On the specification of spatial econometric models

By Jesús Mur

University of Zaragoza

Abstract

In this paper we propose a reflection on the relevance of the concept of unit roots in the spatial dimension. The specialised literature has not paid excessive attention to this, even though its appearance meant changes in econometric methodology. In the paper we highlight that this is not an intuitive concept and that it does not adapt well to the type of models usually employed in the spatial context. Subsequently we focus our attention on the topic of *deterministic trends* associated with the scale factor that intervenes in autoregressive spatial processes. The incidence of this type of element results in a far from negligible risk of *spurious correlation* that should be taken into account.

Address:

Departamento de Análisis Económico
Universidad de Zaragoza
Gran Vía, 2-4
Zaragoza (50005). SPAIN
e-mail: jmur@posta.unizar.es
Phone: +34-976-761815
Fax: +34-976-761996

42th CONGRESS OF THE EUROPEAN REGIONAL SCIENCE ASSOCIATION

DORTMUND. AUGUST 27th-31st 2002

1.- Introduction

In a recent paper, Fingleton (1999) shows his surprise at the scant impact that the concepts of integration and cointegration have had in the field of spatial econometrics: '... *it is apparent that some new developments that have introduced added rigor into mainstream econometrics, most notably unit roots and cointegration, have not had counterparts in the spatial regression literature'* (p.1). Other colleagues with a traditional econometric background show similar surprise. Such a reaction can be understood because, even if we admit a certain methodological immaturity in spatial econometrics, it is difficult to understand why a concept that has substantially altered the way of carrying out Econometrics has hardly been reflected in the spatial field. From my point of view, I believe there are objective reasons to explain the situation, including the following:

(a)- The concept of integration is developed in a univariate context, but does not adapt well to the multivariate scenario of the spatial dimension.

(b)- The spatial dynamic shows certain peculiarities with respect to the temporal, among which instantaneousness can be highlighted.

(c)- The theoretical relevance of this subject in the spatial dimension, to say the least, still needs to be demonstrated.

It must be remembered in relation to the last point that, although the problem of unit roots is already present in the econometric literature of the seventies (the antecedents date back to Yule, 1926), we have to wait until the work of Nelson and Plosser (1982) for the question to gain prominence. In this work it is shown that the majority of economic series have unit roots and that this fact is relevant for macroeconomic analysis. After them, this type of literature really took off, so it would not be exaggerated to talk about a before and after Nelson and Plosser. However, in the field of *spatial regression literature*, nothing remotely similar has occurred, given that not even the concept of integration seems evident.

The object of this paper is to take up the discussion proposed by Fingleton in the work mentioned above, adopting a slightly more critical position. In section two the fundamental concepts related to the hypotheses of unit roots and cointegration are analysed with a view to their application in a spatial context. The conclusions reached in this section are not excessively favourable to this supposition. In the third section we present the results of a small simulation designed around the concept of spatial deterministic trend. The work finishes with a section of conclusions and final recommendations.

2.- Spatial Unit Roots and Spurious Correlation

The main characteristic of an integrated time series is precisely that it is obtained by *integrating* the contemporary noise with the whole set (possibly infinite) of previous noises. That is:

$$y_t = y_{t-1} + \varepsilon_t \Rightarrow y_t = \sum_{j=0} \varepsilon_{t-j}; \quad \varepsilon_t \sim \text{iidN}(0, \sigma^2) \quad \forall t \qquad (1)$$

which results in erratic series whose variance is not stationary given that it grows with time ($V(y_t)=\sigma^2 t$). The memory of these processes is infinite (that is $Cov(y_t;y_{t-s})=\sigma^2(1-s/t)$) so that any shock will have permanent effects on the series. The consequences are not limited to non-stationarity and non-ergodicity but extend to essential aspects both of the Law of Large Numbers and the Central Limit Theorem, not applicable in these circumstances.

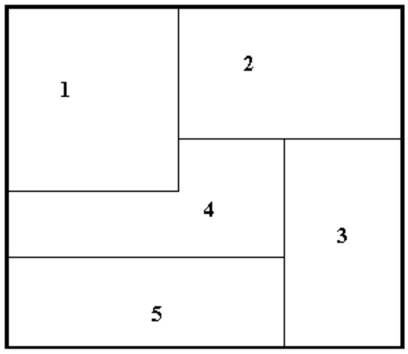
The spatial dimension introduces some difficulties into the previous discussion, as is shown in the seminal work of Fingleton(1999). It is illustrative that the first problem that the author is faced with is that of the specification of a data generating process (DGP) in space which permits the inclusion of a unit root. The solution proposed is well elaborated but not particularly intuitive: '*To avoid circularity in the spatial context, a version of matrix* **W** *is used which has zeros defining an unconnected central cell, just as the time series analogy* $M_{1,T}=0$' (p.5). In the work a regular lattice is used with a variable number of regional units and a contiguity matrix (row standardised) that reflects rook-type movements. The introduction of an unconnected central cell ('... *the unconnected central cell is defined by setting* $W_{ij}=0$ *for* $i=(n+1)/2$, $j=1,...,n$' p.5) breaks with the principle of symmetry in the weight matrix (not essential). Furthermore, and more importantly, it enlarges the range of variation of the autoregressive parameter from the interval $\lambda^{-1}_{M(-)} < \rho < 1$, (relevant under the hypothesis of circularity) to $\lambda^{-1}_{M(-)} < \rho < 1+\eta$ with $\lambda_{M(-)}$ being the highest negative eigenvalue of **W** and η a strictly positive factor (it can be demonstrated that η becomes smaller as the size of the lattice system increases). The final consequence is that the value 1 is no longer a singularity point of the matrix **W** and can, therefore, be used in the simulation of the spatial autoregressive process (SAR):

$$y = \rho \mathbf{W} y + \varepsilon \Rightarrow y = [I - \rho \mathbf{W}]^{-1} \varepsilon = \mathbf{B}^{-1} \varepsilon; \quad \varepsilon \sim N(0, \sigma^2 \mathbf{I}) \qquad (2)$$

Making ρ equal to 1 in the previous expression, we obtain what could be interpreted as a spatial random walk (SRW), equivalent to that specified in (1) for the time dimension.

However, in this case I think that between expressions (1) and (2) the differences predominate over the similarities. To illustrate this question we are going to use the system shown in Figure 1.

Figure 1: A basic regional system.



The contiguity matrix obtained from Fingleton's proposal, using region 3 as unconnected, is:

$$\mathbf{W} = \begin{bmatrix} 0 & 0.50 & 0 & 0.50 & 0 \\ 0.33 & 0 & 0.33 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0 & 0.25 \\ 0 & 0 & 0.50 & 0.50 & 0 \end{bmatrix}$$

The highest eigenvalues are 0.76 (positive) and –0.50 (negative), so that the admissible range of variation of the parameter ρ in equation (2) becomes (-2.0000; 1.3203). If we make ρ=1, we obtain:

$$y = \rho \mathbf{W} y + \varepsilon = y_L + \varepsilon \qquad (3)$$

with y_L= $\mathbf{W}y$ being the spatial lag of the series y. The reduced form of the latter system of equations can be expressed as:

$$\left.\begin{array}{l} y_1 = 1.81\varepsilon_1 + 1.29\varepsilon_2 + \varepsilon_3 + 1.52\varepsilon_4 + 0.38\varepsilon_5 \\ y_2 = 0.86\varepsilon_1 + 1.71\varepsilon_2 + \varepsilon_3 + 1.14\varepsilon_4 + 0.29\varepsilon_5 \\ y_3 = \varepsilon_3 \\ y_4 = 0.76\varepsilon_1 + 0.86\varepsilon_2 + \varepsilon_3 + 1.90\varepsilon_4 + 0.47\varepsilon_5 \\ y_5 = 0.38\varepsilon_1 + 0.43\varepsilon_2 + \varepsilon_3 + 0.95\varepsilon_4 + 1.24\varepsilon_5 \end{array}\right\}$$

The next question is to verify that the series is centred on zero (Ey=0) and is heteroskedastic (making σ^2=1 we obtain $V(y_1)$=8.39, $V(y_2)$=6.06, $V(y_3)$=1, $V(y_4)$=6.17, $V(y_5)$=3.77). Fingleton (1999) also advises that on increasing the number of regions all the indicators associated with the variance grow as well. That is to say, equation 3 replicates (at least in its formal aspect) the typical structure of a random walk and produces series with symptoms of integration.

However, it is inevitable to point out that the reduced form obtained for the SRW of (3) has little to do with the mechanism of accumulation of shocks characteristic of the temporal case (developed within the triangular structure of the system of equations of the reduced form). The variance is heteroskedastic, but its relationship with the sample size, or with the position of each region in the system, is not at all evident. It is true that, as we add new regions, there is a progressive increase in the variances associated with the variables of the SRW. Nevertheless, the growth shown by the variance indicators is vastly superior to that of the sample size (the proportion in Fingleton's work is approximately four to one).

These observations raise doubts about the interpretation of the DGP specified in (3), above all because what have been understood as symptoms of integration also admit other types of interpretations. For example, SAR processes are heteroskedastic by definition, whether or not there is a unit root in the equation. Furthermore, the explosive behaviour shown by the variance indicators can be explained by recourse to the diagonalisation of the contiguity matrix. Effectively, this matrix (square though, as it has been specified, asymmetric) can be decomposed into:

$$\mathbf{W} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \Rightarrow \mathbf{B} = \mathbf{I} - \rho\mathbf{W} = \mathbf{Q}(\mathbf{I} - \rho\mathbf{\Lambda})\mathbf{Q}^{-1} = \mathbf{Q}\mathbf{\Delta}\mathbf{Q}^{-1} \qquad (4)$$

The covariance matrix of vector y of expression (2) can be expressed as:

$$V(y) = \sigma^2(\mathbf{B'B})^{-1} = \sigma^2\mathbf{Q}\mathbf{\Delta}^{-1}\mathbf{Q}^{-1}(\mathbf{Q'})^{-1}\mathbf{\Delta}^{-1}\mathbf{Q'} = \sigma^2\mathbf{Q}\mathbf{\Delta}^{-1}(\mathbf{Q'Q})^{-1}\mathbf{\Delta}^{-1}\mathbf{Q'} \qquad (5)$$

The variance of observation y_r is a quadratic form on the matrix $(\mathbf{Q'Q})^{-1}$:

$$V(y_r) = \sigma^2 q_r^*(\mathbf{Q'Q})^{-1} q_r^{*'} \qquad (6)$$

with $q_r^* = \begin{bmatrix} \frac{q_{1r}}{1-\rho\lambda_1} & \frac{q_{2r}}{1-\rho\lambda_2} & \cdots & \frac{q_{Rr}}{1-\rho\lambda_R} \end{bmatrix}$ being the r^{th} row of matrix $\mathbf{Q}\Delta^{-1}$, where q_{rs} is the element of row r and column s of matrix $\mathbf{Q}$. Developing this quadratic form we obtain:

$$V(y_r) = \sigma^2 \sum_{m;n} q_{mr}^* q_{nr}^* m_{mn} = \sigma^2 \sum_{m;n} \frac{q_{mr} q_{nr}}{(1-\rho\lambda_m)(1-\rho\lambda_n)} m_{mn} \qquad (7)$$

with m_{mn} the element of row m and column n of matrix $(\mathbf{Q'Q})^{-1}$. If we maintain $\rho=1$ at the same time as we increase the sample size, one of the roots will approximate to 1 from the right (for example, the p^{th}: $\lambda_p \to 1$), and the final result will be a group of variances with a tendency to become explosive.

A simpler solution consists of opting not to standardise the contiguity matrix, following the observation of Kelejian and Robinson (1992) when they criticise the interpretation of ρ in (2) as an autocorrelation coefficient. In accordance with their reasoning, there is no need to impose restrictions on this parameter, except those related to the exclusion of the R singularity points associated with the contiguity matrix. The parametric space of the coefficient becomes the real line, excluding the R singularity points. Specifically, the probability that one of these discontinuities is located on the point ρ=1 is remote, so it is not necessary to alter the usual specification.

This proposal solves the formal aspect of the problem, though it does not dissipate the doubts with respect to the interpretation of the SRW of (3) in the spatial dimension. For example, going back to system of Figure 1 and using a first order binary contiguity matrix, the reduced form associated with the SRW is:

$$\left.\begin{array}{l} y_1 = 1.14\varepsilon_1 + 0.29\varepsilon_2 - 0.71\varepsilon_3 - 0.14\varepsilon_4 - 0.85\varepsilon_5 \\ y_2 = 0.29\varepsilon_1 + 0.57\varepsilon_2 - 0.42\varepsilon_3 - 0.29\varepsilon_4 - 0.71\varepsilon_5 \\ y_3 = -0.71\varepsilon_1 - 0.42\varepsilon_2 + 0.57\varepsilon_3 - 0.29\varepsilon_4 + 0.29\varepsilon_5 \\ y_4 = -0.14\varepsilon_1 - 0.29\varepsilon_2 - 0.29\varepsilon_3 + 0.14\varepsilon_4 - 0.14\varepsilon_5 \\ y_5 = -0.85\varepsilon_1 - 0.71\varepsilon_2 + 0.29\varepsilon_3 - 0.14\varepsilon_4 + 1.14\varepsilon_5 \end{array}\right\}$$

Examining the structure of this system of equations, it becomes difficult to even think about seeking something comparable to a spatial stochastic trend. Another aspect to be considered is the evolution of the variance when there are changes in the sample size or in parameter ρ. In this case the results are simpler because matrix **W** maintains its symmetry and its eigenvectors are orthogonal. The covariance matrix of (5) simplifies as:

$$V(y) = \sigma^2 \mathbf{Q} \boldsymbol{\Delta}^{-2} \mathbf{Q}' \qquad (8)$$

and the variance of observation r reduces to:

$$V(y_r) = \sigma^2 \sum_{m=1}^{R} \frac{q_{mr}^2}{(1-\rho\lambda_m)^2} \qquad (9)$$

It is not easy to determine the behaviour of this ratio, though some evidence can be used. In the first place, the sum of the terms of the numerator is equal to 1 because the eigenvectors are orthogonal ($\sum_{m=1}^{R} q_{mr}^2 = 1$). Furthermore, on excluding the singularity points of **W**, all the denominator terms are finite at the same time as their sum is limited (for ρ=1 it can be verified that $\sum_{m=1}^{R}(1-\rho\lambda_m)^2 = \sum_{m=1}^{R}\left(1+\rho^2\lambda_m^2 - 2\rho\lambda_m\right) = R+\rho^2 S_0 \cong 5R$ and $S_0 = \sum_{m,n} w_{mn} = \sum_m \lambda_m^2 \cong 4R$). The consequence is that the variance of (9) is not a function of R. In reality, the determining element of this variance is the proximity of ρ to one of the singularity points of **W**.

To sum up, my position on the real incidence of the problem of stochastic trends in a context of spatial series is one sceptic. It is true that there are symptoms that could be related to the phenomenon of unit roots, though not exclusively with it. In fact, there are other types of arguments that offer an equally satisfactory explanation.

In the analysis of time series it is convenient to pay attention also to the problem of *deterministic trends* because they produce similar results. Frequently, the profiles of an integrated series with drift and of a stationary series around a linear trend tend to get confused, although the statistical implications of one or the other type of process are very different. This requires the discussion about the stochastic structure that underlies the series to be flexible enough to fit different specifications of the DGP (Maddala and Kim, 1998).

The situation in a context of spatial series is peculiar because, as has been discussed, the root unit hypothesis may be questioned and the concept of deterministic trend has never achieved excessive popularity. One of the best-known models in this area is that of *trend surfaces* (Ripley, 1981) but, in any case, it must be recognised that the proposals are scarce, perhaps because their usefulness as a modelling tool has been quite limited.

Throughout this work we are going to use a different sense of the concept of spatial trend. To be exact, we are going to link this term to the non-stochastic component associated with the scale of a spatial autoregressive process. If we introduce a common scale factor for all the regions into the SAR of expression (2):

$$y = \boldsymbol{\delta} + \rho \mathbf{W} y + \varepsilon \Rightarrow y = [\mathbf{I} - \rho \mathbf{W}]^{-1}(\boldsymbol{\delta} + \varepsilon) = \boldsymbol{\delta}^* + \mathbf{B}^{-1}\varepsilon = \delta l^* + \mathbf{B}^{-1}\varepsilon \qquad (10)$$

with δ'=[δ, δ, ..., δ] of the order (Rx1), $\delta^*=\mathbf{B}^{-1}\delta$ y $\mathbf{l}^*=\mathbf{B}^{-1}l$ and $\mathbf{l}^*=\mathbf{B}^{-1}l$, and l is a unit vector also of order (Rx1). The component δ (or δ^* in the reduced form) is non-stochastic and plays a role similar to a trend on the time axis**.**

It must be remembered that, although the scale factor included in the structural form is the same for all regions (δ), the expected value of the series changes for each observation. Series (10) will not be mean stationary ($Ey=\delta\mathbf{l}^*$) unless the scale and/or the autoregressive parameter are zero (another possibility is that the series is a spatial moving average in which case it will always be mean stationary). This trending factor, $\delta\mathbf{l}^*$, is the fundamental path of the series through space around which an autoregressive component centred on zero ($E[\mathbf{B}^{-1}\varepsilon]=0$) is superimposed. The sum of both produces the series that is finally observed, *y*. That is to say, the first component seems to bring together the typical features of a trend.

Apart from the above, the aspect that has most interested me with respect to this factor is that its presence, in SAR type series, leads naturally to the concept of *spurious regression*. Granger and Newbold (1974) begin their famous article saying that '*It is very common to see reported in applied econometric literature time series regression equations with an apparently high degree of fit, as measured by the coefficient of multiple correlation* R^2 *or the corrected coefficient* $\bar{R}^2$, *but an extremely low value for the Durbin-Watson statistic. We find it very curious that whereas virtually every textbook on econometric methodology contains explicit warnings of the dangers of autocorrelated errors, this phenomenon crops up so frequently in well-respected applied work. (...). However, we will suggest that cases with much less extreme values may well be entirely spurious.*'(p.111).

The panorama that we find in a spatial context presents many similarities to that described by Granger and Newbold. Here also it is frequent to find applications with a good fit and clear symptoms of autocorrelation. The usual practice consists of reusing these indications to purge the starting relationship, modifying the dynamic structure of the error term or that of the equation itself. It is convenient to highlight that a model with evident symptoms of misspecification is rarely simply discarded, but rather, first, every attempt is made to correct these symptoms. This stance of absolute confidence in the theoretical specification is not good and can lead to statistically irrelevant or meaningless regressions in which the mechanism of common trends becomes relevant.

Let us suppose, for simplicity's sake that we are using a linear relationship between two variables, y_1 and y_2 such as the following:

$$y_{1r} = \alpha + \beta y_{2r} + u_r \tag{11}$$

These variables have been generated by unconnected SAR processes:

$$\left.\begin{aligned} y_1 = \delta_1 + \rho_1 \mathbf{W}_1 y_1 + \varepsilon_1 \Rightarrow y_1 = \delta_1 \mathbf{B}_1^{-1} l + \mathbf{B}_1^{-1}\varepsilon_1;\quad \mathbf{B}_1^{-1} = (\mathbf{I} - \rho_1 \mathbf{W}_1)^{-1} \\ y_2 = \delta_2 + \rho_2 \mathbf{W}_2 y_2 + \varepsilon_2 \Rightarrow y_2 = \delta_2 \mathbf{B}_2^{-1} l + \mathbf{B}_2^{-1}\varepsilon_2;\quad \mathbf{B}_2^{-1} = (\mathbf{I} - \rho_2 \mathbf{W}_2)^{-1} \end{aligned}\right\} \tag{12}$$

The terms ε_1 and ε_2 are white noises ($\varepsilon_j \sim N\left[0, \sigma_j^2 \mathbf{I}\right]$) mutually independent. Given that the variables y_1 and y_2 are not related, the estimation of an equation like that of (11) should corroborate the hypothesis of a lack of relationship, accepting that β=0. However, the result will habitually be the opposite when we introduce certain combinations of parameters. To prove this we only need to observe that the correlation coefficient between the two variables:

$$r_{y_1 y_2} = \frac{\sum_r \left(y_{1r} - \bar{y}_1\right)\left(y_{2r} - \bar{y}_2\right)}{\sqrt{\sum_r \left(y_{1r} - \bar{y}_1\right)^2 \sum_r \left(y_{2r} - \bar{y}_2\right)^2}} \tag{13}$$

tends to unity when both share the same spatial deterministic trend. Introducing the following matrix notation:

$$\sum_r \left(y_{1r} - \bar{y}_1\right)\left(y_{2r} - \bar{y}_2\right) = \left[y_1 - \bar{\boldsymbol{y}}_1\right]'\left[y_2 - \bar{\boldsymbol{y}}_2\right]$$

$$\sum_r \left(y_{jr} - \bar{y}_j\right)^2 = \left[y_j - \bar{\boldsymbol{y}}_j\right]'\left[y_j - \bar{\boldsymbol{y}}_j\right];\ j = 1,2$$

the probabilistic limit of (13) can be expressed as:

$$\text{plim}\, r_{y_1 y_2} = \frac{\text{plim}(\left[y_1 - \bar{\boldsymbol{y}}_1\right]'\left[y_2 - \bar{\boldsymbol{y}}_2\right])/\text{R}}{\sqrt{\text{plim}(\left[y_1 - \bar{\boldsymbol{y}}_1\right]'\left[y_1 - \bar{\boldsymbol{y}}_1\right])/\text{R}}\sqrt{\text{plim}(\left[y_2 - \bar{\boldsymbol{y}}_2\right]'\left[y_2 - \bar{\boldsymbol{y}}_2\right])/\text{R}}} \tag{14}$$

Moreover, it could be shown that:

$$\text{plim}(\left[y_1 - \bar{\boldsymbol{y}}_1\right]'\left[y_2 - \bar{\boldsymbol{y}}_2\right])/\text{R} = \delta_1 \delta_2 \text{Cov}(b_1, b_2) \tag{15a}$$

$$\text{plim}(\left[y_j - \bar{\boldsymbol{y}}_j\right]'\left[y_j - \bar{\boldsymbol{y}}_j\right])/\text{R} = \delta_j^2 \text{V}(b_j) + \sigma_j^2 c_j \tag{15b}$$

where $b_j = \mathbf{B}_j^{-1} l$; $c_j = \frac{1}{R}\left[\sum_r \left(1 - \rho_j \lambda_r\right)^{-2} - \sum_r q_r^2 \left(1 - \rho_j \lambda_r\right)^{-2}\right]$ and q_r is the element r of the vector $q = \mathbf{Q}' l$, with $\mathbf{Q}$ the matrix of eigenvectors of $\mathbf{W}$. Introducing (15a) and (15b) into (14):

$$\text{plim}\, r_{y_1 y_2} = \frac{\delta_1 \delta_2 \text{Cov}(b_1, b_2)}{\sqrt{\delta_1^2 \text{V}(b_1) + \sigma_1^2 c_1}\sqrt{\delta_2^2 \text{V}(b_2) + \sigma_2^2 c_2}} \tag{16}$$

The convergence limit of both variables, will be different to zero unless Cov(b_1,b_2)=0 (or δ_j=0, j=1,2). It will, in fact take values close to unity when the characteristics of the two SAR processes are similar. For example, if we make $\delta_j=\delta$ and $\rho_j=\rho$, then $b_j=b$, V(b_j)=V(b), $\sigma_j^2=\sigma^2$ and Cov(b_1,b_2)=V(b) so that the probabilistic limit of the coefficient is reduced to:

$$\text{plim}\, r_{y_1 y_2} = \frac{1}{\sqrt{1 + (\sigma/\delta)^2 (c / V(b))}} \tag{17}$$

which will tend to unity as the scale factor (δ) increases with respect to the noise variance (σ^2).

3.- Some Monte Carlo evidence

The graphs included in Figure 1 can be useful to evaluate the real incidence of the problem posed in the previous section. These graphs come from a small simulation exercise on whose results I will now comment. During the experiment pairs of series with an SAR structure have been generated, replicating the equations of (12). The vectors ε_j, mutually independent, have been obtained from a gaussian distribution centred on zero and with unit variance. During the exercise, different sample sizes and scale factors, as well as distinct degrees of spatial dependence have been simulated. Then, each pair of series has been related through a linear model which has been estimated by LS. In the graphs of Figure 1 the average R^2, obtained after 1000 iterations, for each pair of parameters is represented. The horizontal axis of each graph corresponds to the value of the autoregressive coefficient used in the SAR process of the right-hand side variable of the regression, while that corresponding to the left-hand side is represented on the vertical axis.

The results shown in Figure 1 indicate that passing from a scale factor of zero in both variables to a scale of one hundred has a strong impact on the coefficient of determination. In the first case, the R^2 rarely surpasses 0.05 while in the second values superior to 0.80 are the most habitual. That is to say, the risk of spurious correlation is minimum when no scale factor intervenes in the DGP of the variables. However, this risk increases dramatically when high scale series are combined.

The surprising aspect of the simulation is that a perverse behaviour is observed in the sample size. Indeed, for a scale of one hundred, and as the sample size increases, the shape of the graphs tends to get concentrated on the cross-diagonal around which the pairs of parameters of the same sign and close in value are situated. That is, it is the combinations of parameters that lead to the result of (17). As was pointed out above, the average correlation coefficient that is observed for this type of series tends to unity as the scale increases. However, the graphs also indicate that the risk of spurious correlation is higher for small sized samples where regressions with an R^2 coefficient superior to 0.80 predominate. This result is indeed peculiar because, in a context of time series, the risk of spurious correlation tends to grow with the sample size.

If the risk of falling into artificial regressions exists, as is shown in Figure 1, the important question is how to detect this situation. Granger and Newbold (1974), in a time context, offer the clue of the Durbin-Watson statistic, indicating that a high value of this is synonymous with misspecification. Nevertheless, this advice is not very useful in a spatial context where the habitual practice is to adjust the specification to accommodate these symptoms of misspecification.

FIGURE 1

In any case, and following Granger and Newbold, in the simulation we have obtained some specification tests to examine whether their information can be made use of in any way. Specifically, we have analysed the following tests:

Moran Test: $$\mathrm{I} = \frac{\mathrm{R}}{\mathrm{S}_0}\frac{\hat{u}'\mathbf{W}\hat{u}}{\hat{u}'\hat{u}};\ \mathrm{S}_0 = l'\mathbf{W}l \tag{18}$$

LM-LAG Test: $$\mathrm{LM-LAG} = \frac{1}{\mathrm{R}\,\hat{\mathrm{J}}_{\rho-\beta}}\left(\frac{\hat{u}'\mathbf{W}y}{\hat{\sigma}^2}\right)^2 \tag{19}$$

LM-EL Test: $$\mathrm{LM-EL} = \frac{\left(\frac{\hat{u}'\mathbf{W}\hat{u}}{\hat{\sigma}^2} - \frac{\mathrm{T}_1}{\mathrm{R}\,\hat{\mathrm{J}}_{\rho\text{-}\beta}}\frac{\hat{u}'\mathbf{W}y}{\hat{\sigma}^2}\right)^2}{\mathrm{T}_1 - \frac{\mathrm{T}_1^2}{\mathrm{R}\,\hat{\mathrm{J}}_{\rho\text{-}\beta}}};\ T_1 = 2\,S_0 \tag{20}$$

LM-LE Test: $$\mathrm{LM-LE} = \frac{\left(\frac{\hat{u}'\mathbf{W}y}{\hat{\sigma}^2} - \frac{\hat{u}'\mathbf{W}\hat{u}}{\hat{\sigma}^2}\right)^2}{\mathrm{R}\,\hat{\mathrm{J}}_{\rho\text{-}\beta}\text{-}\mathrm{T}_1} \tag{21}$$

SARMA Test: $$SARMA = \frac{\left(\frac{\hat{u}'\mathbf{W}y}{\hat{\sigma}^2} - \frac{\hat{u}'\mathbf{W}\hat{u}}{\hat{\sigma}^2}\right)^2}{\mathrm{R}\,\hat{\mathrm{J}}_{\rho\text{-}\beta}\text{-}\mathrm{T}_1} + \frac{1}{T_1}\left(\frac{\hat{u}'\mathbf{W}\hat{u}}{\hat{\sigma}^2}\right)^2 \tag{22}$$

with $\hat{\sigma}^2$ and $\hat{\beta}$ being the LS estimation (or ML) estimators and $\hat{u}$ the corresponding residual series, $\mathrm{R}\,\hat{\mathrm{J}}_{\rho-\beta} = \mathrm{T}_1 + (\hat{\beta}'\mathbf{X}'\mathbf{WMWX}\hat{\beta})/\hat{\sigma}^2$ and **M** the matrix (**I**-**X**(**X'X**)$^{-1}$**X'**). As is well-known, the asymptotic distribution of the standardised Moran's I is an N(0,1); the three Lagrange Multipliers that follow (LM-LAG, LM-EL and LM-LE) have an asymptotic $\chi^2(1)$ distribution and the final SARMA test, another Lagrange Multiplier, is a $\chi^2(2)$. The characteristics of the five tests are relatively well documented (see, for example, Anselin and Florax, 1995), being of common use in the applied literature.

Because of the characteristics of the problem addressed, it seems that the solution should come from the robust tests. Model (11) is wrongly specified because a spatial lag of the endogenous variable (y_1) has been omitted. Variable y_2, included in the right-hand side of the equation, is a *proxy* for this lag, which results in a high R^2 when the DGP of the two variables includes a scale factor. That is, this is not about problems with the error term (ε_t is a white noise) but with the dynamic structure of the equation. These symptoms of misspecification should be captured by the LM-LE test but should not have any impact on the LM-EL. The other tests included in the experiment will also reflect these symptoms, possibly not in a coherent form, so the checking of the equation can become confused and not very effective.

The most relevant results about the behaviour of the five tests obtained in carrying out the simulation are summarised in Figures 2 to 6, where the same structure is used as in Figure 1. In this case, instead of the R^2 of the regression, the percentages of rejections of the null hypothesis by the test to which it is dedicated are shown in the corresponding Figure.

FIGURES 2 - 6

Underlying these graphs, there are, at least, two types of tendencies. One is related to the number of observations and the other to the scale factor. With respect to the first, the performance of the tests to changes in the sample size is well known, and the results reflected tend to corroborate the expectations. However, the collection of Figures also clearly shows that the impact of the scale factor is not the same for the five tests.

For example, Moran's test behaves strangely when high scale factors are used in both variables. The power maps estimated for the case of (δ=0) conform to what was expected: it is not a very trustworthy test in reduced sample sizes (R=17) though its ability to detect problems in the specification, associated both with the error term and with the dynamic structure of the model, is high for large samples. Nevertheless, these graphs tend to get distorted when a scale factor is introduced into the DGP of the variables. The improvements in the power of the test, due to the increase in the number of observations, are no longer regularly shared out through the whole parametric space, but rather, we find the appearance of *worms*, regions in which the test loses power unexpectedly. This result seems to be to do with the strong *non-stationarity in the mean* that is characteristic of SAR series with a high scale factor.

In contrast, the LM-LAG and SARMA tests (Figures 3 and 6) are affected positively by the inclusion of a scale factor in the DGP of the variables, in such a way that their estimated power improves with the size of the scale. The SARMA test has more difficulties to detect errors in the specification when the scale factor is zero and the sample size employed is of reduced dimensions. However, the estimated power of this test improves on increasing either of the two aspects (size or scale).

Finally, Figures 4 and 5 show a quite disappointing behaviour of the robust tests. As has been said above, the LM-EL test should not detect signs of misspecification in the error term of the model, as, in fact, occurs when the scale factor is zero. In this case, although the estimated power tends to be superior to the standard of 5% (it oscillates between 3% and 10%), it does not reach worrying levels except for very extreme combinations of parameters. The good behaviour of this test, in series with a reduced scale factor, suffers from the weakness of its complement, the LM-LE test. In Figure 6 it can be seen that the estimated power of this latter test hardly surpasses 20% in large zones of the parametric space, even having employed large sized samples.

The scale helps to improve the estimated power of the LM-LE test, in such a way that with a factor of δ=100 its performance is fully satisfactory, as can be seen in Figure 5 (the estimated power in the darkest zone is, in fact, unity). On the other hand, the LM-EL test behaves in the same way when it should not, under any circumstances, do so. In Figure 4, and under the heading δ=100, irregular estimated power maps can be seen where the darkest zones tend to spread as the sample size increases (as before, the estimated power at these points is unity for the case of R=120).

The conclusion that can be drawn from all of the above is that, using these tests, there is a probability that the analyst will not conclude that he is proposing a spurious relationship, and that this probability is no small matter. The tests will give signs of misspecification with respect to the static equation, ever sharper as the scale factor increases. However, the relevant problem is to identify clearly the real cause of misspecification. If we conclude that it is necessary to introduce a dynamic structure into the equation, the risk of spurious correlation disappears because the spatial lag of the endogenous variable will expel the erroneous variable from the equation. On the other hand, if the analyst feels satisfied with the fit of the equation, and prefers to correct the symptoms of misspecification by acting on the structure of the error, the risk will become reality.

4.- Final conclusions

In this work we wanted to take up a discussion that has not received excessive attention in spatial econometric literature, namely that of unit roots. This concept does not adjust well to the type of models that are habitually employed in a spatial context, characterised by instantaneousness and multidimensionality, which to a certain extent justify why it has been dealt with so little.

The position maintained in this paper is one of scepticism with respect to the problem of *stochastic trends*. For this reason, attention has been centred on the topic of *deterministic trends*, associated with the scale factor that intervenes in SAR processes. The incidence of this type of elements results in a serious risk of *spurious correlation* which every analyst should keep in mind. In this sense we could make the following observations:

- The exploratory analysis of spatial data has been popularised among the practitioners of spatial econometrics as a way of introducing the problem. Nevertheless, this analysis should grow in importance to reach a stage where the coherence of the whole econometric specification is discussed.
- There is an evident risk of posing spurious regressions between non-related variables. This risk increases with the size of the variables.
- It is dangerous to develop a specification based exclusively on obtaining a good fit between variables, because the probability of obtaining meaningless equations is multiplied.
- The tests of spatial dynamics (residual or substantive) commonly used will show symptoms of misspecification if the equation is spurious. The problem is that the information will tend to be confusing.
- We should question the real usefulness of models with residual autocorrelation because they can mask spurious relationships between non-related variables.

References

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

Fingleton, B. (1999): Spurious Spatial Regression: Some Monte Carlo Results with a Spatial Unit Root and Spatial Cointegration. *Journal of Regional Science*, 39, 1-19.

Granger, C.W.J. and P. Newbold. (1974): Spurious Regressions in Econometrics. *Journal of Econometrics*. 2, 111-120.

Kelejian, H. and D. Robinson (1995): Spatial Correlation: A Suggested Alternative to Autoregressive Models. In L. Anselin and R. Florax (eds): *New Directions in Spatial Econometrics* (pp 75-95). Berlin: Springer.

Maddala, G.S. and I.M. Kim (1998): *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.

Nelson, C.R. and C.I. Plosser. (1982): Trends and Random Walks in Macroeconomic Time Series. Some Evidence and Implications. *Journal of Monetary Economics*. 10, 139-162

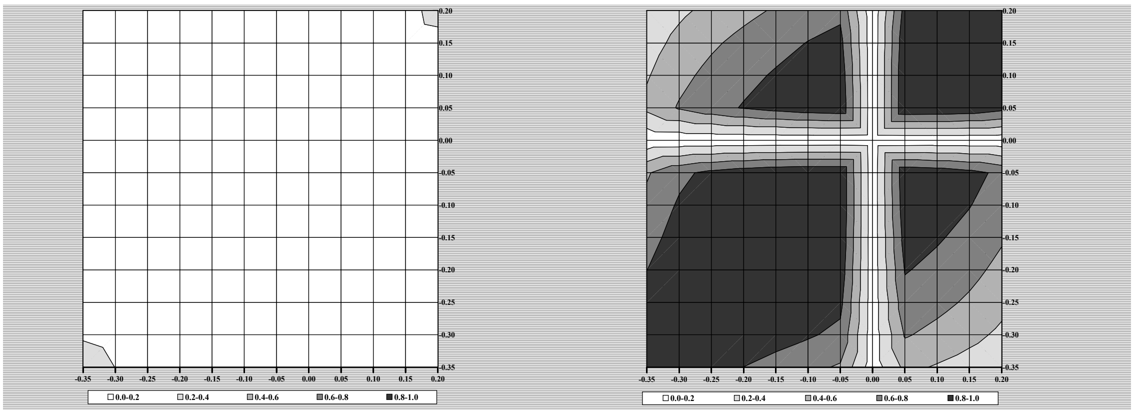
Ripley, B.D. (1981): *Spatial Statistics*. New York: Wiley.

Yule, G.U. (1926): Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series. *Journal of the Royal Statistical Society*, 89, 1-64.

FIGURE 1:. Average R^2 for pairs of independent SAR series.

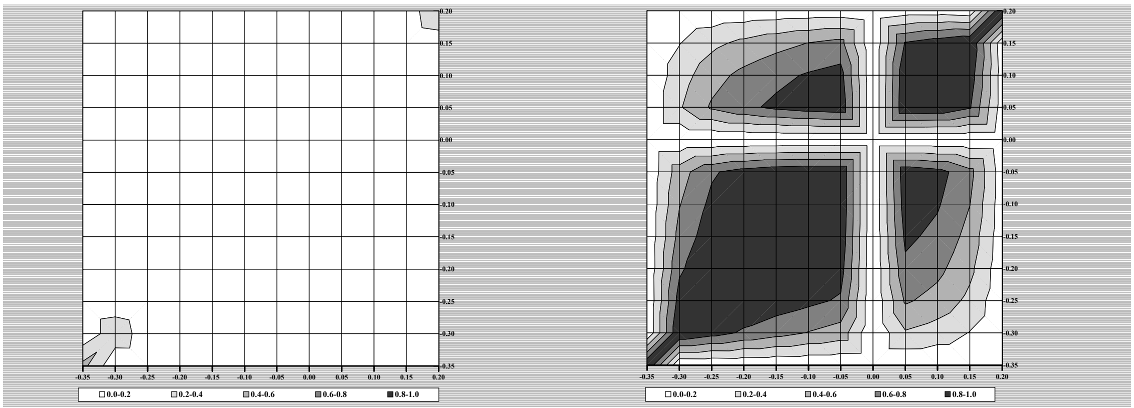
Sampling size: R=17.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=74.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=120.

Scale: δ =0 **Scale: δ =100**

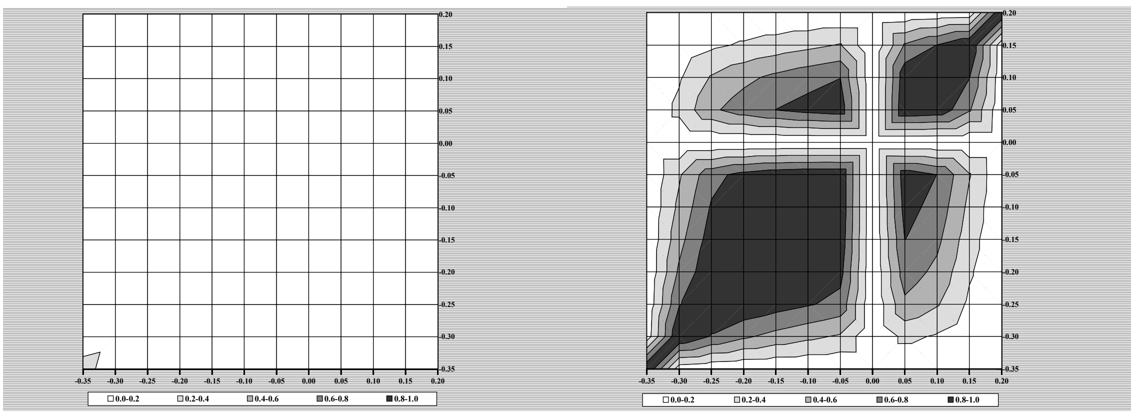
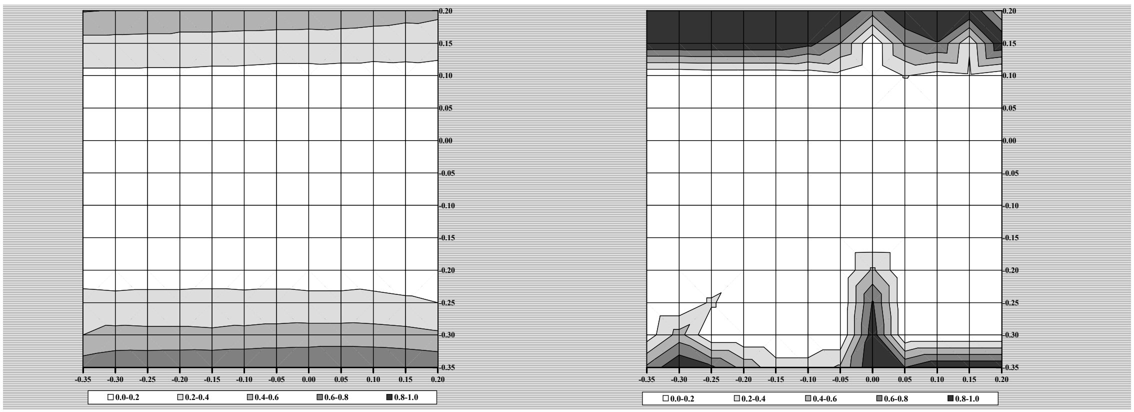


FIGURE 2: Misspecification analysis. Moran's Test. Percentage of rejections of the null hypothesis.

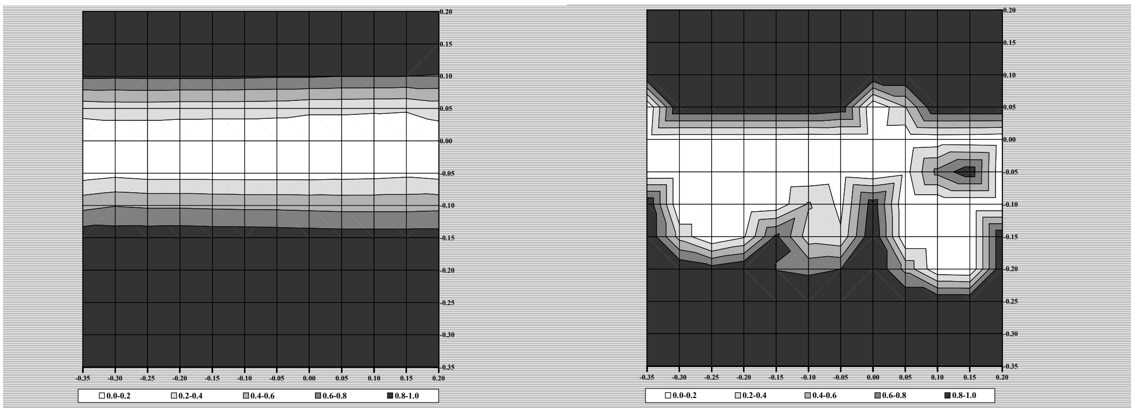
Sampling size: R=17.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=74.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=120.

Scale: δ =0 **Scale: δ =100**

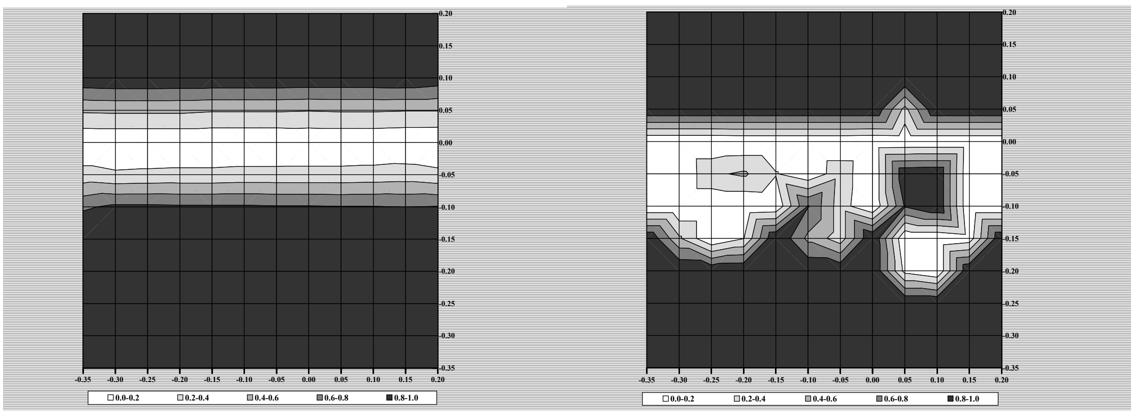
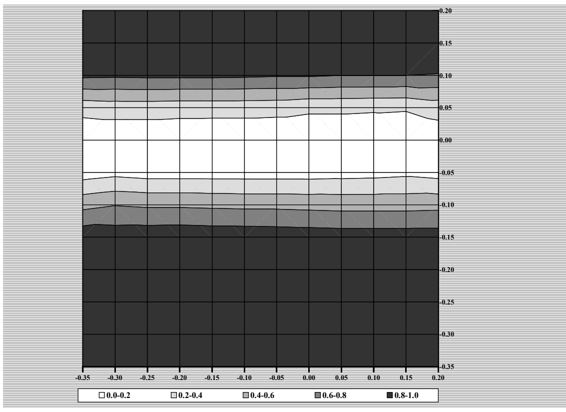


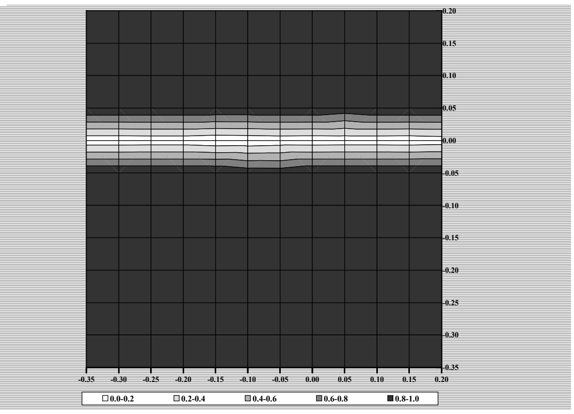
FIGURE 3: Misspecification analysis. LM-LAG Test. Percentage of rejections of the null hypothesis.

Sampling size: R=17.

Scale: δ =0

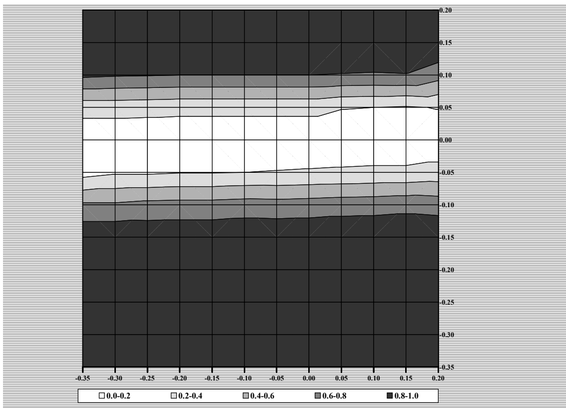


Scale: δ =100

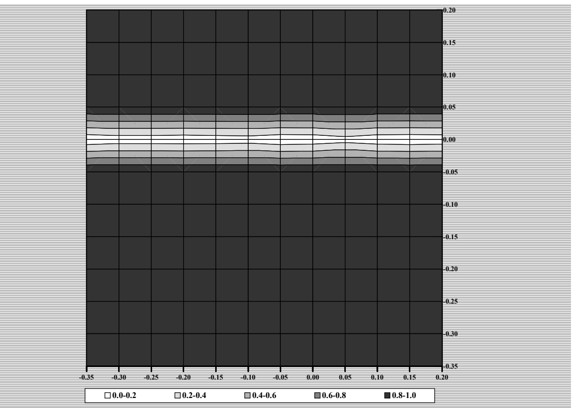


Sampling size: R=74.

Scale: δ =0

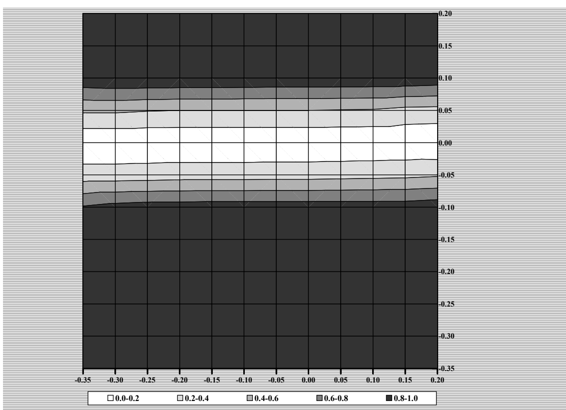


Scale: δ =100



Sampling size: R=120.

Scale: δ =0



Scale: δ =100

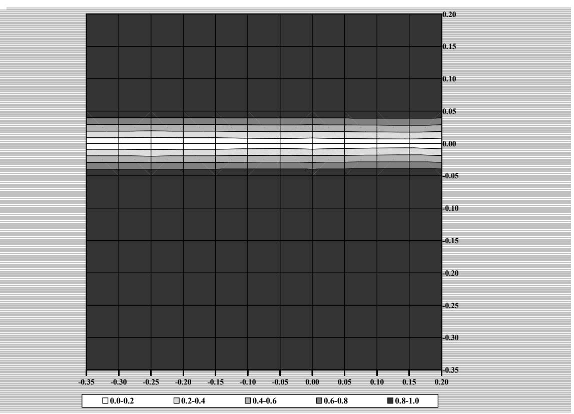
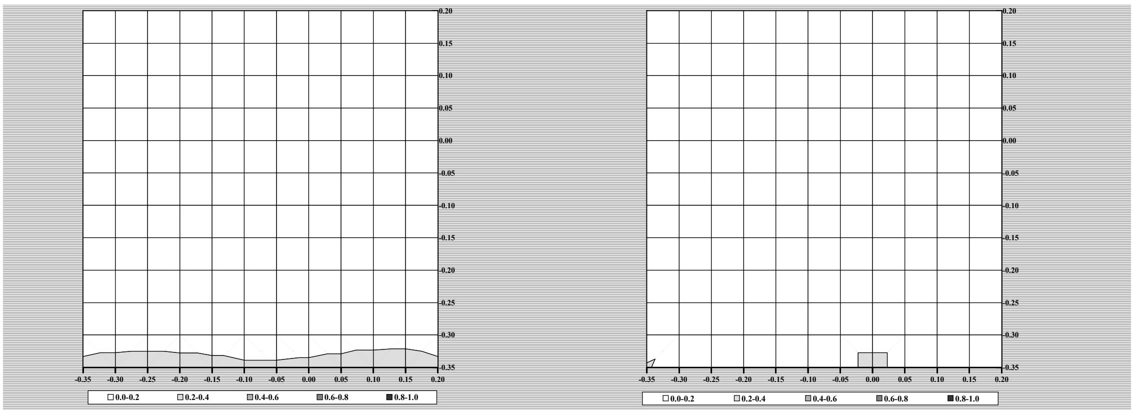


FIGURE 4: Misspecification analysis. LM-EL Test. Percentage of rejections of the null hypothesis.

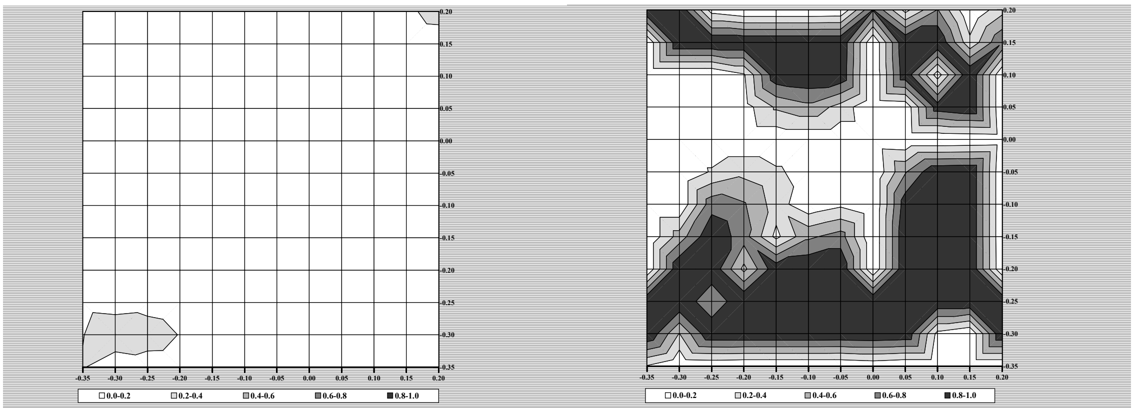
Sampling size: R=17.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=74.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=120.

Scale: δ =0 **Scale: δ =100**

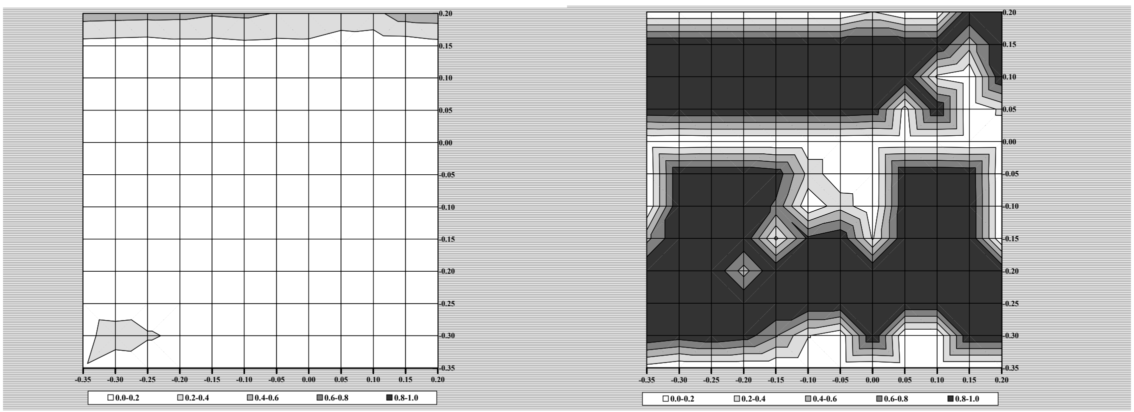
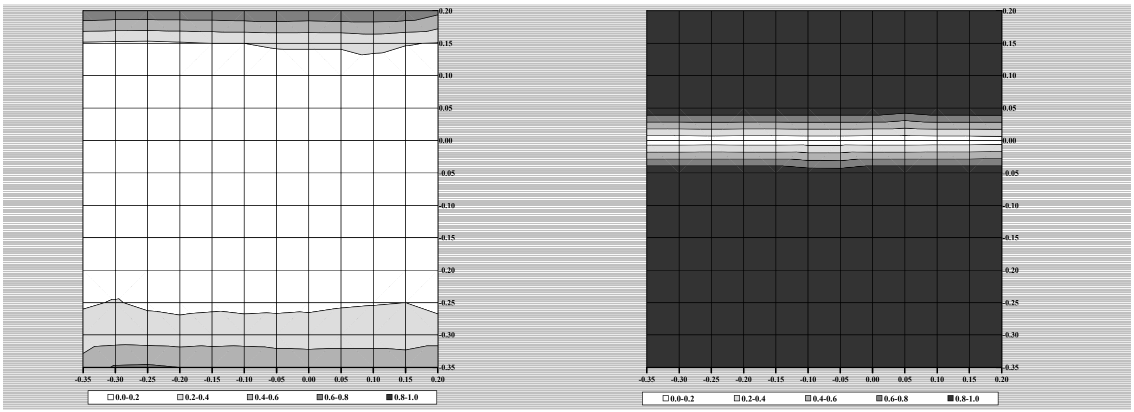


FIGURE 5: Misspecification analysis. LM-LE Test. Percentage of rejections of the null hypothesis.

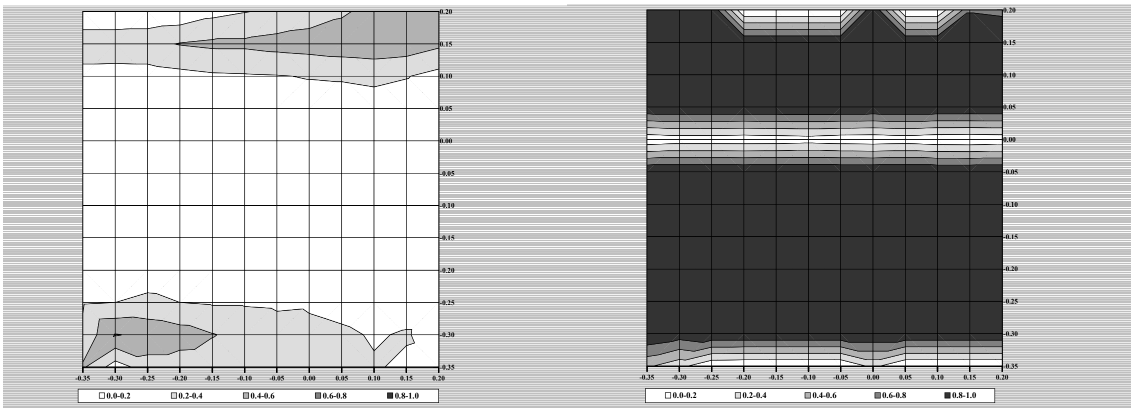
Sampling size: R=17.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=74.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=120.

Scale: δ =0 **Scale: δ =100**

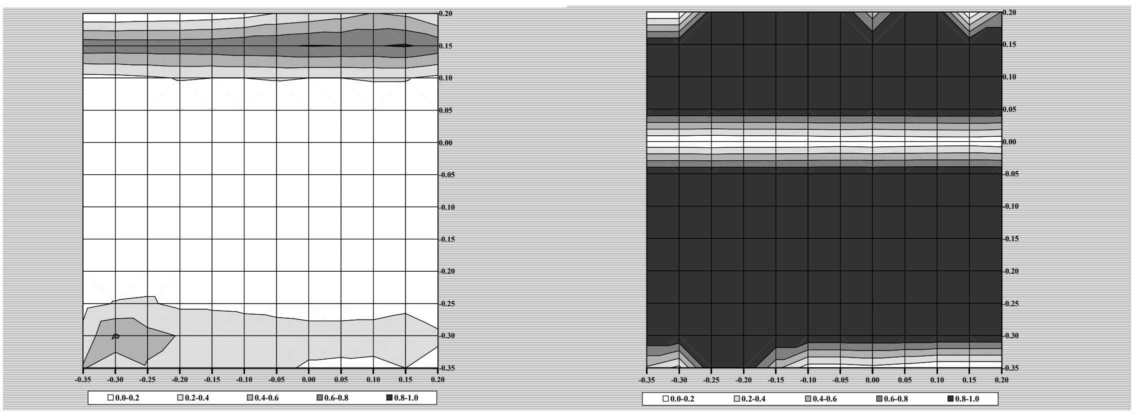
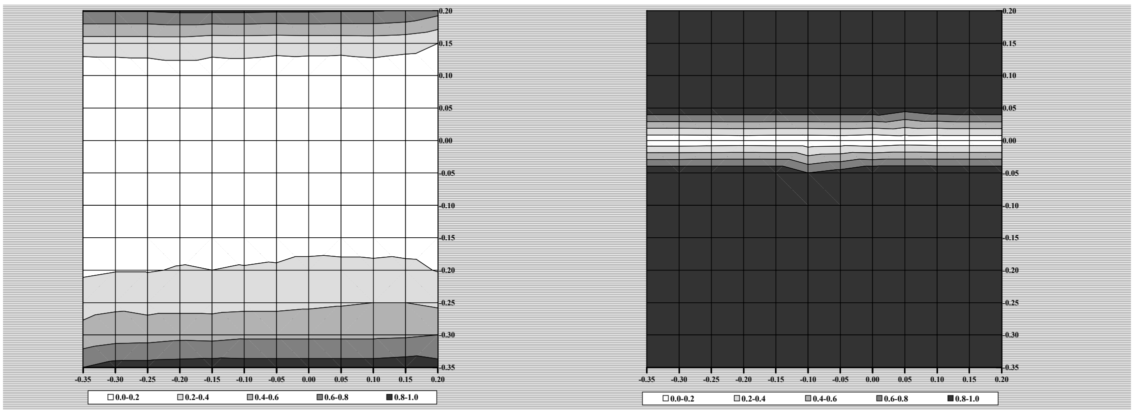


FIGURE 6: Misspecification analysis. SARMA Test. Percentage of rejections of the null hypothesis.

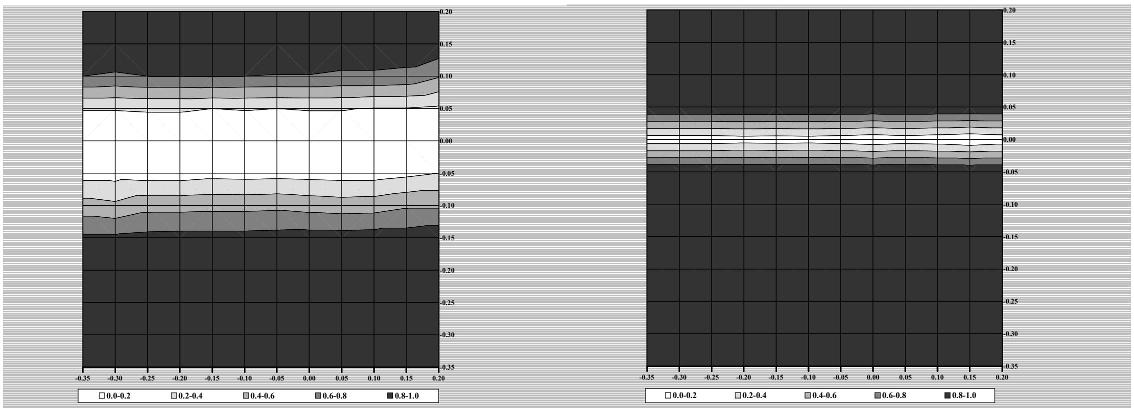
Sampling size: R=17.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=74.

Scale: δ =0 **Scale: δ =100**



Sampling size: R=120.

Scale: δ =0 **Scale: δ =100**