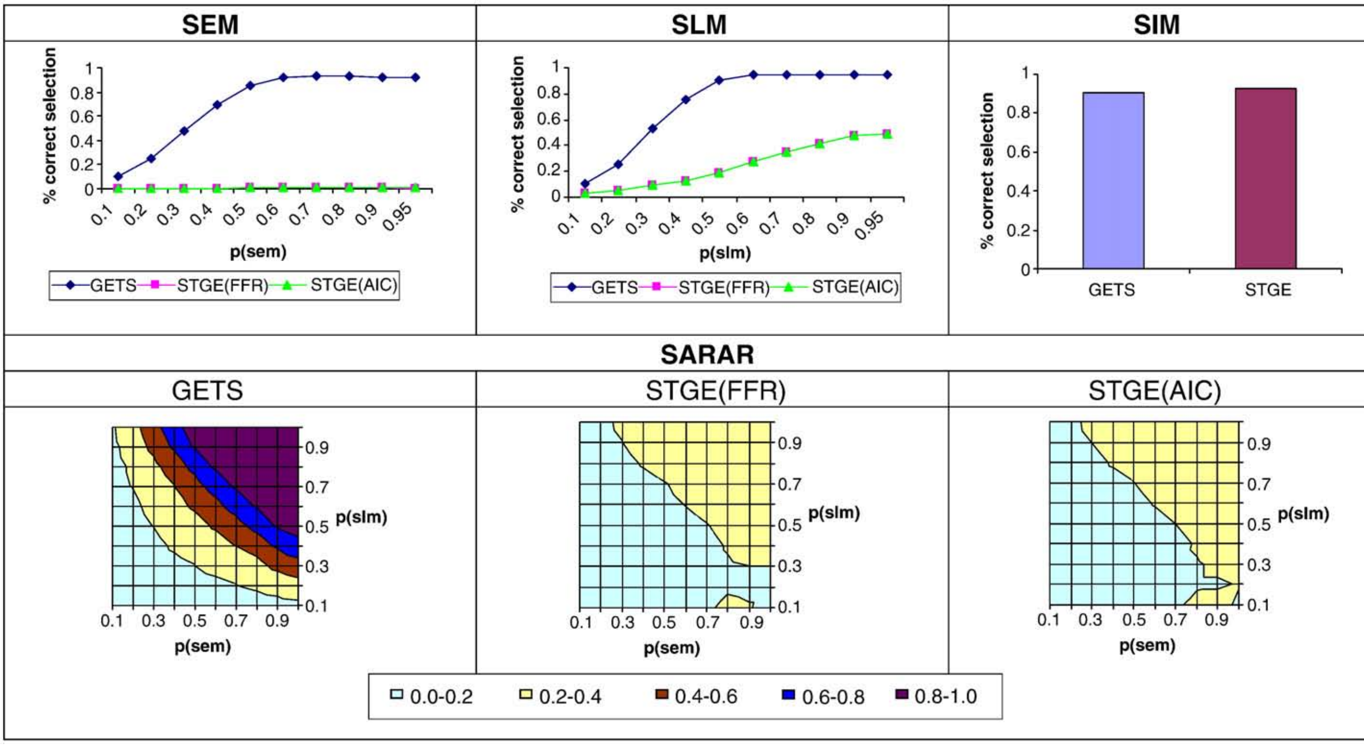


**Fig. 4.** Probability of a correct selection for the different DGP under heteroskedasticity type 1 (H1), ($R=100$, row-standardised weighting matrix).

d. Two types of contiguity matrix have been specified, for all the cases. A binary matrix, using rook-type contacts in a regular lattice system of (7×7), (10×10) or (15×15), and the corresponding row-standardised matrix.
e. In each case, 10 values of the parameters $\rho$ and/or $\theta$ have been simulated, distributed regularly over the positive side of the range of admissible values for each parameter.
f. Each combination has been repeated 1000 times.

### 4.2. Results of the Monte Carlo experiments[5]

The Monte Carlo experiment has provided us with a huge amount of information about the performance of the two selection strategies. In order to simplify the discussion, we will present our results by means of figures and response surface regressions. To begin with, Figs. 3–9 compare the performance of the two strategies for the main cases of interest (only for the 10×10 lattice using a row-standardised matrix). In Fig. 3, we simulated the DGP under ideal conditions (that is, homoskedasticity and normality in the errors, and strict exogeneity of the regressor), whereas in the other figures we introduced some anomaly.

From our point of view, the results are quite clear: the *Gets* and the *Stge* strategies are hardly distinguishable, when the DGP verifies all the standard assumptions (see Fig. 3), but, when the DGP operates with some distortion, the *Gets* strategy seems superior. It is important to note that the difference in favour of *Gets* appears when the distortions suffered by the DGP refer to non-normality or to heteroskedasticity with a spatial pattern. Both strategies behave quite similarly in the case of a non-spatial heteroskedasticity (see Fig. 6), or the advantage corresponds to *Stge* (as in the case of endogeneity in a SEM model, Fig. 9).

In addition to these figures, and according to Florax and de Graaff (2004), we estimated the following equation for each of the four DGP used in the simulation:

$$\log\left(\frac{p}{1-p}\right) = \gamma + z\pi + \zeta \tag{6}$$

where p is the empirical probability of choosing the right DGP using the corresponding selection approach, $\gamma$ is an intercept term, $z$ the design matrix, $\pi$ is a vector of parameters and $\zeta$ is a white noise error term. Table 1 shows the results corresponding to the *Gets* approach.

In this instance, as in the following cases, the category of reference against which the dummy variables should be interpreted is the DGP indicated in each column, simulated under ideal conditions and using the binary weighting matrix. For example, the parameter associated with the variable DH1, reflects the impact (with respect to the case of reference) produced by the existence of a problem of heteroskedasticity of type H1, when using the binary weighting matrix. The parameter associated with DTIP[6] measures what happens when we substitute the binary matrix of the case of reference by its row-standardised version. Finally, the parameter associated with the composed variable DH1 * DTIP shows the consequences of introducing heteroskedasticity of type H1 into the DGP, using the row-standardised weighting matrix instead of the original binary one. That is, the upper part of the table corresponds to the binary contiguity matrix and the bottom to the row-standardised matrix.

In general, the results are what one would expect. The *Gets* strategy works better when the DGP operates in ideal conditions, while heteroskedasticity and non-normality are damaging. The effect of the former is unequivocally harmful if it has a spatial pattern (cases H1 or H2) and increases when it acts in models with a dynamic structure in the main equation. In the case of non-normality, the

[5] Part of the debate between Hendry (2006) and FFR (2006) focuses on the necessity of "standardizing null rejection frequencies" (Hendry, 2006, pp. 311) when comparing the 'power' of different approaches through Monte Carlo. The reply of FFR (2006) is that "the practitioner typically uses asymptotically motivated cut-off points" (p. 302). In our case, we have decided to follow the suggestion of FFR, mainly because this paper tends to replicate the conditions of their study. The use of size-adjusted empirical power functions does not change, substantially, the results of the comparison that we present. The detailed results of the simulation may be requested from the authors.

[6] Defined in Note (a) of Tables 1 and 3.

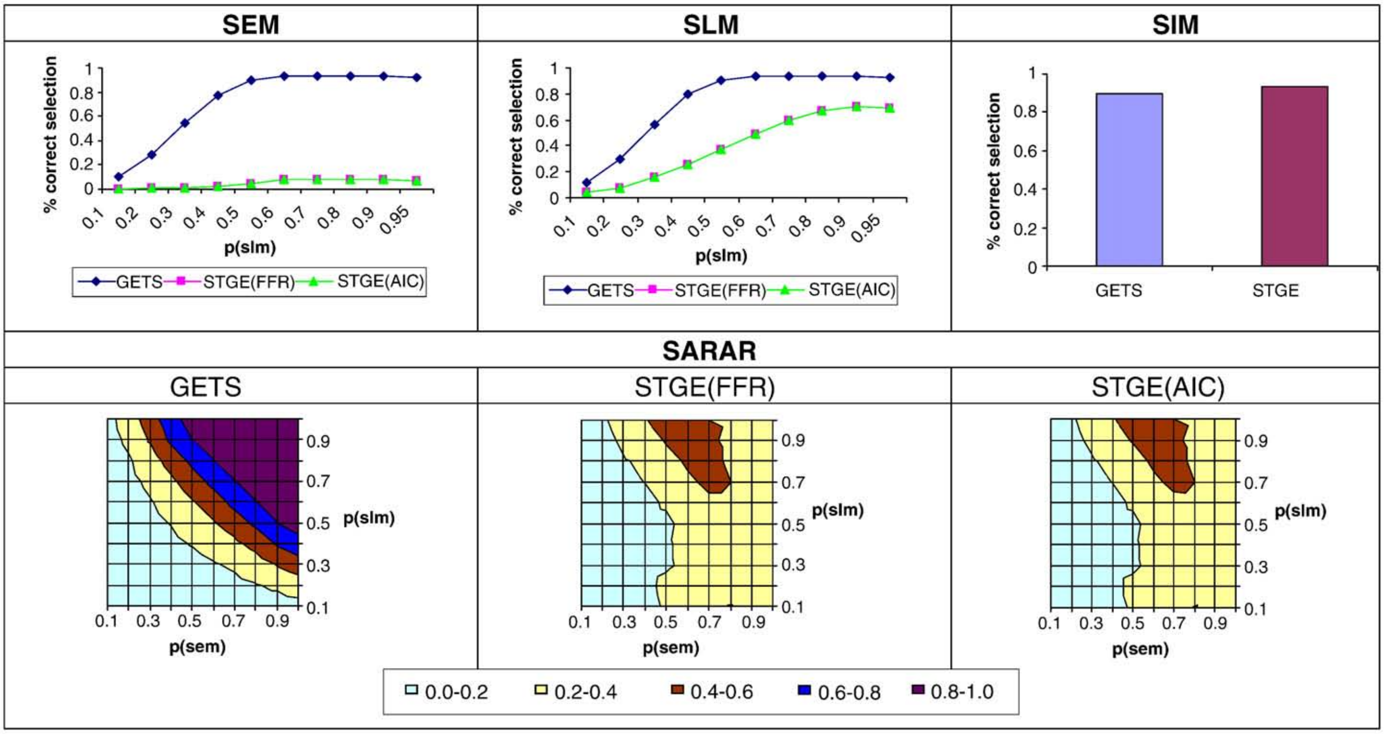


**Fig. 5.** Probability of a correct selection for the different DGP under heteroskedasticity type 2 (H2), ($R$ = 100, row-standardised weighting matrix).

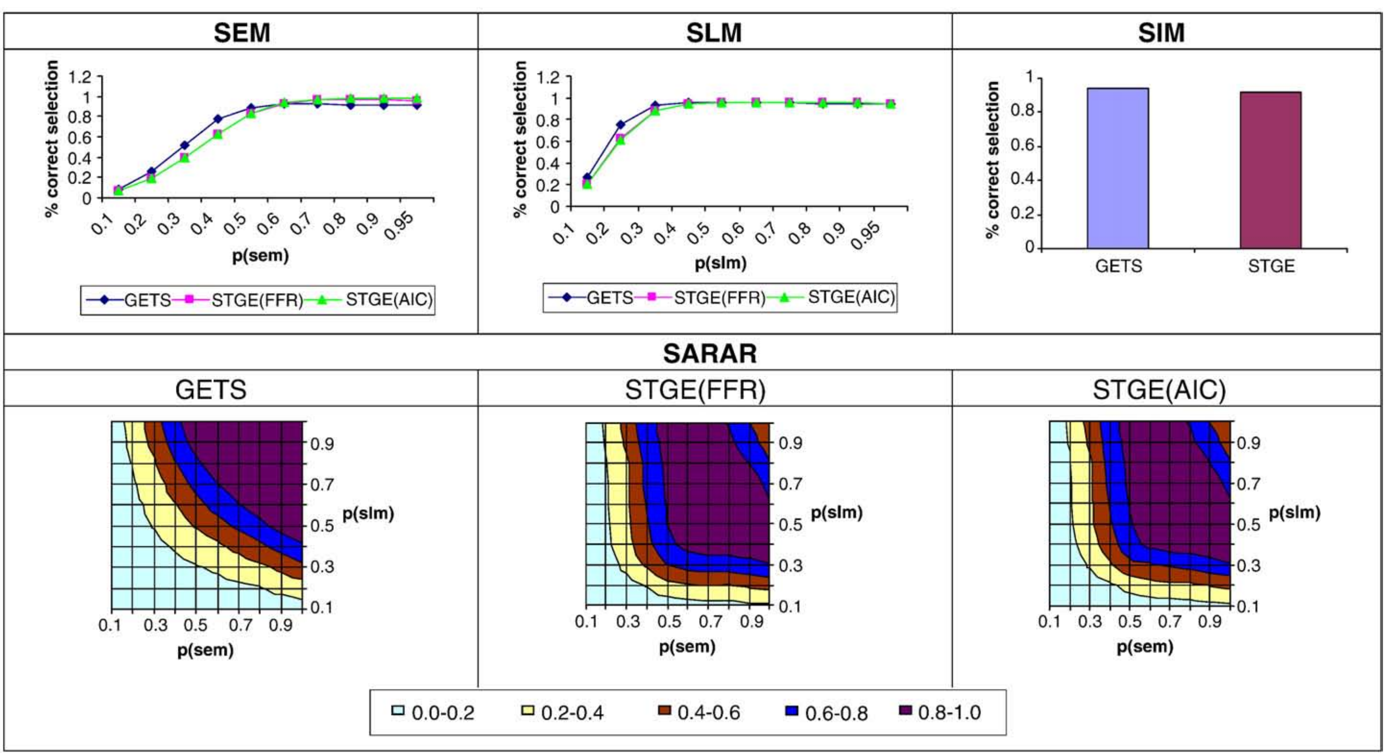


**Fig. 6.** Probability of a correct selection for the different DGP under heteroskedasticity type 3 (H3), ($R$ = 100, row-standardised weighting matrix).

presence of outliers in the sample (measured by means of DSTU[7]) seems to be more harmful than the asymmetry in the error distribution function (associated with DLOGN[8]). Surprisingly, endogeneity has an ambiguous effect: it is very prejudicial when it appears in SEM models (in fact, the SIM model tends to be selected), but it is beneficial in SLM or mixed type, SARAR, models. Furthermore, the results of this strategy improve greatly with the sample size and with the intensity of the cross-sectional dependence, especially in SEM models.

[7] Defined in Note (a) of Tables 1 and 3.
[8] Defined in Note (a) of Tables 1 and 3.

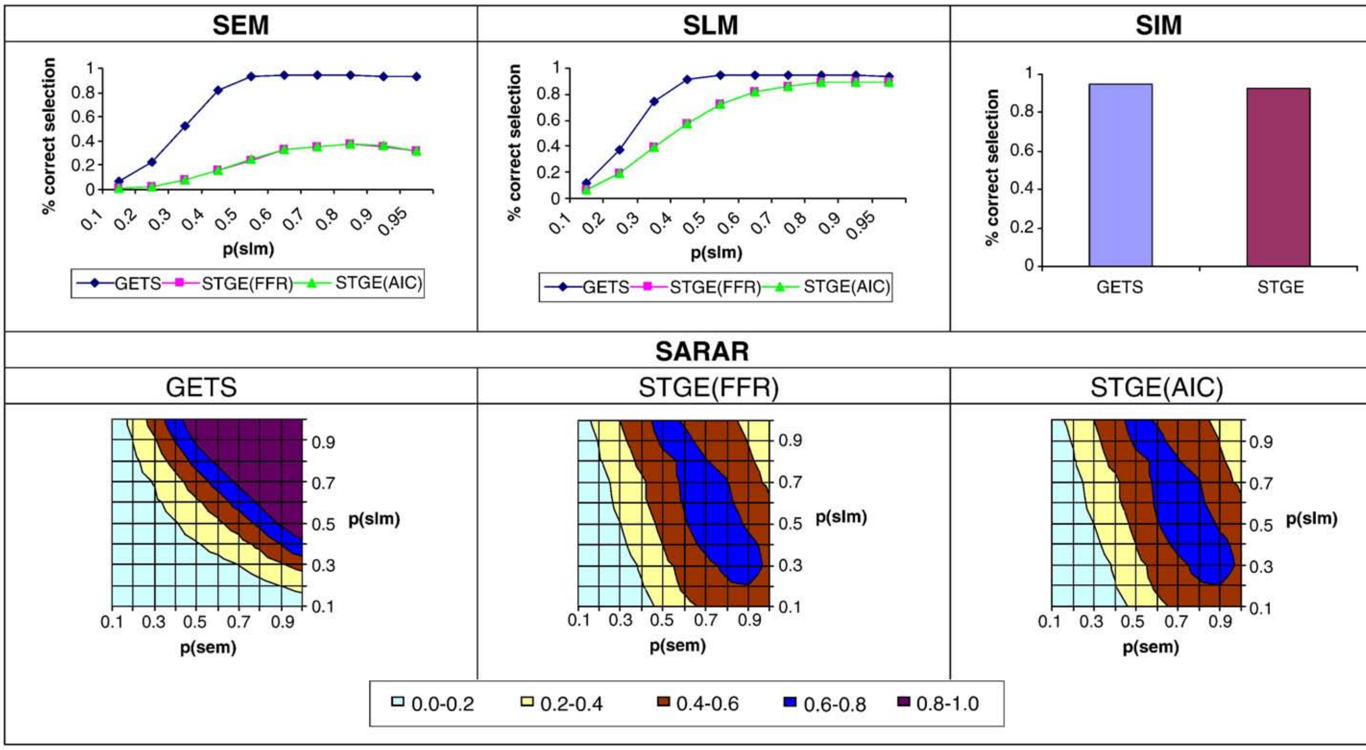


**Fig. 7.** Probability of a correct selection for the different DGP under log-normal distribution of the error term, ($R$ = 100, row-standardised weighting matrix).

The type of matrix that has intervened in the DGP, standardised or not, is another aspect that affects the behaviour of the strategy, as the results of the $F$ test of the equality of the coefficients show. The use of a row-standardised matrix tends to slightly improve the efficiency of the *Gets* approach, although the adjustment is a bit more tortuous. The effect associated with the size of the sample appears less important than in the binary case, while that of the intensity of the cross-sectional dependence, in the process that we are trying to identify, acquires greater importance.

With respect to the *Stge*, the first point to highlight is that, when the weighting matrix is of a binary type, we have up to four different criteria that we can use to resolve the doubt that arises when both

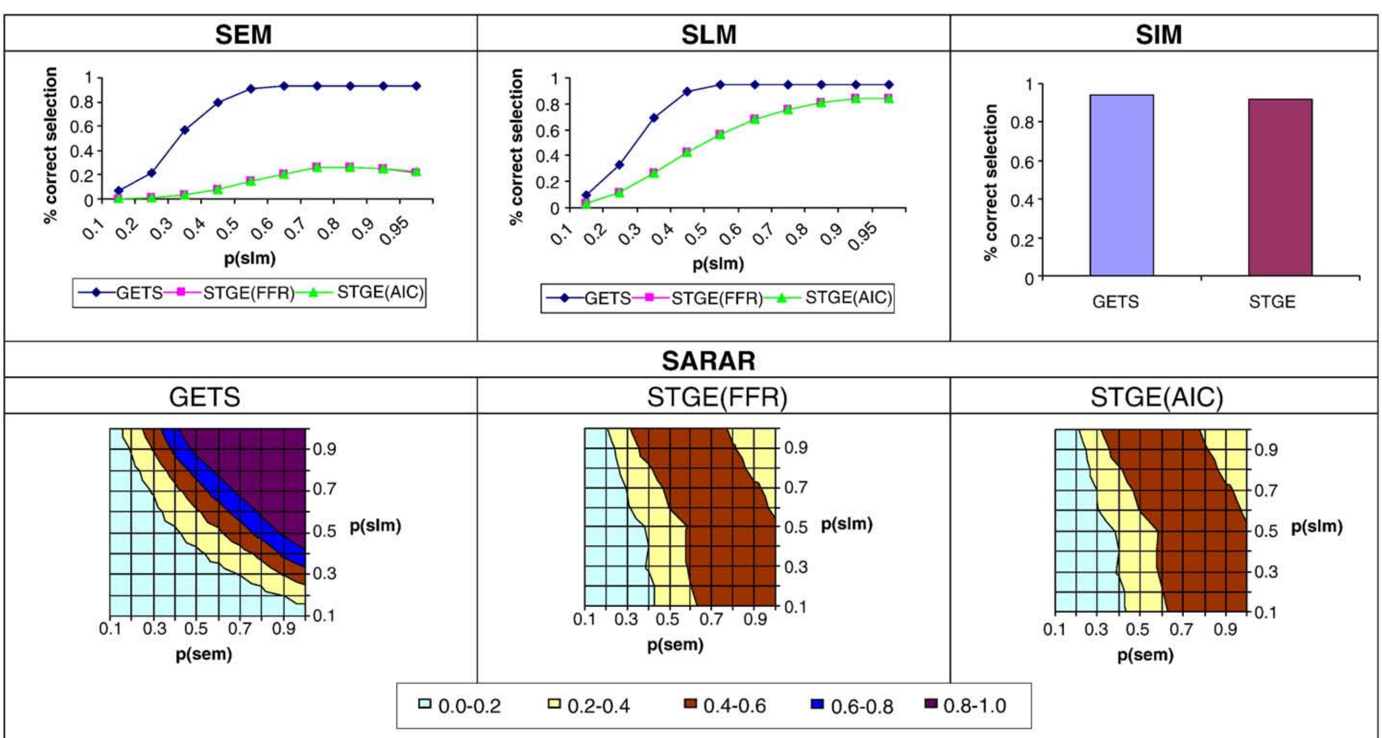


**Fig. 8.** Probability of a correct selection for the different DGP under a $t$-student distribution of the error term, ($R$ = 100, row-standardised weighting matrix).

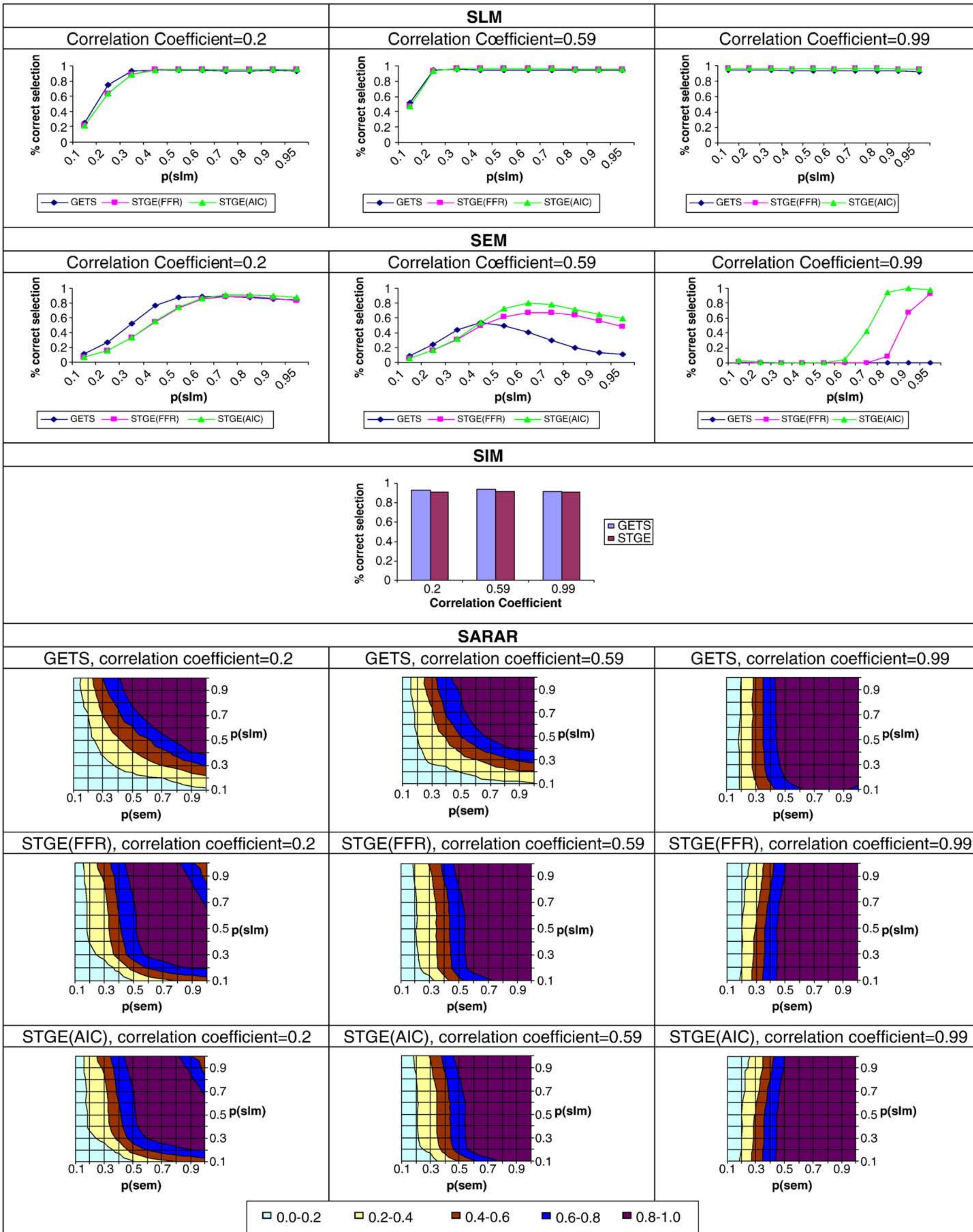


**Fig. 9.** Probability of a correct selection for the different GDP under endogenerity problems ($R=100$, row-standardised weighting matrix).

**Table 1**
Response surfaces estimates

| | SEM | | SLM | | SARAR | | SIM | |
|---|---|---|---|---|---|---|---|---|
| | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio |
| $\gamma$ | 0.389* | 9.281 | 0.563* | 17.516 | −0.558* | −41.984 | 0.835* | 158.669 |
| Correl | −0.819* | −15.809 | 0.046 | 1.277 | 0.160* | 12.311 | −0.020* | −2.779 |
| DH1 | −0.060 | −1.258 | −0.102* | −2.769 | −0.084* | −6.781 | −0.014 | −1.939 |
| DH2 | −0.079 | −1.659 | −0.093* | −2.525 | −0.053* | −4.268 | −0.028* | −3.800 |
| DH3 | −0.086 | −1.801 | 0.020 | 0.551 | 0.030* | 2.421 | −0.006 | −0.804 |
| DLOGN | −0.414* | −8.678 | −0.022 | −0.593 | −0.021 | −1.695 | 0.013 | 1.739 |
| DTSTU | −0.070 | −1.482 | −0.095* | −2.583 | −0.049* | −3.972 | −0.013 | −1.803 |
| $\rho$ | – | – | 1.154* | 9.383 | 2.549* | 56.905 | – | – |
| $\theta$ | 1.474* | 9.181 | – | – | 3.193* | 71.275 | – | – |
| Sample size | 0.001* | 3.601 | 0.001* | 4.097 | 0.002* | 39.594 | 0.000* | 3.776 |
| DTIP | 0.051 | 0.871 | −0.160* | −3.539 | 0.016 | 0.834 | 0.093* | 12.539 |
| Correl*DTIP | −0.934* | −18.418 | 0.274* | 7.582 | 0.214* | 16.491 | −0.017* | −2.356 |
| DH1*DTIP | −0.097* | −2.051 | 0.005 | 0.136 | 0.008 | 0.636 | −0.027* | −3.640 |
| DH2*DTIP | −0.092 | −1.948 | 0.017 | 0.453 | −0.021 | −1.678 | −0.027* | −3.686 |
| DH3*DTIP | −0.099* | −2.083 | 0.130* | 3.529 | 0.015 | 1.238 | 0.007 | 0.945 |
| DLOGN*DTIP | −0.086 | −1.808 | 0.054 | 1.468 | −0.026* | −2.074 | 0.009 | 1.218 |
| DTSTU*DTIP | −0.092 | −1.938 | 0.034 | 0.927 | −0.026* | −2.108 | 0.006 | 0.809 |
| $\rho$*DTIP | – | – | 0.498* | 15.456 | 0.705* | 59.996 | – | – |
| $\theta$*DTIP | 0.565* | 13.492 | – | – | 0.823* | 70.070 | – | – |
| Sample size*DTIP | 0.000* | 2.765 | 0.000* | 2.823 | 0.001* | 30.037 | 0.000* | 4.080 |
| $n$ | 660 | | 660 | | 5400 | | 66 | |
| Adjus. $R^2$ | 0.692 | | 0.407 | | 0.804 | | 0.932 | |
| Log likelihood | 42.78 | | 257.68 | | 1885.37 | | 208.35 | |
| $F_{VA}$ | 88.03 (0.000) | | 27.67 (0.000) | | 1170.54 (0.000) | | 60.87 (0.000) | |
| $F$(tip = bin) | 9.45 (0.000) | | 5.99 (0.000) | | 474.27 (0.000) | | 2.15 (0.050) | |

GETS strategy.
Correl: correlation level between the error term and the regressor (endogeneity measure). DH1: takes a value of 1 in the case of a H1 heteroskedasticity pattern; DH2: takes a value of 1 in the case of a H2 heteroskedasticity pattern; DH3: takes a value of 1 in the case of a H3 heteroskedasticity pattern; DLOGN: takes a value of 1 in the case of log-normal distribution; DTSTU: takes a value of 1 in the case of t-student distribution. DTIP: Takes a value of 1 when using the row-normalised weight matrix; $F_{VA}$: tests that all the coefficients in the regression, except the intercept, are zero (*p*-value in parentheses); $F$(tip = bin): tests that the parameters corresponding to the binary case are equal to the parameters associated with the row-standardised case (*p*-value in parentheses).
An asterisk indicates that the parameter is significant at the 5% level; *t*-ratios are calculated using the White heteroskedasticity-consistent standard errors.

robust Multipliers are significant (the so-called FFR criterion, the Akaike (1973) AIC statistic, and the tests of Vuong (1989) and of Clarke (2003). There are only two options (FFR and AIC) when a row-standardised matrix intervenes in the DGP.

Without elaborating on the details, our results indicate that using the Clarke test to resolve this doubt is not very satisfactory when it acts on models whose structure is exclusively SEM, while the Vuong test has many problems under SARAR structures. The difficulties of both tests increase when the DGP does not develop under strictly ideal conditions. In Table 2, we include the *F* tests of the equality of the coefficients associated with each criterion, obtained from the response surface regressions estimated, respectively, for the three DGPs under consideration in this case. Obviously, the weighting matrix that has acted on these simulations is of the binary type.

**Table 2**
F test for the equality of coefficients in the response surfaces estimates

| | Vuong | | Clarke | | FFR | |
|---|---|---|---|---|---|---|
| | *F* statistic | *p*-value | *F* statistic | *p*-value | *F* statistic | *p*-value |
| *SEM* | | | | | | |
| Akaike | 3.29 | 0.00 | 102.38 | 0.00 | 1.59 | 0.11 |
| Vuong | – | – | 94.11 | 0.00 | 0.51 | 0.87 |
| Clarke | – | – | – | – | 95.58 | 0.00 |
| *SLM* | | | | | | |
| Akaike | 0.02 | 1.00 | 0.17 | 0.99 | 0.05 | 1.00 |
| Vuong | – | – | 0.14 | 0.99 | 0.10 | 0.99 |
| Clarke | – | – | – | – | 0.35 | 0.96 |
| *SARAR* | | | | | | |
| Akaike | 78.19 | 0.00 | 15.53 | 0.00 | 0.91 | 0.53 |
| Vuong | – | – | 102.93 | 0.00 | 38.62 | 0.00 |
| Clarke | – | – | – | – | 18.38 | 0.00 |

*STGE* strategy and binary weighting matrix.

The four criteria are practically indifferent when they act with data coming from an SLM model: all the *F* tests support the supposition of the equality of the coefficients. The Clarke test separates from the other three criteria, when the data come from an SEM. In the case of SARAR, there is a strong discrepancy between the FFR and the AIC criteria (which are very similar to each other) and the tests of Vuong and of Clarke, the latter two being clearly inferior. To sum up, although the selection tests (of Vuong and of Clarke) present some specific advantages, the AIC and FFR criteria appear to be more robust for developing a *Stge* selection strategy using a binary matrix.

The substitution of the binary matrix with a row-standardised one, apart from excluding the last two tests (of Vuong and of Clarke), does not substantially change the results: the FFR and AIC criteria offer very similar results (only in cases of strong endogeneity does the AIC criterion appear preferable to the FFR). For this reason, we will now focus on the functioning of the *Stge* approach, using the FFR criterion to resolve the doubtful situations mentioned. In Table 3, we can see the estimated response surfaces for the four DGP that we are considering.

The estimations are very similar to those corresponding to the *Gets* approach in Table 1. For example, it is evident that the *Stge* approach works better when all the habitual assumptions hold in the DGP. Heteroskedasticity and non-normality have negative consequences that are much stronger in this case, and especially in structures that are exclusively SEM. Asymmetry in the error distribution function also seems to have a lower impact. There is still ambiguity as to the importance of the relations of endogeneity in the model simulated: it is harmful in SEM structures but beneficial in SLM and SARAR mechanisms. Furthermore, and as in the *Gets* case, the type of matrix used in the DGP (row-standardised or not) is another aspect that affects the behaviour of the *Stge* strategy, improving slightly in the row-standardised case.

Finally, Table 4 allows us to directly compare the working of the two criteria. The left-hand-side variable, explained in each Tobit regression, is the percentage of correct selections of the DGP indicated in each column.

**Table 3**
Response surfaces estimates

| | SEM | | SLM | | SARAR | | SIM | |
|---|---|---|---|---|---|---|---|---|
| | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio |
| γ | 0.341* | 6.828 | 0.625* | 16.740 | −0.271* | −20.230 | 0.905* | 216.29 |
| Correl | −0.506* | −9.499 | 0.060* | 2.196 | 0.111* | 7.458 | −0.007 | −1.291 |
| DH1 | −0.762* | −19.010 | −0.247* | −5.047 | −0.140* | −10.937 | 0.036* | 5.885 |
| DH2 | −0.685* | −18.175 | −0.140* | −3.226 | −0.021 | −1.787 | 0.002 | 0.251 |
| DH3 | −0.078 | −1.690 | 0.020 | 0.710 | −0.002 | −0.153 | −0.002 | −0.592 |
| DLOGN | −0.691* | −13.230 | −0.036 | −1.032 | −0.018 | −1.407 | 0.013* | 2.865 |
| DTSTU | −0.480* | −9.910 | −0.156* | −3.478 | −0.060* | −3.987 | 0.012 | 1.276 |
| ρ | – | – | 1.554* | 10.246 | 3.627* | 78.053 | – | – |
| θ | 2.140* | 12.984 | – | – | 1.078* | 20.980 | – | – |
| Sample size | 0.001* | 5.333 | 0.000* | 3.061 | 0.002* | 38.238 | 0.000 | 1.321 |
| DTIP | 0.116 | 1.584 | −0.122* | −2.110 | 0.097* | 4.765 | 0.012* | 2.378 |
| Correl*DTIP | −0.636* | −10.446 | 0.030 | 0.861 | 0.181* | 11.444 | −0.007 | −1.139 |
| DH1*DTIP | −0.746* | −18.567 | −0.605* | −21.639 | −0.366* | −28.202 | 0.025* | 4.809 |
| DH2*DTIP | −0.705* | −18.339 | −0.440* | −12.725 | −0.280* | −22.504 | 0.010 | 1.713 |
| DH3*DTIP | −0.063 | −1.091 | −0.038 | −0.958 | −0.009 | −0.592 | 0.007 | 1.750 |
| DLOGN*DTIP | −0.503* | −13.849 | −0.243* | −5.914 | −0.127* | −9.533 | 0.014* | 3.540 |
| DTSTU*DTIP | −0.487* | −9.060 | −0.360* | −6.799 | −0.153* | −10.130 | 0.013* | 2.404 |
| ρ*DTIP | 0.321* | 6.919 | 0.472* | 11.835 | 0.640* | 46.048 | – | – |
| θ*DTIP | – | – | – | – | 0.203* | 14.517 | – | – |
| Sample size*DTIP | 0.001* | 4.957 | 0.001* | 6.390 | 0.002* | 37.663 | 0.000* | −2.302 |
| *n* | 660 | | 660 | | 5400 | | 66 | |
| Adjus. $R^2$ | 0.644 | | 0.625 | | 0.741 | | 0.549 | |
| Log likelihood | 34.40 | | 232.81 | | 1381.95 | | 223.44 | |
| $F_{VA}$ | 68.49 (0.000) | | 63.29 (0.000) | | 1796.14 (0.000) | | 6.46 (0.000) | |
| *F*(tip = bin) | 21.18 (0.000) | | 27.94 (0.000) | | 619.67 (0.000) | | 2.53 (0.030) | |

FFR approach and *STGE* strategy.
Correl: correlation level between the error term and the regressor (endogeneity measure). DH1: takes a value of 1 in the case of a H1 heteroskedasticity pattern; DH2: takes a value of 1 in the case of a H2 heteroskedasticity pattern; DH3: takes a value of 1 in the case of a H3 heteroskedasticity pattern; DLOGN: takes a value of 1 in the case of log-normal distribution; DTSTU: takes a value of 1 in the case of t-student distribution. DTIP: Takes a value of 1 when using the row-normalised weight matrix; $F_{VA}$: tests that all the coefficients in the regression, except the intercept, are zero (*p*-value in parentheses); *F*(tip = bin): tests that the parameters corresponding to the binary case are equal to the parameters associated with the row-standardised case (*p*-value in parentheses).
An asterisk indicates that the parameter is significant at the 5% level; *t*-ratios are calculated using the White heteroskedasticity-consistent standard errors.

The regressors are a set of dummy variables, whose interpretation is similar to that of Tables 1 and 3, the sample size, the autocorrelation coefficients ($\rho$ and/or $\theta$) and the intercept ($\gamma$). The upper part of the table contains the cases resolved with *Gets*, and the lower part contains those resolved with *Stge*. Each parameter estimated, except the intercept for the *Stge* case, γ(*Stge*), should be interpreted autonomously.

The first result that we can highlight is that the two strategies are not equal. The *F* test of the equality of estimated coefficients ($F_{(\text{gets = stge})}$ in the

**Table 4**
*GETS* versus *STGE* strategies

| | *GETS* vs. *STGE* and FFR approach. Binary weighting matrix | | | | | | | | *GETS* vs. *STGE* and FFR approach. Row-standardised weighting matrix | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SEM | | SLM | | SARAR | | SIM | | SEM | | SLM | | SARAR | | SIM | |
| | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio | Coef | *t*-ratio |
| γ(*Gets*) | 0.39* | 8.19 | 0.56* | 17.64 | −0.56* | −42.60 | 0.84* | 187.15 | 0.44* | 8.87 | 0.40* | 9.08 | −0.54* | −45.29 | 0.93* | 206.66 |
| Correl (*Gets*) | −0.82* | −16.67 | 0.05* | 2.18 | 0.16* | 10.15 | −0.02* | −2.46 | −0.94* | −16.57 | 0.27* | 5.32 | 0.21* | 14.03 | −0.02* | −2.50 |
| DH1(*Gets*) | −0.06 | −1.42 | −0.10* | −2.80 | −0.08* | −7.00 | −0.01* | −2.36 | −0.10* | −2.29 | 0.01 | 0.10 | 0.01 | 0.80 | −0.03* | −2.33 |
| DH2(*Gets*) | −0.08 | −1.90 | −0.09* | −2.73 | −0.05* | −4.40 | −0.03* | −2.82 | −0.09* | −2.16 | 0.02 | 0.34 | −0.02* | −1.98 | −0.03* | −6.86 |
| DH3(*Gets*) | −0.08* | −2.04 | 0.02 | 0.93 | 0.03* | 2.27 | −0.01 | −1.47 | −0.10* | −2.29 | 0.13* | 2.85 | 0.02 | 1.41 | 0.01 | 1.68 |
| DLOGN(*Gets*) | −0.41* | −7.14 | −0.02 | −0.72 | −0.02 | −1.64 | 0.01* | 2.93 | −0.09 | −1.87 | 0.05 | 1.09 | −0.03* | −2.33 | 0.01 | 1.14 |
| DTSTU(*Gets*) | −0.07 | −1.55 | −0.10* | −2.57 | −0.05* | −3.68 | −0.01* | −3.04 | −0.09* | −1.98 | 0.03 | 0.65 | −0.03* | −2.27 | 0.01 | 1.09 |
| *ρ*(*Gets*) | – | – | 1.15* | 8.96 | 2.55* | 47.49 | – | – | – | – | 0.50* | 11.05 | 0.70* | 54.74 | – | – |
| *θ*(*Gets*) | 1.48* | 8.72 | – | – | 3.19* | 64.48 | – | – | 0.56* | 11.57 | – | – | 0.82* | 68.65 | – | – |
| Sample size (*Gets*) | 0.00* | 3.55 | 0.00* | 5.28 | 0.00* | 36.65 | 0.00* | 4.73 | 0.00* | 2.58 | 0.00* | 2.27 | 0.00* | 31.04 | 0.00 | 0.45 |
| γ(*Stge*) | −0.05 | −0.71 | 0.06 | 1.25 | 0.29* | 15.30 | 0.07* | 11.40 | 0.01 | 0.17 | 0.10 | 1.60 | 0.37* | 18.98 | −0.01* | −2.06 |
| Correl (*Stge*) | −0.50* | −9.52 | 0.06* | 2.20 | 0.11* | 7.46 | −0.01 | −1.29 | −0.64* | −10.44 | 0.03 | 0.86 | 0.18* | 11.44 | −0.01 | −1.14 |
| DH1(*Stge*) | −0.76* | −19.01 | −0.25* | −5.05 | −0.14* | −10.94 | 0.04* | 5.89 | −0.75* | −18.55 | −0.61* | −21.64 | −0.37* | −28.22 | 0.02* | 4.81 |
| DH2(*Stge*) | −0.68* | −18.19 | −0.14* | −3.23 | −0.02 | −1.79 | 0.00 | 0.25 | −0.71* | −18.33 | −0.44* | −12.72 | −0.28* | −22.52 | 0.01 | 1.71 |
| DH3(*Stge*) | −0.08 | −1.68 | 0.02 | 0.71 | 0.00 | −0.15 | 0.00 | −0.59 | −0.06 | −1.10 | −0.04 | −0.96 | −0.01 | −0.59 | 0.01 | 1.75 |
| DLOGN(*Stge*) | −0.69* | −13.32 | −0.04 | −1.03 | −0.02 | −1.41 | 0.01* | 2.87 | −0.50* | −13.85 | −0.24* | −5.91 | −0.13* | −9.54 | 0.01* | 3.54 |
| DTSTU(*Stge*) | −0.48* | −9.91 | −0.16* | −3.48 | −0.06* | −3.99 | 0.01 | 1.28 | −0.49* | −9.06 | −0.36* | −6.80 | −0.15* | −10.14 | 0.01* | 2.40 |
| *ρ*(*Stge*) | – | – | 1.55* | 10.25 | 1.08* | 20.98 | – | – | – | – | 0.47* | 11.84 | 0.20* | 14.51 | – | – |
| *θ*(*Stge*) | 2.14* | 13.02 | – | – | 3.63* | 78.05 | – | – | 0.32* | 6.92 | – | – | 0.64* | 46.05 | – | – |
| Sample size (*Stge*) | 0.00* | 5.36 | 0.00* | 3.06 | 0.00* | 38.24 | 0.00 | 1.32 | 0.00* | 4.95 | 0.00* | 6.39 | 0.00* | 37.66 | 0.00* | −2.30 |
| *n* | 660 | | 660 | | 5400 | | 66 | | 660 | | 660 | | 5400 | | 66 | |
| Adjus. $R^2$ | 0.67 | | 0.45 | | 0.77 | | 0.93 | | 0.67 | | 0.58 | | 0.77 | | 0.49 | |
| Log like. | 57.20 | | 329.25 | | 1543.83 | | 212.77 | | 21.17 | | 177.96 | | 1704.02 | | 217.49 | |
| $F_{VA}$ | 86.52 (0.00) | | 31.33 (0.00) | | 967.30 (0.00) | | 5.63 (0.00) | | 75.42 (0.00) | | 56.39 (0.00) | | 1011.64 (0.00) | | 5.37 (0.00) | |
| *F*(*gets* = *stge*) | 63.54 (0.00) | | 1.45 (0.17) | | 49.44 (0.00) | | 7.65 (0.00) | | 60.80 (0.00) | | 21.53 (0.00) | | 202.47 (0.00) | | 8.98 (0.00) | |

See Table 1 for the notation.

table) always rejects this supposition, except when the DGP is of the SLM type and works with a binary weighting matrix. Secondly, we have already seen that the two approaches function differently depending on whether the weighting matrix is binary or row-standardised. Furthermore, looking at the results of the differential factor $\gamma(stge)$ of Table 4, we can say that, under ideal conditions, *Gets* and *Stge* are two almost indifferent strategies when the DGP is of a SEM or a SLM type; however, the *Stge* approach appears to be slightly better for SARAR cases. The last aspect of Table 4 that we wish to underline is the greater sensitivity of the *Stge* approach to the anomalies introduced into the DGP. Heteroskedasticity and non-normality of the errors have more severe consequences in this approach than in the *Gets* approach. On the contrary, the impact of endogeneity seems to be more acute in the *Gets* approach.

## 5. Conclusions

With this paper, we wish to vindicate the importance of a stage that has not occupied a very prominent position in the specification of a spatial econometric model. It should be acknowledged that, during recent years, many results on estimation and testing have been published in the literature on spatial econometrics, though references as to how to discriminate between rival models have been scarce.

Day-to-day practice has eventually consolidated an approach that works reasonably well; it is simple, easy to obtain, and quite reliable. This approach may be characterized as *Specific-to-General*, *Stge*, in the terms used by Charemza and Deadman (1997), but it is not the only available option. It is in this sense that we consider the possibilities of a strategy based on a *General-to-Specific*, *Gets*, approach, our interest justified by the fact that Florax, Folmer and Rey (2003), in a previous and extensive simulation study, found that the *Stge* strategy was preferable in a spatial context, but that the opposite tends to hold in mainstream econometrics. The debate between Hendry (2006) and Florax, Folmer and Rey (2006) encouraged our curiosity.

The main results of the Monte Carlo study conducted in our paper may be summarized as follows:

– Nowadays, there are sufficient techniques in the literature on spatial econometrics to arrange well defined *Stge* and/or *Gets* specification strategies, the results of which will depend, mainly, on the combination of techniques. In any case, in our experiment, the performance of both strategies is quite satisfactory.
– The ‘autoregressive distributed lag model of the first order’ defined by Bivand (1984) is a reasonable starting point from which to develop a *Gets* approach in a spatial context. There are some other possibilities but, in general, this model is a rather good first specification.
– The weighting matrix (row-standardised or binary) has an appreciable impact on the performance of the two strategies. In general, the results of both tend to improve when used in a DGP with a row-standardised weighting matrix.
– The sample size and the intensity of the spatial dependence relationships are crucial elements for an acceptable performance, under both approaches. From our experience, a sample size of at least 100 observations is necessary.
– Both approaches behave very well when the DGP has been specified under ideal conditions. The differences between the two are very small and not significant when the DGP is of a SEM or a SLM type, whereas *Stge* is somewhat better for SARAR cases.
– *Stge* is much more sensitive to the existence of anomalies in the DGP. Heteroskedasticity with a spatial pattern, or departures from the assumption of normality, undermine the performance of both approaches. However, *Gets* tends to be a little more robust.
– The presence of an endogenous regressor in the DGP has an ambiguous impact on the functioning of both strategies. It is extremely damaging for the right selection of SEM models, in either of the two approaches, but endogeneity may be helpful when the DGP is of a SLM or SARAR type.

Finally, we would like to emphasise the limitations of a study of this type. There are many questions that, for reasons of space and of clarity in the interpretation of the results, we have not taken into account. The discussion of these aspects, especially with regard to what happens when the DGP is not specified under ideal conditions, are part of our future research agenda.

## Acknowledgements

An earlier version of this paper was presented at the Workshop on Spatial Econometrics. Kiel Institute for the World Economy, Kiel, 8–9 April 2005. We appreciate the constructive comments received from a large number of people, including Serge Rey, Maribel Ayuda and anonymous reviewers. This article was supported by the project SEJ2006-02328/ECON of the Ministerio de Ciencia y Tecnología del Reino de España.

## Appendix A. A brief introduction to the Vuong and Clarke tests

In this Appendix, we present a brief introduction to Vuong and Clarke tests. The first may be seen as an evolution of the Cox test, whereas the second stands out for its simplicity. In both cases, we are comparing two non-nested models. The null hypothesis of the two tests is the same, i.e., that both models are indifferent (or the data cannot discriminate between them), whereas the alternative indicates that there is one model superior to the other. They differ in the form of comparing the models.

In Cox's (1961) test, there exist two families of conditioned density functions: $f_\theta = \{f_{Y|X}(\theta); \theta \in \Theta \subset \Re^p\}$ and $g_\gamma = \{g_{Y|X}(\gamma); \gamma \in \Gamma \subset \Re^q\}$, and the objective is to test one against the other. The null hypothesis corresponds to one of the families, while the other is the alternative ($H_0$:$f_\theta$ vs. $H_A$:$g_\gamma$). The Cox statistic is a centred and typified version of the traditional Likelihood Ratio:

$$C_0(f) = \frac{\mathrm{LR}_n\left(\tilde{\theta}_n; \tilde{\gamma}_n\right) - E_Z E_f\left(LR_n\left(\tilde{\theta}_n; \tilde{\gamma}_n\right)\right)}{\sqrt{V\left[\mathrm{LR}_n\left(\tilde{\theta}_n; \tilde{\gamma}_n\right) - E_Z E_f\left(LR_n\left(\tilde{\theta}_n; \tilde{\gamma}_n\right)\right)\right]}} \underset{\text{as}}{\sim} N(0,1), \tag{A.1}$$

where $\tilde{\theta}_n$ and $\tilde{\gamma}_n$ are the respective maximum likelihood (ML) estimations of $\theta$ and $\gamma$, and $\mathrm{LR}_n$ is the Likelihood Ratio: $\mathrm{LR}_n(\tilde{\theta}_n; \tilde{\gamma}_n) = L_f(\tilde{\theta}_n) - L_f(\tilde{\gamma}_n)$. As a second step, the test should be repeated using the other density function, $g_\gamma$, in the null hypothesis.

Vuong (1989) uses this approach but redefines the content of the null and alternative hypotheses. The former is now associated with a situation of indifference between the models, while the alternative is bilateral and identifies the model most favoured by the data. Analytically:

$$\left.\begin{array}{l} H_0 : E^0\left[\lg \dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right] = 0 \\ H_A : \begin{cases} H_f : E^0\left[\lg \dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right] > 0 \\ H_g : E^0\left[\lg \dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right] < 0 \end{cases} \end{array}\right\}. \tag{A.2}$$

The change of perspective allows us to obtain additional results regarding convergence in probability and in distribution, with respect to the $LR_n$ statistic of Eq. (A.1). The most important results are summarised in the following expression:

$$\left.\begin{array}{l}\sqrt{n}\left\{\dfrac{LR_n\left(\tilde{\theta}_n;\tilde{\gamma}_n\right)}{n}-E^0\left[\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right]\right\}\xrightarrow{D}N(0;\omega^2)\\ \omega^2=V^0\left[\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right]=E^0\left[\left(\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right)^2\right]-\left(E^0\left[\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right]\right)^2\end{array}\right\}. \quad (A.3)$$

Finally, Theorem 5.1 of Vuong (1989, p. 318) establishes that: "…. if $F_\theta$ and $G_\gamma$ are strictly non-nested, then:

(i) under $H_0$ : $n^{-1/2}LR_n\left[\tilde{\theta}_n;\tilde{\gamma}_n\right]/\tilde{\omega}_n\xrightarrow{D}N(0;1)$

(ii) under $H_f$ : $n^{-1/2}LR_n\left[\tilde{\theta}_n;\tilde{\gamma}_n\right]/\tilde{\omega}_n\xrightarrow{as}+\infty$

(iii) under $H_g$ : $n^{-1/2}LR_n\left[\tilde{\theta}_n;\tilde{\gamma}_n\right]/\tilde{\omega}_n\xrightarrow{as}-\infty$ (…)".

Given that the two models are different, the acceptance of the null should be interpreted as meaning that the available evidence does not permit us to discriminate between them. With slight variations, the test can also be used in the case where the models are nested, or overlapped. Finally, for the application of the test we need the full iid (independent and identically distributed log-likelihoods) clause and the habitual conditions of regularity that guarantee the existence of the ML estimators. Rivers and Vuong (2002) extend the use of the Vuong test for GMM (generalized method of moments), for non-maximum likelihood estimators, and for models with weakly-dependent heterogeneous data.

The Distribution-Free Test of Clarke is even more intuitive because it "(…) applies a modified paired sign test to the differences in the individual log-likelihoods from two non-nested models. Whereas the Vuong test determines whether or not the average log-likelihood ratio is statistically different from zero, the proposed test determines whether or not the median log-likelihood ratio is statistically different from zero. If the models are equally close to the true specification, half the individual log-likelihood ratios should be greater than zero and half should be less than zero" (Clarke, 2004, p. 6). In other words, the reasoning reproduces the discussion of Vuong, only substituting the average with the median, so that:

$$H_0:\ \text{Median}\left[\lg\frac{f(Y_t|X_t;\theta)}{g(Y_t|Z_t;\gamma)}\right]=0\Rightarrow H_0:\ \Pr\left[\lg\frac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}>0\right]=0.5 \quad (A.4)$$

$$H_A:\left\{\begin{array}{l}H_f:\ \text{Median}\left[\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right]>0\\ H_g:\ \text{Median}\left[\lg\dfrac{f(Y_t|X_t;\theta)}{g(Y_t|X_t;\gamma)}\right]<0\end{array}\right.$$

In this case, we need only: (i) to estimate the model corresponding to the family $f_\theta=\{f_{Y|X}(\theta);\theta\in\Theta\subset\Re^p\}$, retaining the individual log-likelihoods $\{f_t=\lg f(Y_t|X_t;\tilde{\theta}_n);\ t=1,2,\ldots,n\}$; (ii) to estimate the model corresponding to the family $g_\gamma=\{g_{Y|X}(\gamma);\gamma\in\Gamma\subset\Re^q\}$, maintaining the individual log-likelihoods $\{g_t=\lg g(Y_t|X_t;\tilde{\gamma}_n);\ t=1,2,\ldots,n\}$; (iii) to obtain the differences between the log-likelihoods $\{d_t=f_t-g_t;\ t=1,2,\ldots,n\}$; and, (iv) to solve the test through a Binomial ($n$;0.5) distribution using $B$, the number of positive differences $d_t$, as the testing statistic. The conditions for using the Clarke test are weak: only independence between the log-likelihood ratios is required, and, also, that each of them comes from a continuous population ('not necessarily the same,' Clarke, 2004, p.6).

Obviously, in both cases, the requirement of independence between the individual log-likelihoods can be problematic if we are thinking of comparing the SLM of Eq. (3) against the SEM of Eq. (2): these models are characterised by cross-sectional dependence. However, to circumvent this clause, it is sufficient to filter the variables of the model, using the eigenvectors of the weighting matrix $W$, assuming that this matrix is symmetric (for the details, see Trivez and Mur, 2004). If matrix $W$ is not symmetric, the filtering process does not work and the tests of Vuong and of Clarke are not applicable to this case.

## Appendix B. Misspecification tests used in the analysis

The four tests described in Eqs. (B.1)–(B.4) refer to a static model, such as: $y=X\beta+u$. This model has been estimated by least-squares (LS), where $\hat{\sigma}^2$ and $\hat{\beta}$ correspond to the LS estimations, and $\hat{u}$ corresponds to the residual series. These tests are as follows (see Florax and de Graaff, 2004, for the details):

$$\text{LMERR Test}:\quad \text{LM-ERR}=\left(\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2\frac{1}{T_1};\quad T_1=tr[W'W+WW] \quad (B.1)$$

$$\text{LMEL Test}:\quad \text{LM-EL}=\frac{\left(\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}-\frac{T_1}{R\hat{J}_{\rho-\beta}}\frac{u'Wy}{\hat{\sigma}^2}\right)^2}{\frac{T_1}{R\hat{J}_{\rho-\beta}}\left(R\hat{J}_{\rho-\beta}-T_1\right)} \quad (B.2)$$

$$\text{LMLAG Test}:\quad \text{LM-LAG}=\frac{1}{R\hat{J}_{\rho-\beta}}\left(\frac{\hat{u}'Wy}{\hat{\sigma}^2}\right)^2 \quad (B.3)$$

$$\text{LMLE Test}:\quad \text{LM-LE}=\frac{\left(\frac{\hat{u}'Wy}{\hat{\sigma}^2}-\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2}{R\hat{J}_{\rho-\beta}-T_1}. \quad (B.4)$$

Moreover, $R\hat{J}_{\rho-\beta}=T_1+(\hat{\beta}'X'WMWX\hat{\beta})/\hat{\sigma}^2$ and $M=[I-X(X'X)^{-1}X']$. The next two tests use the ML estimation of a spatial model. The first refers to an SLM model, and tests whether there is residual spatial dependence in the errors of the equation:

$$\text{RS}_{\lambda|\rho}\ \text{Test}:\quad \text{RS}_{\lambda|\rho}=\frac{\left(\frac{\tilde{u}'W\tilde{u}}{\tilde{\sigma}^2}\right)^2}{T_1-T_A^2V[\tilde{\rho}]}, \quad (B.5)$$

where $T_A=tr[(WW+W'W)(I-\tilde{\rho}W)^{-1}]$ and $V(\tilde{\rho})$ is the ML estimation of the variance of the parameter $\rho$ in Eq. (3), under the null hypothesis of no residual spatial dependence (Anselin and Bera, 1998). The second test refers to an SEM model, and tests whether we have omitted a spatial lag of the endogenous variable in the main equation of the model:

$$\text{RS}_{\rho|\lambda}\ \text{Test}:\quad \text{RS}_{\rho|\lambda}=\frac{\left[\tilde{u}'\left(I-\tilde{\theta}W\right)'\left(I-\tilde{\theta}W\right)Wy\right]^2}{H_{\lambda\lambda}}, \quad (B.6)$$

where $H_{\lambda\lambda}$ is the ML estimation of the variance of the restriction ($\rho=0$) obtained in the model of Eq. (2), and under the null hypothesis of no spatial lag in the main equation. The six Lagrange Multipliers, Eqs. (B.1)–(B.6), have an asymptotic $X^2(1)$ distribution.

## Appendix C. The *Hybrid* and the *Classical* strategy of FFR (2003)

These strategies differ in the way that they solve the doubts that appear when the LMERR and the LMLAG are statistically significant. In the so-called 'classical strategy,' we compare the value of the two statistics, in order to select the model associated with the more significant value: the SEM if LMERR>LMLAG and the SLM if LMERR<LMLAG. The *hybrid* strategy replicates the previous one, but uses the *robust* Multipliers: the LMEL and the LMLE, instead of the LMERR and the LMLAG. Below, we show that both inequalities are equivalent. That is: LMERR>LMLAG ⇔ L-MEL>LMLE. The proof is simple, using the expressions of Appendix B:

$$\text{LMEL}>\text{LMLE}\Rightarrow\frac{\left(\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}-\frac{T_1}{R\hat{J}_{\rho-\beta}}\frac{\hat{u}'Wy}{\hat{\sigma}^2}\right)^2}{T_1-\frac{T_1^2}{R\hat{J}_{\rho-\beta}}}>\frac{\left(\frac{\hat{u}'Wy}{\hat{\sigma}^2}-\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2}{R\hat{J}_{\rho-\beta}-T_1} \quad (C.1)$$

Simplifying common terms, Eq. (C.1) reads as:

$$\Rightarrow \frac{R\hat{J}_{\rho-\beta}}{T_1}\left(\hat{u}'W\hat{u}-\frac{T_1}{R\hat{J}_{\rho-\beta}}\hat{u}'Wy\right)^2 > \left(\hat{u}'Wy-\hat{u}'W\hat{u}\right)^2 \tag{C.2}$$

After solving for the squares and grouping terms, we obtain:

$$\Rightarrow \frac{R\hat{J}_{\rho-\beta}}{T_1}\left(\hat{u}'W\hat{u}\right)^2 > \left(\hat{u}'Wy\right)^2 \tag{C.3}$$

That is:

$$\begin{aligned} &\text{LMEL} > \text{LMLE} \Rightarrow \frac{1}{T_1}\left(\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2 > \frac{1}{R\hat{J}_{\rho-\beta}}\left(\frac{\hat{u}'Wy}{\hat{\sigma}^2}\right) \\ &\text{LMEL} > \text{LMLE} \Leftrightarrow \text{LMERR} > \text{LMLAG} \end{aligned} \tag{C.4}$$

## References


Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In Petrov, B. and F. Cáski (eds.) Second International Symposium on Information Theory, pp. 267–281. Budapest: Akademiai Kaidó. Reprinted in Kort, S. and N. Johnson (eds): Breakthroughs in Statistics, volume I, pp. 599–624. Berlin: Springer.

Anselin, L., 1986. Non-nested tests of the weight structure in spatial autoregressive models: some Monte Carlo results. Journal of Regional Science 26, 267–284.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer, Dordrecht.

Anselin, L., Florax, R., 1995. Small sample properties of tests for spatial dependence in regression models: some further results. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Springer, Berlin, pp. 21–74.

Anselin, L., Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 237–289.

Anselin, L., Bera, A., Florax, R., Yoon, M., 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26, 77–104.

Anselin, L., Florax, R., Rey, S., 2004. Econometrics for spatial models: recent advances. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics. Springer, New York, pp. 1–4.

Aznar, A., 1989. Econometric Model Selection: A New Approach. Kluwer, Dordrecht.

Baltagi, B., Li, D., 2001. Double length artificial regression for testing spatial dependence. Econometric Reviews 20, 31–40.

Bivand, R., 1984. Regression modelling with spatial dependence: an application of some class selection and estimation methods. Geographical Analysis 16, 25–37.

Blommestein, H., 1983. Specification and estimation of spatial econometric models. A discussion of alternative strategies for spatial economic modelling. Regional Science and Urban Economics 13, 251–270.

Box, G., 1980. Sampling and Bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society. Series A 143, 383–430.

Burnham, K., Anderson, D., 2002. Model Selection and Multimodel Inference, 2nd edition. Springer, Berlin.

Burridge, P., 1981. Testing for a common factor in a spatial autoregression model. Environment and Planning A 13, 795–800.

Campos, J., Ericsson, N., Hendry, D., 2005. General-to-Specific Modelling. Edward Elgar, London.

Charemza, W., Deadman, D., 1997. New Directions in Econometric Practice. Edward Elgar, Cheltenham.

Chow, A., 1983. Econometrics. McGraw-Hill, New York.

Clarke, K., 2003. Nonparametric model discrimination in international relations. Journal of Conflict Resolution 47, 72–93.

Clarke, K., 2004. A simple distribution-free test for nonnested hypothesis. Working Paper. Department of Political Science, University of Rochester.

Cox, D., 1961. Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 105–123.

Cox, D., 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society. Series B 24, 406–424.

Cox, D., Wermuth, N., 1996. Multivariate dependencies. Models, Analysis and Interpretation. Chapman & Hall, London.

Danilov, D., Magnus, J., 2004. On the harm that ignoring pretesting can cause. Journal of Econometrics 122, 27–46.

Dastoor, N., McAleer, M., 1989. Some new comparisons of joint and paired tests for non-nested models under local hypotheses. Econometric Theory 5, 83–94.

Davidson, J., 2000. Econometric Theory. Blackwell, Oxford.

Davidson, R., MacKinnon, J., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 88, 781–793.

Fisher, G., McAleer, M., 1979. On the interpretation of the Cox tests in econometrics. Economics Letters 4, 145–150.

Florax, R., de Graaff, T., 2004. The performance of diagnostics tests for spatial dependence in linear regression models: a meta-analysis of simulation studies. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Berlin, pp. 29–65.

Florax, R., Folmer, H., Rey, S., 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. Regional Science and Urban Economics 33, 557–579.

Florax, R., Folmer, H., Rey, S., 2006. A comment on specification searches in spatial econometrics: the relevance of Hendry's methodology: a reply. Regional Science and Urban Economics 36, 300–308.

Fotheringham, A., Charlton, M., Brunsdon, C., 1999. Geographically weighted regression. A natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30, 1905–1927.

Gourieroux, C., Monfort, A., 1989. Statistique et modèles économétriques. Economica, Paris.

Granger, C., King, M., White, H., 1995. Comments on testing and the use of model selection criteria. Journal of Econometrics 67, 173–187.

Hausman, D., 1992. The Inexact and Separate Science of Economics. Cambridge University Press, Cambridge.

Hendry, D., 2006. A comment on "Specification searches in spatial econometrics: the relevance of Hendry's methodology". Regional Science and Urban Economics 36, 309–312.

Hendry, D., Richard, J., 1982. On the formulation of empirical models in dynamic econometrics. Journal of Econometrics 20, 3–33.

Hendry, D., Richard, J., 1989. Recent developments in the theory of encompassing. In: Cornet, B., Tulkens, H. (Eds.), Contributions to Operations Research and Economics. MIT Press, Cambridge, pp. 393–440.

Hendry, D., Krolzig, H., 2005. The properties of automatic gets modelling. Economic Journal 115, 32–61.

Hepple, L., 1995a. Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds. Environment and Planning A 27, 447–469.

Hepple, L., 1995b. Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms. Environment and Planning A 27, 615–644.

Hoover, K., Perez, S., 2004. Truth and robustness in cross-country growth regressions. Oxford Bulletin of Economics and Statistics 66, 765–798.

King, M., McAleer, M., 1987. Further results on testing AR (1) against MA (1) disturbances in the linear regression model. Review of Economic Studies 54, 649–663.

Leamer, E., 1978. Specification Searches: Ad-hoc Inference with Nonexperimental Data. John Willey, New York.

LeSage, J., 1997. Bayesian estimation of spatial autoregressive models. International Regional Science Review 20, 113–129.

LeSage, J., 2004. A family of geographically weighted regression models. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Berlin, pp. 241–264.

Lütkepohl, H., 2007. General-to-specific or specific-to-general modelling? An opinion on current econometric terminology. Journal of Econometrics 136, 319–324.

Morgan, M., 1990. The History of Econometric Ideas. Cambridge University Press, Cambridge.

Mur, J., Lopez, F., Angulo, A., 2008. Symptoms of instability in models of spatial dependence. Geographical Analysis 20, 189–211.

Pace, K., LeSage, J., 2004. Spatial autoregressive local estimation. In: Getis, A., Mur, J., Zoller, H. (Eds.), Spatial Econometrics and Spatial Statistics. Palgrave, Londres, pp. 31–51.

Popper, K., 1979. Objective knowledge. An Evolutionary Approach. Oxford University Press, Oxford.

Ripley, B., 1996. Pattern Recognition and Neuronal Networks. Cambridge University Press, Cambridge.

Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. Econometrics Journal 5, 1–39.

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, 1273–1291.

Schwarz, C., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Spanos, A., 1986. Statistical Foundations of Econometric Modelling. Cambridge University Press, Cambridge.

Spiegelhalter, D., Best, J., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B 64, 583–639.

Trivez, J., Mur, J., 2004. Some proposals for discriminating between spatial processes. In: Getis, A., Mur, J., Zoller, H. (Eds.), Spatial Econometrics and Spatial Statistics. Palgrave, Houndmills, pp. 150–175.

Vuong, Q., 1989. Likelihood ratio-tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Zellner, A., 1978. Jeffreys–Bayes posterior odds ratio and the Akaike information criterion for discriminating between models. Economic Letters 1, 337–342.