A closer look at the Spatial Durbin Model

Article · September 2005

Source: RePEc

2 authors:

Jesus Mur
University of Zaragoza

Ana Angulo
University of Zaragoza

EUROPEAN REGIONAL SCIENCE ASSOCIATION.

45th Congress. Amsterdam. August, 23-27, 2005

A closer look at the Spatial Durbin Model

Jesús Mur (*)

Ana Angulo (**)

Department of Economic Analysis

University of Zaragoza.

Gran Vía, 2-4 (50005) Zaragoza, (SPAIN)

Phone: +34-976-761815

(*) jmur@unizar.es

(**) aangulo@unizar.es

Abstract

The spatial Durbin model occupies a key position in Spatial Econometrics. It is the reduced form of a model with cross-sectional dependence in the errors, but it may also be used as the nesting model in a more general approach of model selection. In the first case, it is the equation where we solve the Likelihood Ratio test of Common Factors, LRCOM. The objective in this case is to discriminate between substantive and residual dependence in an, apparently, misspecified equation. Our paper tries to go further into the interpretation of this intermediate equation in both aspects. We include a small Monte Carlo study related to the LRCOM test and present some new results that facilitate the use, and the interpretation, of the Durbin equation in a more general context of discriminating between econometric models in a spatial context.

Keywords: Likelihood Ratio; Spatial Lag Model; Spatial Error Model

JEL Classification: C21, C50, R15

1. Introduction

In recent years, there has been an increase in the concern about questions related to methodology in Spatial Econometrics. The works of Anselin and Florax (1995), Anselin et al. (1996) and Anselin and Bera (1998) have played a fundamental role in the revitalisation of this concern. These works deal, among other things, with the lack of specificity of the traditional tests based on the principle of the Lagrange Multiplier and, consequently, with the difficulty of finding the true model when there are various alternatives. Particularly, they demonstrate that there is a very real probability of obtaining a misspecification if the analyst is not sufficiently careful with the method.

Later, Florax, Folmer and Rey (2003) carried out a more systematic approach to the root of the problem, comparing the results of the methodology of Hendry with those of other more traditional techniques in the field of spatial econometrics. Surprisingly, the latter seem to work better. Mur and Angulo (2005) follow the same lines introducing new elements into a discussion that appears to be more extensive / (an, apparently, more open disccussion). Dubin (2003) focuses on the robustness of various models of interaction to capture structures of spatial autocorrelation in the error term of the equation, in an approach that is reminiscent of that of Florax and Rey (1995). McMillen (2003) demonstrates that, beneath an apparent problem of autocorrelation, what the analyst may really be finding is a misspecification of the model: *'yet autocorrelation is often produced spuriously by model misspecification*' (p.215) is one of his conclusions.

Finally, López-Bazo and Fingleton (2004) openly question the real validity of specifications with structures of dependence in the error term of the equation. According to them, if the externalities that dominate in a spatial context are as Anselin (2003) describes them, it is reasonable to find substantive cross-sectional dependence relationships (that is, where lags in the endogenous variable are necessary on the right-hand side of the equation). However, what dominates in most of the cases analysed by the authors are mechanisms of residual dependence. Using particular but relevant modelling problems, López-Bazo and Fingleton arrive at the same conclusion reached previously by McMillen: really, what habitually underlies models with residual autocorrelation is a problem of misspecification of the equation due, in general, to the omission of relevant variables on the right-hand side.

In this paper, we want to contribute further evidence towards the debate on methodology in Spatial Econometrics that we have hurriedly outlined above. Our objective, in particular, focuses on what the literature calls the '*Spatial Durbin Model*' and on the test of common factors, LRCOM. Both elements are intimately linked, as is remarked in Section 2, and they can become very helpful instruments in the process of specifying an econometric model. In Section 3, we present the main results obtained from a Monte Carlo experiment on the LRCOM test. The paper finishes with a section of conclusions.

2. The Spatial Durbin Model and the test of Common Factors

The Durbin model arises in a very specific context in which, using time series, we need to estimate an econometric model with an AR(1) error term:

$$\left.\begin{array}{l} y_t = x_t'\beta + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{array}\right\} \tag{1}$$

Cochrane and Orcutt (1949) had proposed a stepwise algorithm using successive LS estimations in semi-differentiated variables, $y_t - \rho^{(r-1)} y_{t-1} = \left[x_t - \rho^{(r-1)} x_{t-1}\right]'\beta + \varepsilon_t$, where $\rho^{(r-1)}$ is the estimation of ρ obtained in the (r-1)th iteration. The process, implicitly, begins with a value of ρ equal to zero, which can lead to problems of consistency in some cases. To avoid such inconveniences, Durbin (1960) suggested directly estimating the reduced unrestricted form of (1) by LS:

$$y_t = \rho_{t-1} + x_t'\beta + x_{t-1}'\theta + \varepsilon_t \tag{2}$$

This alternative already guarantees consistent estimators in the first step and is much simpler. The adaptation of these results to the spatial case does not involve any complexity, as Anselin (1980) proposed:

$$\left.\begin{array}{l} y = x\beta + u \\ u = \rho Wu + \varepsilon \end{array}\right\} \Rightarrow y = \rho Wy + x\beta + Wx\theta + \varepsilon \tag{3}$$

where W is the weighting matrix; y, u and ε are vectors of order (Rx1); x is the (Rxk) matrix of observations of the explicative variables; β a (kx1) vector of parameters and ρ the autoregressive parameter of the SAR(1) process that intervenes in the random term of (3).

Following the initial proposal of Durbin, the next step should be the estimation of the reduced form of (3), which creates serious difficulties. The problem with the last expression is that, unlike what happens in (2), the equation cannot be estimated by LS because there is an endogenity relationship: the regressor Wy is contemporaneously dependent on the error term, ε (y_{t-1} is predetermined in (2), which results in biased but still consistent estimators). In this sense, the Durbin model does not help to simplify the problem of the estimation of the autoregressive parameter. However, the equation is useful for us because it highlights other questions relevant to the specification exercise.

Firstly, it clearly shows why substantive spatial dependence tests (the LM-LAG, for example) have so much power when they act in static models in which an autoregressive, SAR(1), process was present in the error term: the reduced form of both types of models is the same, except for the term $Wx\theta$. The opposite is equally applicable: residual dependence tests (the LM-ERR, for example) have high power when they are applied to dynamic structures whose error term is white noise. The already cited works of Anselin and Florax (1995) and of Anselin et al. (1996) try precisely to correct this source of uncertainty.

Another aspect to note is that in (3) it must hold that $\theta=-\rho\beta$. That is, if in the unrestricted equation of (3) we cannot reject the set of k non-linear restrictions, the evidence points to a static process with an SAR(1) error term. The resulting test is the LRCOM of Burridge (1981), which is specified as a traditional Likelihood Ratio test. In order to obtain it, it will be necessary to estimate the ample model by ML:

$$\left.\begin{array}{l} y = \rho Wy + x\beta + Wx\theta + \varepsilon \\ \varepsilon \sim N(0,\sigma^2 I) \end{array}\right\} \tag{4}$$

where, for simplicity, we assume normality in the error term. The log-likelihood function of the model is standard:

$$l(y/\varphi_A) = -\frac{R}{2}\ln 2\pi - \frac{R}{2}\ln\sigma^2 - \frac{\left[By - x\beta - Wx\theta\right]'\left[By - x\beta - Wx\theta\right]}{2\sigma^2} + \ln\left|B\right| \tag{5}$$

with $\varphi_A = [\beta,\theta,\rho,\sigma^2]'$; |B| is the determinant of the Jacobian term whose logarithm is easily obtained: $\ln|B| = |I - \rho W| = \sum_{r=1}^{R}(1-\rho\lambda_r)$, with $\{\lambda_r$, r=1, 2, ..., R$\}$ being the set of R eigenvalues of matrix W. The optimisation of (5) presents no difficulties.

Let's look now at the restricted model:

$$\left.\begin{array}{l} y = x\beta + u \\ u = \rho W u + \varepsilon \end{array}\right\} \tag{6}$$

whose log-likelihood function is also known:

$$l(y/\varphi_0) = -\frac{R}{2}\ln 2\pi - \frac{R}{2}\ln\sigma^2 - \left[\frac{(y-x\beta)'B'B(y-x\beta)}{2\sigma^2}\right] + \ln|B| \tag{7}$$

with $\varphi_0 = [\beta,\rho,\sigma^2]'$. Formally, the LRCOM test can be expressed as:

$$\left.\begin{array}{l} H_0: \rho\beta + \theta = 0 \\ H_A: \rho\beta + \theta \neq 0 \end{array}\right\} \Rightarrow \text{LRCOM} = 2\left[l(y/\tilde{\varphi}_A) - l(y/\tilde{\varphi}_0)\right] \sim \chi^2(k) \tag{8}$$

If the null hypothesis is accepted, Model (6) should be specified, while we will maintain the unrestricted dynamic model, (4), if this hypothesis is not admissible. So that this test does not give rise to errors, it should be applied once the habitual tests of autocorrelation, residual or substantive, have allowed the rejection of the static model without spatial effects[1]. That is, the LRCOM test is useful in cases in which there is enough evidence to maintain that the parameter ρ is different to zero; the problem is that we do not know exactly in which specification, that of (4) or that of (6).

The situation described overlaps with that corresponding to the LM-LE and LM-EL tests, as shown in Anselin et al. (1996), which are robust to local misspecification errors in the null hypothesis. However, these tests pose an additional operative problem which is that, to reach a definitive conclusion as to the type of spatial effects in the specification, there must not be any contradictions between them (for example, accepting, or rejecting, simultaneously their respective null hypotheses). This circumstance, according to the results presented in Mur and Trívez (2003) or in Mur and

---

[1] In that case, if ρ and θ are equal to zero, both in (4) and in (6), the null hypothesis will continue to hold.

Angulo (2005), cannot be totally discarded. In any case, these tests (the LRCOM and the LM-LE and LM-EL) are instruments that should be used complementarily because they exploit different dimensions of the same problem.

Returning to the Likelihood Ratio, neither is the LRCOM free of difficulties and in no case will it be a final test. For example, if the test leads us to the dynamic model of (4), it seems reasonable to continue the testing process by proposing the subordinate test as to whether external effects associated only with the exogenous variables intervene in the equation. That is, if vector θ is equal to zero or not.

On the other hand, if the null hypothesis of the LRCOM test is accepted, we must consider the additional problem of which is, really, the process that should be introduced into the error term. In (6), an SAR(1) process has been specified, but there is no reason for excluding alternatives such as the moving average type, given that they will produce structures of spatial dependence with similar symptoms. Indeed, using the result of Bivand (1984):

$$\left(I-\rho W\right)^{-1}=\sum_{j=1}^{\infty}\left(-1\right)^{j}\left(\rho W\right)^{j}=I-\rho W+\left(\rho W\right)^{2}-\left(\rho W\right)^{3}+\cdots \qquad (9)$$

it is evident that an SAR(1) process admits an SMA(∞) representation. Conversely, an SMA(1) process equally admits an SAR(∞) representation, a situation that the LRCOM will tend to identify as a generic problem of residual dependence in a static model. In sum, also under the null hypothesis a discrimination exercise should be carried out between, at least, two alternatives, a structure of autoregressive or of moving average dependence.

To conclude this section on the Durbin model and the LRCOM test, we want to incorporate a series of additional arguments that situate the discussion in its proper place. We are, really, dealing with a problem of model selection, which makes it advisable to adopt a broader perspective. For example, the nesting model employed up to now is that specified in (4), which is '*an autoregressive distributed lag model of the first order*', ADL(1,1), in terms of Bivand (1984, p.27). Blommestein (1983) had previously advanced in this same direction suggesting the convenience of adopting an approach like Hendry's, structured in a sequence of ADL(m,n) processes, in spite of '*the increasing complexity (in) specifying higher order (>1) spatial lags in the case of non-*

binary weights and/or irregular lattices' (p. 259). Blommestein's proposal, to initiate the process using an ADL(m,n) sufficiently general so as to try to simplify it through successive nesting tests, is not easily operative in a spatial context.

However, for us, this proposal has some interest because it underlines the additional necessity of checking, as said before, the dynamics of the model to which the LRCOM test has led us. For example, if the data come from an ADL(2,2):

$$\left.\begin{array}{l} y = \rho_1 W_1 y + \rho_2 W_2 y + x\beta + W_1 x\theta_1 + W_2 x\theta_2 + \varepsilon \\ \varepsilon \sim N(0,\sigma^2 I) \end{array}\right\} \tag{10}$$

in which W_1 and W_2 are two weighting matrixes; ρ_1 and ρ_2 two autoregressive parameters and θ_1 and θ_2 vectors of parameters associated to the spatial lags of the exogenous variables. It is relatively simple to show that if, in (10), we introduce a restriction of common factors such as: θ_1=-$\rho_1\beta$, we will obtain a model with an SAR(1) error term in which dynamic elements will continue to act in the main equation:

$$\begin{array}{l} [I - \rho_1 W_1] y = \rho_2 W_2 y + [I - \rho_1 W_1] x\beta + W_2 x\theta_2 + \varepsilon \\ \quad \Rightarrow y = \rho_2 [I - \rho_1 W_1]^{-1} W_2 y + [I - \rho_1 W_1]^{-1} W_2 x\theta_2 + [I - \rho_1 W_1]^{-1} \varepsilon \\ \quad\quad \Rightarrow y = \rho_2 W^* y + W^* x\theta_2 + [I - \rho_1 W_1]^{-1} \varepsilon \end{array} \tag{11}$$

where $W^* = [I - \rho_1 W_1]^{-1} W_2$. The problem will be to find a good approximation to W*. In any case, it is clear that the reduced form of (11) has a dynamic nature with an autoregressive structure in the random term. On the other hand, if we assume as our starting point a static model such as that of (6), but with a moving average random term, we arrive at an ADL(1,1) with an autocorrelated error term:

$$\left.\begin{array}{l} y = x\beta + u \\ u = \varepsilon + \rho W \varepsilon \end{array}\right\}$$

$$\begin{array}{l} u = \varepsilon + \rho W \varepsilon \Rightarrow u = \varepsilon + \rho W u + \rho^2 W^2 u + \cdots \\ \quad \Rightarrow u = \varepsilon + \rho W u + v \end{array} \tag{12}$$

$$\Rightarrow \left.\begin{array}{l} y = x\beta + u = \rho W y + x\beta + W x\theta + \omega \\ \omega = \varepsilon + \pi_1^* W \varepsilon + \pi_2^* W^2 \varepsilon + \cdots \approx \rho^* W \omega + \varepsilon \end{array}\right\}$$

To sum up, if the LRCOM has led us to the dynamic model of (4), it should be tested, among other things, that the error term of the equation is a white noise, using, for example, the Lagrange Multiplier $RS_{\lambda|\rho}$ of Anselin and Bera (2003). Equally, if the LRCOM leads us to a static model with an apparent SAR(1) structure in the random term, the necessity of also including dynamic elements in the main equation should be tested through the $RS_{\rho|\lambda}$ test of Anselin and Bera (2003), also based on the Lagrange Multiplier.

3. The LRCOM in small to medium sample sizes. A Monte Carlo approach.

In the previous section we have insisted on the important role that the LRCOM test can play in the specification process of a cross-sectional econometric model. As has been said, it is not a final test but only signposts the appropriate direction to take in the specification of the model. Furthermore, the indications of this test should be complemented with those of other instruments before reaching a definitive solution. In short, it is an additional test that occupies a strategic position in the whole process.

For this reason, the scant attention that has been paid to it in the specialised literature seems a little surprising. To cite only the most recent cases, it is not included in the comprehensive simulation exercise carried out by Anselin and Florax (1995), nor is it mentioned in the meta-analysis of Florax and de Graaff (2004) nor in the manuals of Tiefelsdorf (2000) and Griffith (2003). This section aims to partially correct this deficiency by resolving a Monte Carlo exercise in order to be able to evaluate the behaviour of the LRCOM test under different configurations relative to both the data generation mechanism and the sample size.

In this exercise we have taken a simple linear model as a point of reference:

$$y_r = \alpha + \beta x_r + u_r \tag{13}$$

From this, it is straightforward to obtain a Spatial Lag Model, SLM, or a Spatial Error Model, SEM. In matrix terms:

$$\text{SLM}: \begin{cases} y = \rho W y + x\beta + u \\ u \sim iid\left(0;\sigma^2 I\right) \end{cases} \qquad \text{SEM}: \begin{cases} y = x\beta + u \\ u = \rho W u + \varepsilon \\ \varepsilon \sim iid\left(0;\sigma^2 I\right) \end{cases} \tag{14}$$

The main characteristics of the exercise are the following:

(a)- We have used two pairs of values for α and β in (13). The first, (α=10; β=0.5), guarantees an average R^2 of 0.2 without spatial effects while the second, (α=10; β=2.0), raises it to 0.8.

(b)- The observations of x and of the random terms ε and u proceed from a univariate normal distribution with mean zero and a variance, σ^2, equal to one in all cases.

(c)- We have used three different sample sizes: 25, 100 and 225.

(d)- The contiguity matrix, W, has always been specified as of binary type using a rook scheme in a regular lattice of (5x5), (10x10) or (15x15).

(e)- The range of values for parameter ρ depends on the contiguity matrix used in each case. For the matrix of the (5x5) system, the interval is (-0.274; 0.274), for the (10x10) it is (-0.248; 0.248) and for the (15x15) it is (-0.229; 0.229). In each case, 40 values of the parameter, distributed regularly over the whole interval, have been simulated.

(f)- Each combination has been repeated 1000 times.

Next, we will present the main results obtained from the simulation, structured in three cases of interest. In the first two, one of the two models, the SEM or the SLM, was the true one (that is, the data were generated with one of them, the *true model*, whereas the other is, obviously, *false)*. Lastly, the third case is characterised by the true model not belonging to the set of alternatives, because the data were generated with a mixed model:

$$\left.\begin{array}{c} y = \rho_1 Wy + x\beta + u \\ u = \rho_2 Wu + \varepsilon \\ \varepsilon \sim iid\left(0;\sigma^2 I\right) \end{array}\right\} \qquad (15)$$

However, the catalogue of possible decisions is limited to the SEM or to the SLM, both false in this third case.

In all the cases we begin by specifying a model without spatial effects, like that of (13), and we look for evidence of misspecification to try to find the correct model that generated the data. Figure 1 summarises the results obtained in the first case where the data have been generated using an SEM model, whereas Figure 2 presents the results observed in the case of the SLM.

In the figures we show the average R^2 obtained for the 1000 simulations corresponding to each case, the percentage of rejections of the null hypothesis of the absence of spatial effects obtained with the SARMA test and the percentage of rejections of the null hypothesis corresponding to the LRCOM test. The terms '(h)' or '(l)' next to the corresponding sign indicate that these data come from a simulation with a high (h) or a low (l) signal-to-noise ratio. Furthermore, 'unc' or 'con' means that the LRCOM test has been obtained Unconditionally, 'unc', or Conditionally, 'con', with respect to the SARMA test. In the first case, Unconditional approach, the LRCOM test has always been obtained, whether there was evidence of misspecification in the equation or not. In the second case, Conditional approach, the LRCOM test has only been obtained for those cases in which the SARMA test has previously rejected the absence of spatial effects in the estimated equation (that of 13). This means that the percentage represented in the series of the conditional approach comes from a variable number of cases.

The results for the third case appear, in an abridged version, in Table 2. Here we summarise only the percentage of rejections of the null hypothesis of the LRCOM when the data have been generated with the mixed model of (15) and a high signal-to-noise ratio. In a sense, we could say that these percentages measure the propensity to accept an SLM instead of an SEM, when the right model is a mixture of both. Horizontally, we reproduce the range of ρ_1 (the coefficient of Wy in the main equation of 15) and vertically the range of ρ_2 (the autocorrelation coefficient of the error term).

Finally, in Table 1, we present the results obtained when no spatial effects have intervened in the DGP.

TABLE 1: Percentage of rejections of the null hypothesis of each test.

Significance level: 5%

DGP: SEM MODEL with ρ=0 in (14)								
R=25	**High R^2**	**Low R^2**	**R=100**	**High R^2**	**Low R^2**	**R=225**	**High R^2**	**Low R^2**
SARMA	0.046	0.045	**SARMA**	0.049	0.052	**SARMA**	0.060	0.060
LR (unc)	0.075	0.094	**LR (unc)**	0.060	0.057	**LR (unc)**	0.040	0.050
LR (con)	0.370	0.578	**LR (con)**	0.458	0.387	**LR (con)**	0.167	0.667
DGP: SLM MODEL with ρ=0 in (14)								
R=25	**High R^2**	**Low R^2**	**R=100**	**High R^2**	**Low R^2**	**R=225**	**High R^2**	**Low R^2**
SARMA	0.041	0.044	**SARMA**	0.045	0.045	**SARMA**	0.055	0.040
LR (unc)	0.068	0.083	**LR (unc)**	0.050	0.055	**LR (unc)**	0.060	0.051
LR (con)	0.439	0.568	**LR (con)**	0.457	0.422	**LR (con)**	0.471	0.495
DGP: MIXED MODEL with ρ_1= ρ_2=0 in (15)								
R=25	**High R^2**	**Low R^2**	**R=100**	**High R^2**	**Low R^2**	**R=225**	**High R^2**	**Low R^2**
SARMA	0.052	0.084	**SARMA**	0.044	0.068	**SARMA**	0.041	0.049
LR (unc)	0.062	0.045	**LR (unc)**	0.060	0.047	**LR (unc)**	0.050	0.055
LR (con)	0.598	0.124	**LR (con)**	0.677	0.388	**LR (con)**	0.614	0.686

The percentages of the SARMA test correspond to the level of significance estimated for a theoretical level of 5%. The estimates are in the proximity of that value and, generally, within the theoretical interval, (0.036;0.064) for 1000 replications. As was foreseeable, the estimated size of the test tends to stabilise at the theoretical level of 5%, as the number of observations in the sample increases.

The situation is a little more complex in the LRCOM test. The null hypothesis of the test holds for all the cases of the unconditional approach included in Table 1, (if there are no spatial effects, parameter ρ and vector θ will be zero in expression (8) so the restriction will be verified: $\rho\beta$=- θ =0). Obviously, the null hypothesis also holds in all the cases contemplated in Figure 1 because the data have been obtained by using an SEM model. In both situations, the level of significance estimated for the LRCOM test tends to adjust to the theoretical value, with some small anomalies when the sample is of reduced size and the parameter of autocorrelation takes high values (case 1.A in Figure 1).

The results corresponding to the LRCOM in Table 1, in a conditional approach, bear no direct relation to the concept of level of significance because the series were generated without spatial effects. The results show a certain predisposition toward SLM structures. It should be noted that these estimations are obtained with a very reduced number of cases: $1000 \cdot \hat{\varepsilon}$ (where $\hat{\varepsilon}$ is the level of significance estimated for the SARMA test).

That deficiency of information decreases as the parameter of autocorrelation takes higher values, given that the SARMA test detects a greater number of cases with spatial effects, as can be seen in Figure 1. The consequence is that the behaviour of the LRCOM test under both approaches, unconditional and conditional, tends to converge: the inverted U that characterises the series of the unconditional approach narrows as the sample size increases. In Case 1.C with a sample of 225 observations, the series of rejections corresponding to the unconditional approach and a high signal-to-noise ratio maintains a small difference with respect to the series that comes from the conditional approach. The discrepancy oscillates between 3 and 4 points, although it is unexpected and appears stable.

The the power function of the test corresponds to the data reproduced in Figure 2, in which an SLM has intervened in the DGP, and to those of Table 2, obtained with a mixed process. With respect to both cases, we wish to underline the following aspects:

TABLE 2. LRCOM in the case of a mixed DGP. High Signal-to-Noise Ratio.

CASE 3.A: Sample Size is 25

UNCONDITIONAL APPROACH

	ρ_1													
ρ_2	**-0.25**	**-0.21**	**-0.17**	**-0.13**	**-0.10**	**-0.06**	**-0.02**	**0.02**	**0.06**	**0.10**	**0.13**	**0.17**	**0.21**	**0.25**
-0.25	0.71	0.74	0.80	0.79	0.66	0.37	0.12	0.15	0.76	0.98	1.00	1.00	1.00	1.00
-0.21	0.82	0.93	0.94	0.88	0.72	0.39	0.11	0.13	0.69	0.99	1.00	1.00	1.00	1.00
-0.17	0.94	0.97	0.95	0.88	0.70	0.37	0.10	0.09	0.65	0.99	1.00	1.00	1.00	1.00
-0.13	0.98	0.98	0.95	0.87	0.68	0.34	0.08	0.07	0.59	0.99	1.00	1.00	1.00	1.00
-0.10	0.99	0.99	0.95	0.86	0.65	0.32	0.07	0.07	0.59	0.99	1.00	1.00	1.00	1.00
-0.06	1.00	0.99	0.95	0.86	0.63	0.29	0.06	0.06	0.58	0.98	1.00	1.00	1.00	1.00
-0.02	1.00	0.99	0.95	0.87	0.61	0.27	0.06	0.06	0.56	0.97	1.00	1.00	1.00	1.00
0.02	1.00	0.98	0.94	0.85	0.58	0.26	0.06	0.07	0.55	0.95	0.99	1.00	1.00	1.00
0.06	1.00	0.98	0.94	0.83	0.57	0.25	0.06	0.07	0.52	0.92	0.98	1.00	1.00	1.00
0.10	1.00	0.98	0.93	0.81	0.56	0.25	0.06	0.07	0.48	0.88	0.97	0.99	1.00	1.00
0.13	1.00	0.97	0.92	0.78	0.54	0.23	0.07	0.07	0.41	0.82	0.94	0.98	1.00	1.00
0.17	0.99	0.97	0.89	0.74	0.51	0.22	0.07	0.09	0.35	0.74	0.89	0.96	0.99	0.99
0.21	0.98	0.94	0.84	0.70	0.46	0.21	0.08	0.11	0.29	0.58	0.80	0.91	0.97	0.97
0.25	0.94	0.87	0.75	0.60	0.42	0.24	0.13	0.15	0.25	0.43	0.65	0.80	0.90	0.92

CONDITIONAL APPROACH

	ρ_1													
ρ_2	**-0.25**	**-0.21**	**-0.17**	**-0.13**	**-0.10**	**-0.06**	**-0.02**	**0.02**	**0.06**	**0.10**	**0.13**	**0.17**	**0.21**	**0.25**
-0.25	0.71	0.74	0.80	0.80	0.67	0.38	0.13	0.16	0.84	1.00	1.00	1.00	1.00	1.00
-0.21	0.82	0.94	0.94	0.90	0.76	0.45	0.14	0.18	0.88	1.00	1.00	1.00	1.00	1.00
-0.17	0.94	0.98	0.97	0.92	0.80	0.50	0.16	0.19	0.92	1.00	1.00	1.00	1.00	1.00
-0.13	0.98	0.98	0.97	0.94	0.83	0.56	0.19	0.27	0.94	0.99	1.00	1.00	1.00	1.00
-0.10	0.99	0.99	0.98	0.94	0.86	0.62	0.26	0.40	0.95	0.99	1.00	1.00	1.00	1.00
-0.06	1.00	0.99	0.98	0.95	0.89	0.71	0.37	0.54	0.93	0.98	1.00	1.00	1.00	1.00
-0.02	1.00	0.99	0.98	0.96	0.91	0.80	0.57	0.64	0.87	0.97	1.00	1.00	1.00	1.00
0.02	1.00	0.99	0.98	0.96	0.92	0.85	0.63	0.57	0.82	0.96	0.99	1.00	1.00	1.00
0.06	1.00	0.99	0.98	0.96	0.93	0.83	0.61	0.46	0.75	0.93	0.98	1.00	1.00	1.00
0.10	1.00	0.99	0.97	0.96	0.92	0.79	0.47	0.34	0.66	0.88	0.97	0.99	1.00	1.00
0.13	1.00	0.98	0.97	0.96	0.90	0.69	0.36	0.24	0.55	0.83	0.94	0.98	1.00	1.00
0.17	0.99	0.98	0.97	0.95	0.85	0.54	0.21	0.19	0.44	0.75	0.89	0.96	0.99	0.99
0.21	0.98	0.97	0.96	0.93	0.77	0.42	0.17	0.17	0.35	0.60	0.80	0.91	0.97	0.97
0.25	0.96	0.96	0.94	0.85	0.64	0.35	0.17	0.17	0.27	0.45	0.65	0.80	0.90	0.92

CASE 3.B: Sample Size is 100

UNCONDITIONAL APPROACH

	ρ_1													
ρ_2	**-0.23**	**-0.19**	**-0.16**	**-0.12**	**-0.09**	**-0.05**	**-0.02**	**0.02**	**0.05**	**0.09**	**0.12**	**0.16**	**0.19**	**0.23**
-0.23	0.84	0.89	0.95	0.98	0.95	0.79	0.18	0.24	1.00	1.00	1.00	1.00	1.00	1.00
-0.19	0.96	1.00	1.00	1.00	1.00	0.86	0.19	0.23	1.00	1.00	1.00	1.00	1.00	1.00
-0.16	1.00	1.00	1.00	1.00	1.00	0.86	0.18	0.23	1.00	1.00	1.00	1.00	1.00	1.00
-0.12	1.00	1.00	1.00	1.00	1.00	0.85	0.18	0.21	0.99	1.00	1.00	1.00	1.00	1.00
-0.09	1.00	1.00	1.00	1.00	0.99	0.83	0.16	0.17	0.98	1.00	1.00	1.00	1.00	1.00
-0.05	1.00	1.00	1.00	1.00	0.99	0.81	0.15	0.15	0.97	1.00	1.00	1.00	1.00	1.00
-0.02	1.00	1.00	1.00	1.00	0.99	0.78	0.11	0.14	0.96	1.00	1.00	1.00	1.00	1.00
0.02	1.00	1.00	1.00	1.00	0.99	0.75	0.10	0.14	0.95	1.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.00	0.99	0.71	0.11	0.14	0.92	1.00	1.00	1.00	1.00	1.00
0.09	1.00	1.00	1.00	1.00	0.98	0.69	0.12	0.14	0.88	1.00	1.00	1.00	1.00	1.00
0.12	1.00	1.00	1.00	1.00	0.97	0.66	0.12	0.13	0.84	1.00	1.00	1.00	1.00	1.00
0.16	1.00	1.00	1.00	1.00	0.95	0.62	0.12	0.12	0.73	1.00	1.00	1.00	1.00	1.00
0.19	1.00	1.00	1.00	1.00	0.94	0.55	0.11	0.11	0.55	0.95	1.00	1.00	1.00	1.00
0.23	1.00	1.00	1.00	0.99	0.86	0.45	0.11	0.09	0.32	0.68	0.84	0.91	0.98	0.98

CONDITIONAL APPROACH

	ρ_1													
ρ_2	**-0.23**	**-0.19**	**-0.16**	**-0.12**	**-0.09**	**-0.05**	**-0.02**	**0.02**	**0.05**	**0.09**	**0.12**	**0.16**	**0.19**	**0.23**
-0.23	0.84	0.89	0.95	0.98	0.95	0.79	0.18	0.24	1.00	1.00	1.00	1.00	1.00	1.00
-0.19	0.96	1.00	1.00	1.00	1.00	0.86	0.19	0.24	1.00	1.00	1.00	1.00	1.00	1.00
-0.16	1.00	1.00	1.00	1.00	1.00	0.86	0.18	0.24	1.00	1.00	1.00	1.00	1.00	1.00
-0.12	1.00	1.00	1.00	1.00	1.00	0.85	0.19	0.26	1.00	1.00	1.00	1.00	1.00	1.00
-0.09	1.00	1.00	1.00	1.00	0.99	0.85	0.21	0.35	1.00	1.00	1.00	1.00	1.00	1.00
-0.05	1.00	1.00	1.00	1.00	0.99	0.85	0.30	0.57	0.99	1.00	1.00	1.00	1.00	1.00
-0.02	1.00	1.00	1.00	1.00	0.99	0.89	0.47	0.69	0.98	1.00	1.00	1.00	1.00	1.00
0.02	1.00	1.00	1.00	1.00	0.99	0.93	0.66	0.54	0.96	1.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.00	1.00	0.93	0.48	0.32	0.93	1.00	1.00	1.00	1.00	1.00
0.09	1.00	1.00	1.00	1.00	0.99	0.89	0.27	0.20	0.88	1.00	1.00	1.00	1.00	1.00
0.12	1.00	1.00	1.00	1.00	0.99	0.78	0.17	0.14	0.84	1.00	1.00	1.00	1.00	1.00
0.16	1.00	1.00	1.00	1.00	0.97	0.66	0.13	0.12	0.73	1.00	1.00	1.00	1.00	1.00
0.19	1.00	1.00	1.00	1.00	0.94	0.55	0.11	0.11	0.55	0.95	1.00	1.00	1.00	1.00
0.23	1.00	1.00	1.00	0.99	0.86	0.45	0.11	0.09	0.32	0.68	0.84	0.91	0.98	0.98

TABLE 2. LRCOM in the case of a mixed DGP. High Signal-to-Noise Ratio. (continued)

CASE 3.C: Sample Size is 225

UNCONDITIONAL APPROACH

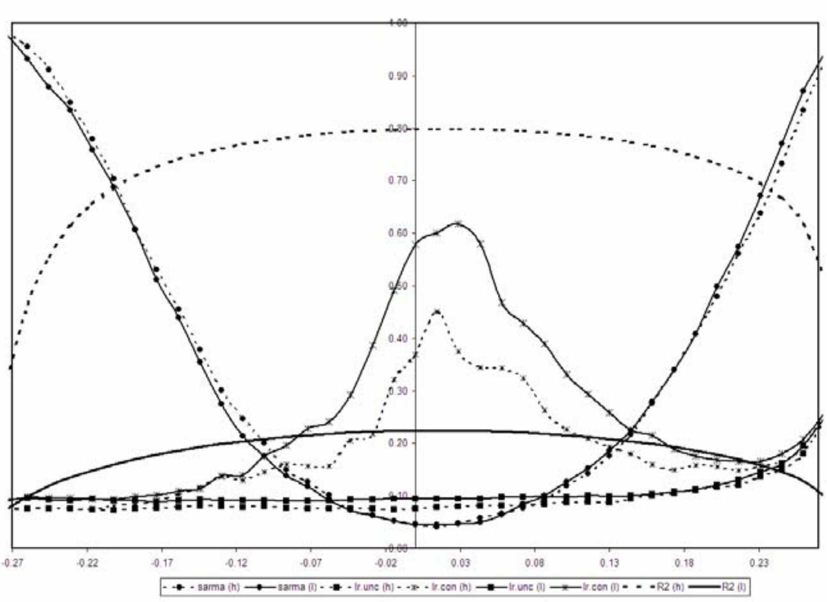
	ρ_1													
ρ_2	**-0.22**	**-0.19**	**-0.15**	**-0.12**	**-0.08**	**-0.05**	**-0.02**	**0.02**	**0.05**	**0.08**	**0.12**	**0.15**	**0.19**	**0.22**
-0.22	0.56	0.41	0.39	0.34	0.30	0.22	0.17	0.21	0.35	0.63	1.00	1.00	1.00	1.00
-0.19	0.37	0.43	0.44	0.43	0.35	0.29	0.26	0.22	0.07	1.00	1.00	1.00	1.00	1.00
-0.15	0.37	0.47	0.51	0.50	0.45	0.33	0.31	0.06	0.49	1.00	1.00	0.99	1.00	1.00
-0.12	0.43	0.52	0.55	0.55	0.54	0.41	0.42	0.02	0.71	1.00	0.97	0.97	1.00	1.00
-0.08	0.41	0.53	0.61	0.60	0.60	0.44	0.41	0.00	0.73	0.97	0.83	0.97	1.00	1.00
-0.05	0.42	0.59	0.65	0.70	0.68	0.57	0.27	0.02	0.71	0.96	0.58	0.99	1.00	1.00
-0.02	0.43	0.60	0.69	0.76	0.73	0.61	0.19	0.13	0.62	0.95	0.33	0.98	1.00	1.00
0.02	0.43	0.64	0.75	0.79	0.74	0.60	0.07	0.20	0.57	0.92	0.21	0.99	1.00	1.00
0.05	0.49	0.67	0.80	0.82	0.76	0.55	0.04	0.26	0.52	0.80	0.13	0.99	1.00	1.00
0.08	0.55	0.73	0.85	0.84	0.74	0.34	0.03	0.35	0.40	0.55	0.11	0.96	1.00	1.00
0.12	0.60	0.81	0.87	0.87	0.66	0.16	0.01	0.28	0.28	0.25	0.09	0.94	1.00	1.00
0.15	0.72	0.88	0.93	0.88	0.53	0.05	0.09	0.25	0.23	0.10	0.10	0.89	1.00	1.00
0.19	0.90	0.98	0.97	0.75	0.14	0.04	0.19	0.21	0.16	0.09	0.14	0.84	1.00	1.00
0.22	1.00	1.00	0.79	0.17	0.17	0.23	0.20	0.11	0.12	0.10	0.21	0.75	0.99	1.00

CONDITIONAL APPROACH

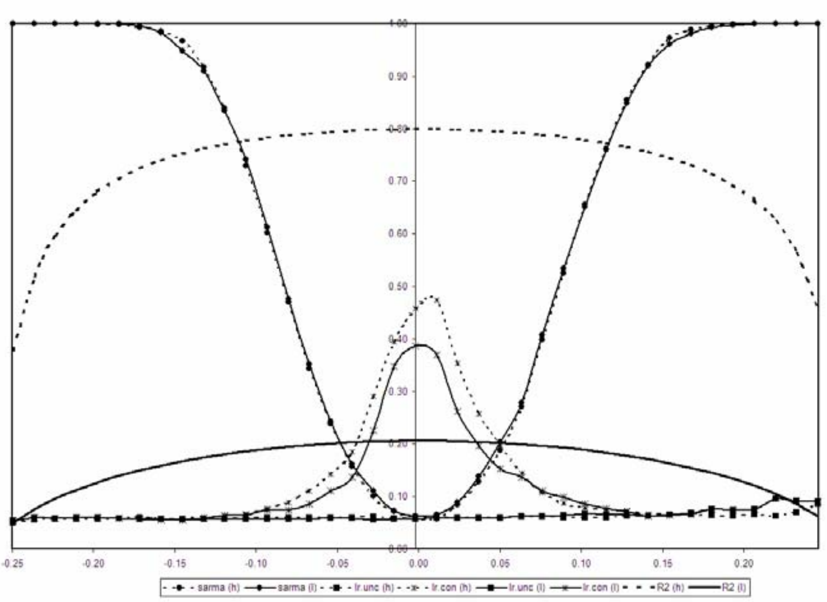
	ρ_1													
ρ_2	**-0.22**	**-0.19**	**-0.15**	**-0.12**	**-0.08**	**-0.05**	**-0.02**	**0.02**	**0.05**	**0.08**	**0.12**	**0.15**	**0.19**	**0.22**
-0.22	0.56	0.41	0.39	0.34	0.30	0.22	0.17	0.21	0.66	1.00	1.00	1.00	1.00	1.00
-0.19	0.37	0.43	0.44	0.43	0.35	0.29	0.26	0.29	0.64	1.00	1.00	1.00	1.00	1.00
-0.15	0.37	0.47	0.51	0.50	0.45	0.33	0.31	0.25	0.77	1.00	1.00	0.99	1.00	1.00
-0.12	0.43	0.52	0.55	0.55	0.54	0.41	0.44	0.40	0.78	1.00	0.97	0.97	1.00	1.00
-0.08	0.41	0.53	0.61	0.60	0.60	0.44	0.51	0.00	0.75	0.97	0.83	0.97	1.00	1.00
-0.05	0.42	0.59	0.65	0.70	0.68	0.57	0.56	0.33	0.72	0.96	0.58	0.99	1.00	1.00
-0.02	0.43	0.60	0.69	0.76	0.73	0.62	0.63	0.59	0.62	0.95	0.33	0.98	1.00	1.00
0.02	0.43	0.64	0.75	0.79	0.74	0.62	0.64	0.51	0.57	0.92	0.21	0.99	1.00	1.00
0.05	0.49	0.67	0.80	0.82	0.76	0.62	0.67	0.52	0.52	0.80	0.13	0.99	1.00	1.00
0.08	0.55	0.73	0.85	0.84	0.75	0.57	0.60	0.47	0.40	0.55	0.11	0.96	1.00	1.00
0.12	0.60	0.81	0.87	0.87	0.69	0.52	0.17	0.33	0.28	0.25	0.09	0.94	1.00	1.00
0.15	0.72	0.88	0.93	0.88	0.70	0.83	0.29	0.27	0.23	0.10	0.10	0.89	1.00	1.00
0.19	0.90	0.98	0.97	0.84	0.74	0.27	0.25	0.21	0.16	0.09	0.14	0.84	1.00	1.00
0.22	1.00	1.00	0.99	0.81	0.44	0.27	0.20	0.11	0.12	0.10	0.21	0.75	0.99	1.00

FIGURE 1: LRCOM in the case of an SEM in the DGP.

CASE 1.A: Sample Size is 25



CASE 1.B: Sample Size is 100



CASE 1.C: Sample Size is 225

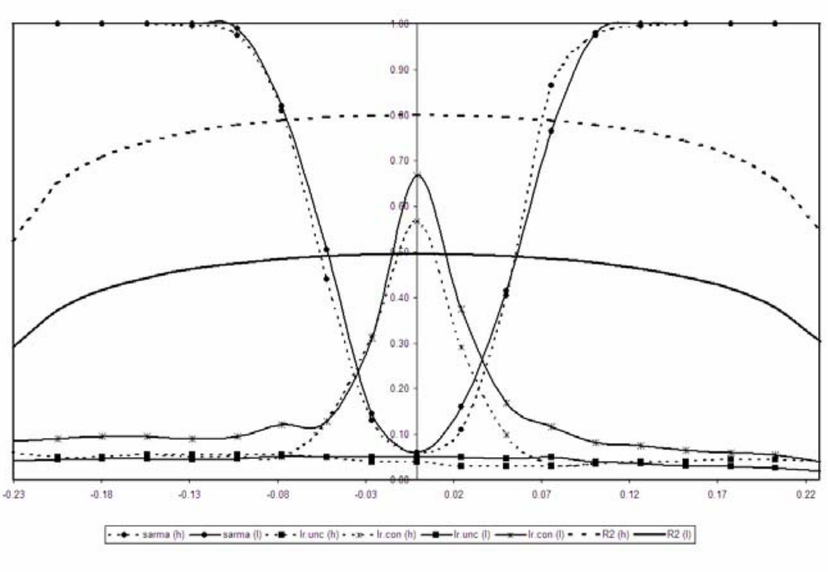
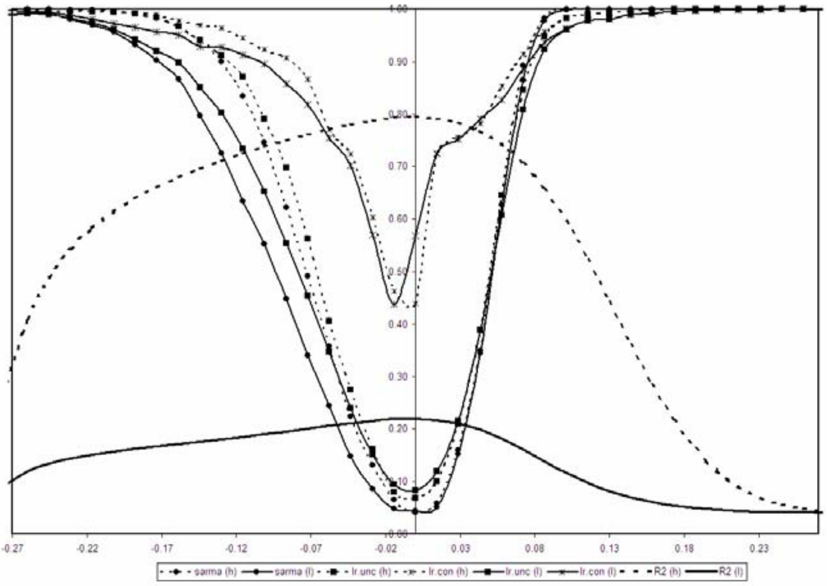
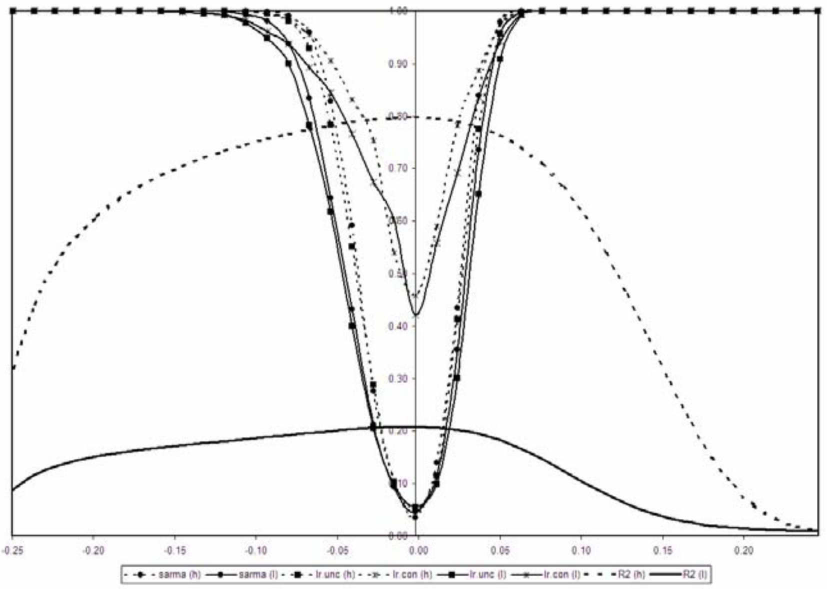


FIGURE 2: LRCOM in the case of an SLM in the DGP.

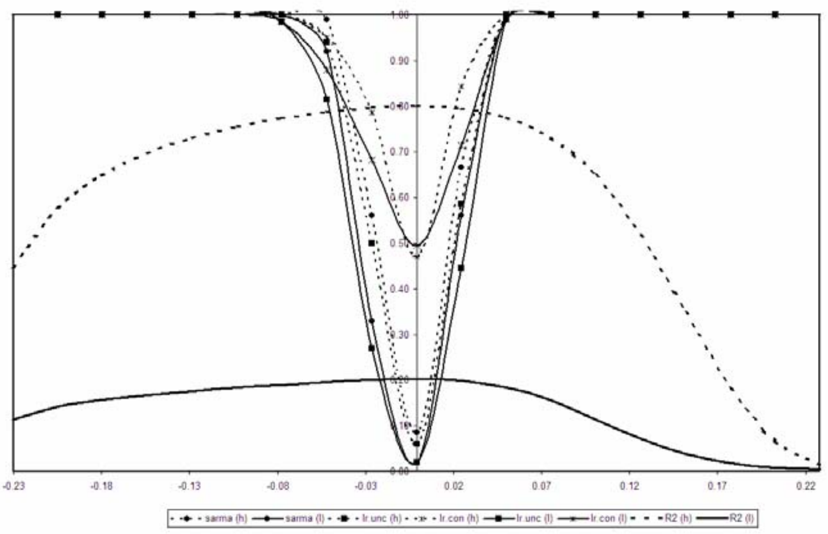
CASE 2.A: Sample Size is 25



CASE 2.B: Sample Size is 100



CASE 2.C: Sample Size is 225



(i) There is a strong effect related to the sample size. The power of the LRCOM test (and also of that of the SARMA) improve substantially with the number of observations.

(ii) There is, equally, an effect associated with the explicative capacity of the model. All the tests work better with a high R^2 coefficient. The differences should not be underestimated because they can reach 30 points in the proximities of the null hypothesis (ρ equal to zero in this case).

(iii) The SARMA and LRCOM tests, in an unconditional approach, evolve in a similar way although the former always moves slightly above the latter. The two tests are not rivals, nor do they refer to the same problem, so this small difference is of no great importance.

(iv) The power functions of both the tests maintain a certain asymmetry with respect to the origin (value zero in ρ). In general, both perform slightly better in the range of positive values of the parameter of autocorrelation. The difference is evident with a sample of 25 observations although it has almost disappeared when the sample grows to 100 observations.

(v) Table 2 reflects a strong preference for processes with a dynamic structure in the main equation. With a grey background we highlight those combinations of parameters in which the LRCOM has selected mostly the SEM model (percentage of rejections under 50%).

(vi) In general, the LRCOM test tends to select the SEM model when the signal coming from the dynamic component of the equation is weak (value of ρ_1 in the proximities of zero). This appears to be more important than the intensity of the signal coming from the equation of the random term (that is, of the value of ρ_2).

(vii) Also in this case there is an effect associated with the sample size. If the number of observations included in the sample is small, the LRCOM test selects, on most occasions, the SLM model. The situation is partially corrected with a sample of 225 observations, although a strong imbalance in the distribution of the decisions taken by the test persists.

4. Conclusions.

The test of Common Factors was introduced into a spatial context at the beginning of the eighties when a lot of the instruments that we use today were still being developed. However, it has never occupied a really important position in the process specifying a model. Habitually, it has been used as an auxiliary test, useful for corroborating conclusions obtained with other instruments. Nevertheless, we believe that it should play a more relevant role as a guide in applied work.

As we have insisted, the LRCOM should not be used as a final test but as an instrument for defining the most adequate direction for the specification process. At least, it should be borne in mind to deal with the requirement of Davidson (2000, p. 168) when, alluding to its time series equivalent, he indicates: '*The point is that although AR(1) errors may well be the correct specification, they impose a common-factor parameter restriction on the equation that requires to be tested. It would nowadays be regarded as bad practice to impose the AR(1) model without testing the implicit restriction*'.

Our position is that, given the peculiarities of the discipline, we can be a little more ambitious. Externalities and dynamic spatial relationships play a strategic role in any model that is specified in this field. Nevertheless, these elements often have an evasive nature that makes them difficult to see. For this reason, it is important to have an instrument that can discriminate between the different alternatives that can be used in order to deal with this type of mechanisms. This is especially so when, as we believe to have demonstrated in the simulation, it seems to be a very reliable test.

Acknowledgements*:* This work has been carried out with the financial support of project SEC 2002-02350 of the Ministerio de Ciencia y Tecnología del Reino de España.

References

Anselin, L., 1980. *Estimation Methods for Spatial Autoregressive Structures*. Regional Science Dissertation and Monograph Series #8. Ithaca: Cornell University.

Anselin L., 2003. Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review,* **26**, 153-166.

Anselin, L. and R. Florax, 1995. Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. In Anselin, L. and R. Florax (eds.) *New Directions in Spatial Econometrics* (pp. 21-74). Berlin: Springer.

Anselin, L., A, Bera, R. Florax and M. Yoon, 1996. Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics,* **26**, 77-104.

Anselin, L. and A, Bera, 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In Ullah, A. and D. Giles (eds.) *Handbook of Applied Economic Statistics* (pp. 237-289). New York: Marcel Dekker.

Bivand, R., 1984. Regression Modeling 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.

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

Cochrane, D. and G. Orcutt, 1949. Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms *Journal of the American Statistical Association,* **44**, 32-61.

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

Dubin, R., 2003. Robustness of Spatial Autocorrelation Specifications: Some Monte Carlo Evidence. *Journal of Regional Science,* **43**, 221-248.

Durbin, J., 1960. Estimation of Parameters in Time-Series Regression Models. *Journal of the Royal Statistical Society B*, **22**, 139-153.

Florax, R. and T. de Graaff, 2004. The Performance of Diagnostics Tests for Spatial Dependence in Linear Regression Models: A Meta-Analysis of Simulation Studies. In Anselin, L., R. Florax and S. Rey (eds.): *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pp. 29-65. Berlin: Springer.

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.

Griffith, D, 2003. *Spatial Autocorrelation and Spatial Filtering*. Berlin: Springer.

López-Bazo, E. and B. Fingleton, 2004. *Empirical Growth Models with Spatial Effects.* Paper presented at the First Seminar Jean Paelinck on Spatial Econometrics, held in Zaragoza (Spain). October 22-23.

McMillen, D., 2003. Spatial Autocorrelation or Model Misspecification. *International Regional Science Review*, **26**, 208-217.

Mur, J. and A. Angulo, 2005. *Model Selection Strategies in a Spatial Setting. Some More Results*. Paper presented at the Workshop on Spatial Econometrics, held in Kiel (Germany). April 8-9.

Mur, J. and J. Trívez, 2003. Unit Roots and Deterministic Trends in Spatial Econometric Models. *International Regional Science Review*, **26**, pp. 289-312.

Tiefelsdorf, M, 2000. *Modelling Spatial Processes. The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I.* Berlin: Springer.