Scale Elements in Spatial Autocorrelation Tests

By

Jesús Mur

F. Javier Trívez

Abstract:

Moran's I is a good test for detecting relationships of cross-sectional dependency in spatial series. However, its behavior is sensitive to the scale of the process in autoregressive series. When the coefficient of variation of the process is high, the power of the test is zero in a not unimportant range of values of the coefficient of autocorrelation.

In this paper we point out some solutions and discard others. Among the latter the redimensioning of the series before resolving the I test stands out. The chapter of proposals, in no case definitive, revolves around the exploitation of the heteroskedastic structure in a series with signs of spatial autocorrelation.

Address:

Department of Economic Analysis.
University of Zaragoza
Gran Vía, 2-4
Zaragoza (50005). SPAIN
E-Mail: jmur@posta.unizar.es
Phone: +34-976-761815
Fax: +34-976-761996

40TH CONGRESS OF THE EUROPEAN REGIONAL SCIENCE ASSOCIATION

BARCELONA. 30 AUGUST-2 SEPTEMBER 2000

1.- Introduction

In 1950 Moran proposed one of the most popular tools in the field of spatial statistics, the I test, which addresses a basic question: is there cross-sectional dependency in the variable or model analyzed? All the available evidence points to this being the most efficient test in this field.

The aim of this paper is not to question its suitability but to highlight certain weaknesses. In particular, we point out a minor question which is the dependency of the test in relation to the scale of the series. In some circumstances, the power of Moran's I tends towards zero when the scale of the series is high. In the following section we present the basic theory related to the test, while in the third we discuss the specific problems derived from scale. In the fourth section we present some alternatives capable of lessening these difficulties, checked in the replication carried out in the fifth section. The work ends with the sixth section dedicated to the presentation of the principal conclusions reached.

2.- Basic Aspects

Moran's test, as Arbia (1989) indicates, is not a coefficient of correlation but a statistic whose objective is to measure the degree of spatial autocorrelation that exists in the spatial distribution of a variable. Its expression is well known:

$$I = \frac{R}{S_0} \frac{\sum_{r,s}^{R} (y_r - \bar{y})\, w_{rs}\, (y_s - \bar{y})}{\sum_{r=1}^{R} \left(y_r - \bar{y}\right)^2} \tag{2.1}$$

with $\{y_r, r = 1, 2, ..., R\}$ the observations of the variable y in the R points of the space and $\bar{y}$ its corresponding sampling mean, w_{rs} (r,s =1, 2, ..., R) the elements of the specified contiguity matrix W, with $S_0 = \sum_{r,s} w_{rs}$. This statistic will take values in the interval $((R/S_0)m_{min}, (R/S_0)m_{max})$, with m_{min} and m_{max} the lowest and highest roots respectively of the matrix DWD with D=[I-(ll'/R)] where l is an (Rx1) vector of ones and I is the identity matrix (De Jong et al, 1984). If the spatial distribution of variable is random, Moran's I will take values close to zero.

Cliff and Ord (1972 and 1981) obtain the moments of the test under the null hypothesis of independence both under the normality assumption and when the distribution of the series is unknown (*randomization assumption*). The same authors show that, once again under the hypothesis of independence, the test tends asymptotically towards the normal distribution. The conditions that guarantee this approximation are relatively weak, as can be seen in Sen (1976). Using numerical integration methods we can obtain the exact distribution function of the test (Sen, 1990, Tiefelsdorf and Boots, 1995).

Furthermore, King (1981) demonstrates that the test built on Moran's I is a Locally Best Invariant (LBI) test in the neighbourhood of the null hypothesis and, under certain conditions, a Uniformly Most Powerful Invariant (UMPI) test. Burridge (1980) obtains its formal equivalent with the Lagrange Multiplier (LM), corroborating the results put forward by Cliff and Ord (1981), while Anselin and Rey (1991) demonstrate, using Monte Carlo methods, its superiority as a test of spatial autocorrelation. All these results confer on Moran's I test a central position in the field of spatial analysis.

Nevertheless, this test presents certain limitations which it is necessary to consider. Moran's I is insensitive to the size and the shape of the units of observation, giving rise to the problem of "*topological invariance*" described by Dacey (1965). This weakness is shared by practically all the spatial autocorrelation tests, with the exception of that proposed by Dacey himself. Another question is that no model is specified in the alternative hypothesis. The only clue the researcher has is that the distribution of the variable in question maintains a certain regularity in relation to the spatial structure reflected in W. Lastly, it is necessary to suppose that, under the assumption of independence, the first order moment of the variable analyzed is homogeneous in space. This is why the sampling mean is used in the definition of test in (2.1). If this assumption were unable to be maintained, the use of the normal should be rethought in favor of the focus on randomization.

3.- The Problems

The principal limitation we find above is the lack of a well-defined alternative hypothesis. When the user rejects the null hypothesis of independence, it is not clear what the relevant model should be. Most applications choose an SAR structure, but there is no reason to justify this preference (why not an SMA?).

This decision is generally taken with very little reasoning. An SAR process implies that the structure of dependencies in the series is of a general type, in the sense that all the regions interact with all the others without exception, however inaccessible they may be from the others. In practice this means that all the covariances should be different to zero. Furthermore when an SAR process is imposed on a spatial series, we are also affirming that its expected value is not homogeneous, a characteristic that most spatial series seem to share. However, the first order moment of these processes should also respond to the spatial structure reflected in the contiguity matrix, a restriction which it does not seem very reasonable to impose in a general way.

These observations have an immediate reflection in the behavior of Moran's I. In particular, the sampling mean that intervenes in its definition will be an unbiased and consistent estimator of the expected moment when the Data Generating Process (DGP) of the series is an SMA:

$$y = \mu l + (I - \delta W)u = \mu l + Bu$$

$$\bar{y} = \frac{l'y}{R} \Rightarrow \begin{cases} E[\bar{y}] = \mu & \rightarrow \lim\limits_{R\to\infty} E[\bar{y}] = \mu \\ V[\bar{y}] = \dfrac{\sigma^2}{R}(1 + \delta^2 \dfrac{l'W^2 l}{R} - 2\delta \dfrac{S_0}{R}) & \rightarrow \lim\limits_{R\to\infty} V[\bar{y}] = 0 \end{cases} \quad (3.1)$$

with l a vector of ones and B=I-δW. The first order moment in the SAR case is not homogeneous so the sampling mean is a statistic lacking in meaning:

$$y = \mu l + \delta W y + u \Rightarrow y = B^{-1}(\mu l + u) \Rightarrow E[y] = \mu B^{-1} l \quad (3.2)$$

It is centered on the sampling mean of the expected values ($E[\bar{y}] = \mu(l'B^{-1}l)/R$), andc can be consistent in relation to this point ($\lim\limits_{R\to\infty} V[\bar{y}] = \sigma^2(l'B^{-2}l)/R^2 = 0$). However, as an estimator of an hypothetically homogeneous factor of scale it leads to errors.

In the specification of Moran's I, the sampling mean of the series is effectively introduced to obtain both the covariance of the numerator and the variance of the denominator. The question we raise now is how Moran's statistic responds to the scale.

The comments above suggest that the inclusion of the sampling mean is justified when the DGP of the series is an SMA. If, for this type of process, we write the I statistic in matrix notation:

$$I = \frac{R}{S_0} \frac{\sum\limits_{r,s}^{R} (y_r - \bar{y}) w_{rs} (y_s - \bar{y})}{\sum\limits_{r=1}^{R} (y_r - \bar{y})^2} = \frac{R}{S_0} \frac{u'BDWDBu}{u'BDBu} \quad (3.3)$$

it is clear that the numerator and the denominator are quadratic forms of a vector of normal N(0,1) variates on a symmetric and singular matrix. The distribution of both quadratic forms will pertain to the chi-squared family, although they will not be independent. Using the results of Yule and Kendall (1950), we can approximate their expected value though:

$$E\left[\frac{N}{D}\right]=\frac{E(N)}{E(D)}\left[1+\frac{V(D)}{E(D)^2}-\frac{Cov(N,D)}{E(N)E(D)}\right]+o(R^{-2}) \tag{3.4}$$

where o(-) means "of smaller order than". Resolving the above expression, the expected value of Moran's I is:

$$E[I]\cong\frac{R}{S_0}\left[\frac{trBDWDB}{trBDB}\right]\left[1+\frac{tr(BDB)(BDB)}{(trBDB)^2}-\frac{tr(BDWDB)(BDB)}{tr(BDWDB)tr(BDB)}\right] \tag{3.5}$$

which can be approximated by:

$$E[I]\cong\frac{R}{S_0}\frac{\sum_{r=1}^{R}\lambda_r(1-\delta\lambda_r)^2}{\sum_{r=1}^{R}(1-\delta\lambda_r)^2}\left[1+\frac{\sum_{r=1}^{R}(1-\delta\lambda_r)^4}{\left[\sum_{r=1}^{R}(1-\delta\lambda_r)^2\right]^2}+\frac{\sum_{r=1}^{R}\lambda_r(1-\delta\lambda_r)^4}{\sum_{r=1}^{R}\lambda_r(1-\delta\lambda_r)^2\sum_{r=1}^{R}(1-\delta\lambda_r)^2}\right] \tag{3.6}$$

with λ_r the r[th] characteristic root of W. For positive values of the parameter δ, the expected value of I will be negative and positive for negative values of the former. The two quotients that appear in the square brackets are positive and less than one (using the Cauchy–Swartz inequality). Their contribution will become less significant as the sample size increases, so that the above expression can be reduced, for large sample sizes, to:

$$E[I]\to\frac{R}{S_0}\frac{\sum_{r=1}^{R}\lambda_r(1-\delta\lambda_r)^2}{\sum_{r=1}^{R}(1-\delta\lambda_r)^2} \tag{3.7}$$

Independently of the accuracy of these approximations, the relevant aspect of all of them is that at no moment is the distribution of Moran's I affected by the scale of the process, being efficiently neutralized by the sampling mean.

The above result allows us to present graphs such as that of Figure 3.1. With the continuous line we represent the expected value of Moran's I according to (3.7) and with the dotted lines the acceptance limits of the null hypothesis of independence at a significance level of 5% (that is 1.96xDT (I), where DT(I) is the standard deviation under the null hypothesis). The reference matrix, of the order (74x74), corresponds to the NUTS II European regional system of 12 member states. The stability interval (if by such we understand that in which it is true that $|\delta\lambda_r|<1$, $\forall$ r) associated with this matrix is (-0.31, 0.17).

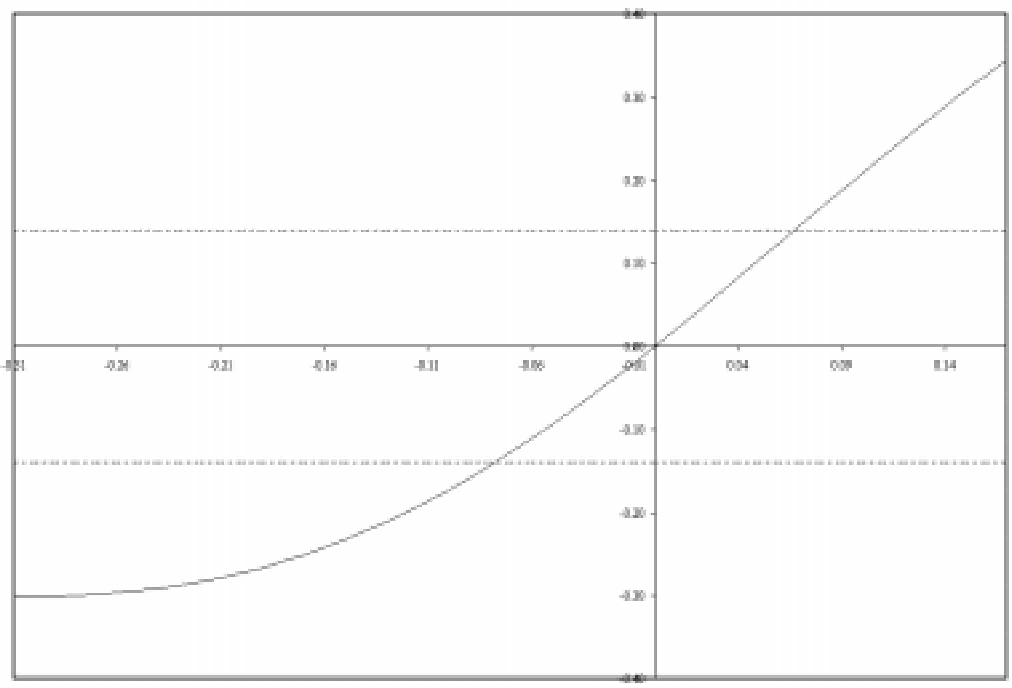
The situation in the SAR case is different, given that, as we mentioned before, the first order moment of the series is not constant, which means that the sampling mean will be a biased estimator of the scale factor. Taking (3.2) as a point of reference, it follows that:

$$E[\bar{y}] = \mu \frac{l'B^{-1}l}{R} \Rightarrow \begin{cases} \text{Si} \quad \delta < 0 \Rightarrow \quad 0 < \dfrac{l'B^{-1}l}{R} < 1 \Rightarrow \quad E[\bar{y}] < \mu \\ \text{Si} \quad \delta > 0 \Rightarrow \quad \dfrac{l'B^{-1}l}{R} > 1 \Rightarrow \quad E[\bar{y}] > \mu \end{cases} \tag{3.8}$$

In this type of process, Moran's I can be developed as:

$$I = \frac{R}{S_0} \frac{(u+\mu l)'B^{-1}DWD\,B^{-1}(u+\mu l)}{(u+\mu l)'B^{-1}D\,B^{-1}(u+\mu l)} \tag{3.9}$$

Figure 3.1: Expected value of Moran's I for SMA processes



Using the approximation of (3.4) again, its expected value can be expressed as:

$$E[I] \cong \frac{R}{S_0}\left[\frac{\mathrm{tr}\,B^{-2}DWD + c^2 l'B^{-1}DWDB^{-1}l}{\mathrm{tr}\,B^2D + c^2 l'B^{-1}DB^{-1}l}\right]\left[1+\frac{\mathrm{tr}\,B^{-2}DB^{-2}D + 4c^2 l'B^{-1}DB^{-2}DB^{-1}l}{(\mathrm{tr}B^{-2}D+c^2 l'B^{-1}DB^{-1}l)^2}\right.$$
$$\left. - \frac{\mathrm{tr}\,B^{-1}DB^{-2}DWDB^{-1} + 4c^2 l'B^{-1}DWDB^{-2}DB^{-1}l}{(\mathrm{tr}\,B^{-2}DWD + c^2 l'B^{-1}DWDB^{-1}l)(\mathrm{tr}\,B^{-2}D + c^2 l'B^{-1}DB^{-1}l)}\right] \quad (3.10)$$

with $c=\mu/\sigma^2$, the coefficient of variation of the process. The above result is intractable, although the probalistic limit of (3.9) can be considered as an approximation:

$$\mathrm{plim}I = \frac{\lim_{R\to\infty}\left[\mathrm{tr}(B^{-1}DWDB^{-1})/S_0\right] + c^2 \lim_{R\to\infty}\left[(l'B^{-1}DWDB^{-1}l))/S_0\right]}{\lim_{R\to\infty}\left[\mathrm{tr}(B^{-1}DB^{-1})/R\right] + c^2 \lim_{R\to\infty}\left[(l'B^{-1}DB^{-1}l)/R\right]} = \frac{n_1 + c^2 n_2}{d_1 + c^2 d_2} \quad (3.11)$$

The terms d_1 and d_2 are positive for any δ, while n_1 and n_2 will be negative for $\delta<0$ and positive if the opposite is true. Given that c^2 will also be positive, the presence of a factor of scale in the DGP of the series will not affect the sign of the test. However, as the scale increases the I statistic will tend towards the quotient (n_2/d_2), terms strictly associated with the scale (and with the error in the estimation of the first order moment).

This situation can be represented on a graph as it appears in Figure 3.2. The continuous line represents the quotient (n_2/d_2) obtained for the same contiguity matrix used in Figure 3.1. This is the limit of Moran's I when we increase the factor of scale of the series indefinitely. The dotted lines represent the acceptance limits of the null hypothesis of independence at a significance level of 5%.

Figures 3.1 and 3.2 are apparently similar, though in the first we are representing an expected value around which the finally observed value of the I statistic will fluctuate (SMA case), while in the second we represent the convergence limit of the same statistic (SAR case). That is to say that, in the first case, Moran's I will have a certain capacity to detect relationships of spatial autocorrelation even when the coefficient of the process is close to zero, while in the second there is a zone of values in which the test will not have any power at all (if the scale is high).

In Figures 3.3 and 3.4 we reproduce the results of a small simulation (100 replications in each experiment) carried out to confirm this effect. The contiguity matrix used is the same (that corresponding to the European regional system of 12 member states, binary and of the order 74x74). Random series have been obtained from an N(0,1) distribution, and then transformed in SMA or SAR processes with a factor of scale ranging between 0 and 1000. In the graphs the percentage of rejections of the null hypothesis of incorrelation is reflected, obtained in each case using Moran's test.

Figure 3.2: Limit of the expected value of Moran's I for SAR processes.

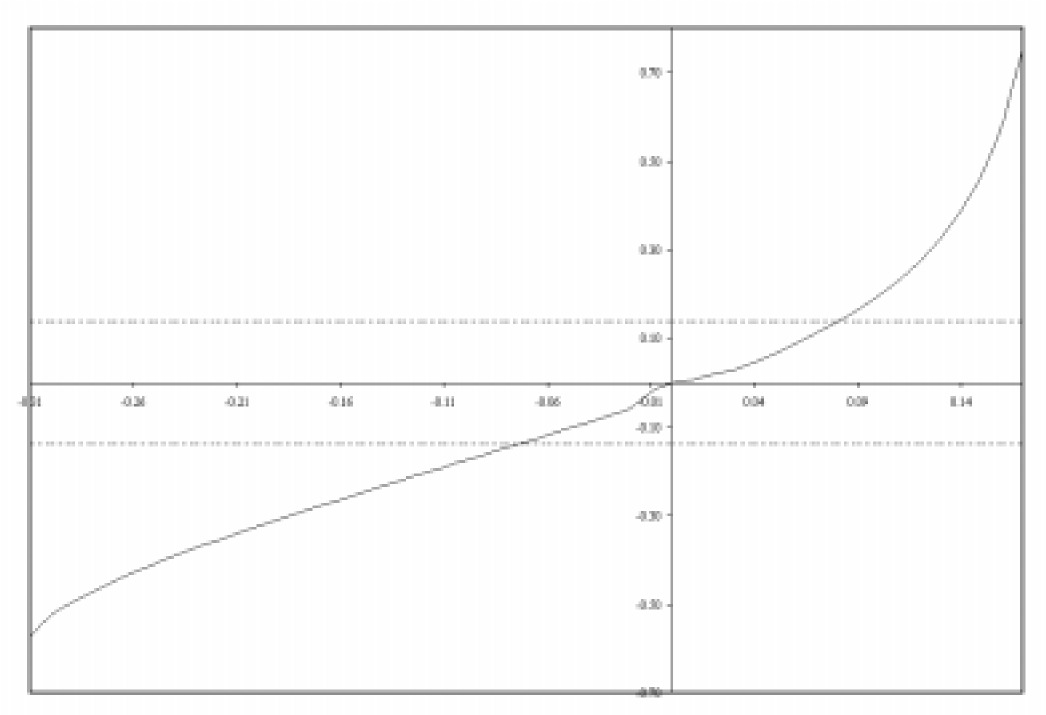


Figure 3.3: Estimated power function of Moran's I. SMA case.

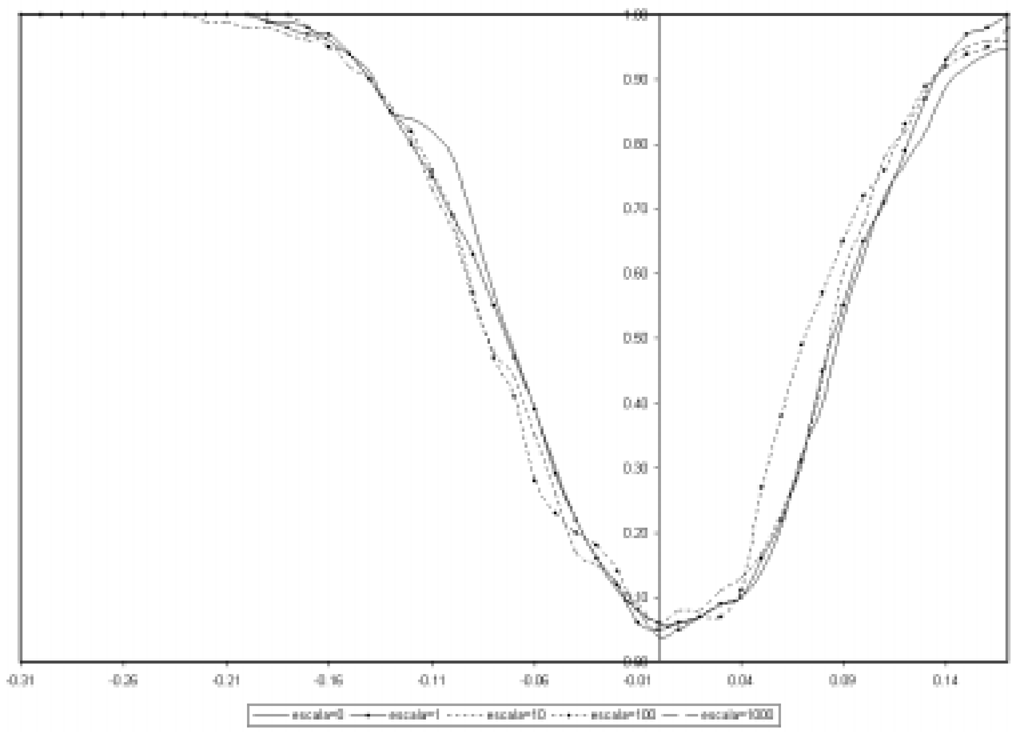
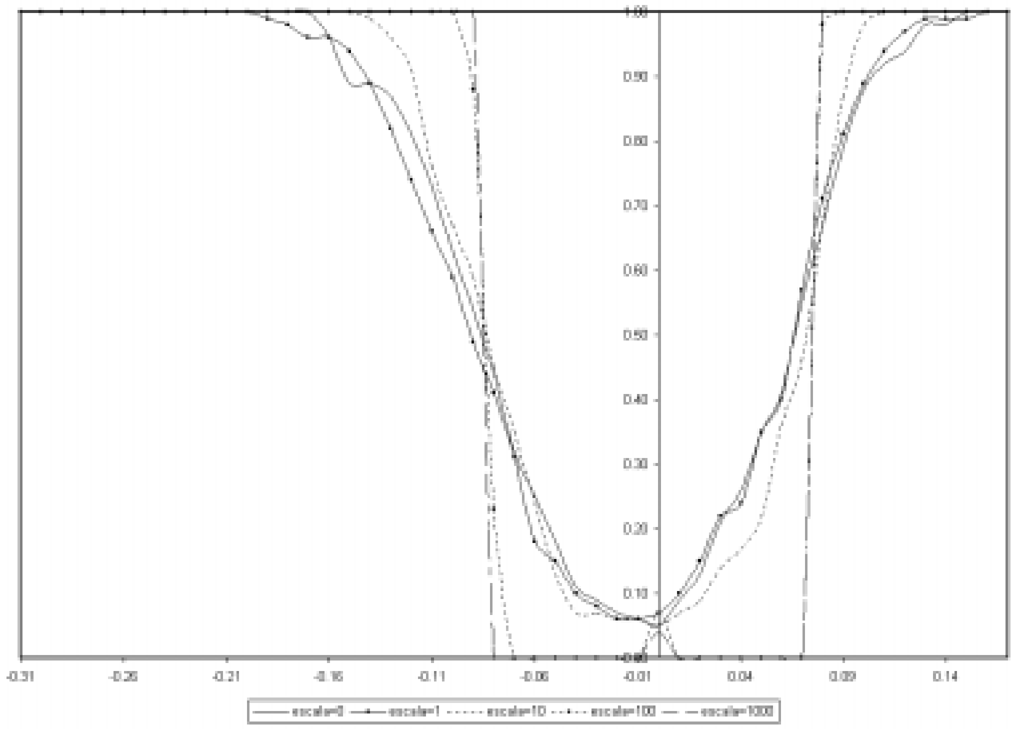


Figure 3.4: Estimated power function of Moran's I. SAR case.



These graphs tend to corroborate previous comments. Moran's I does not show any special sensitivity to the factor of scale when it is applied to SMA processes, while its impact is evident in SAR structures. In this case, and when the scale of the series is 1000, the power of the test is zero for values of the coefficient of autocorrelation between –0.08 and 0.07. In the rest of the parametric space the power is one.

4.- Some Proposals

The problem noted in the previous section affects a restricted zone of the parametric space (dependent on the contiguity matrix) and occurs only when the series is of SAR type. It is not a critical problem but does create certain inconveniences. Some solutions appear obvious, such as using a logarithmic transformation on the series before resolving the test. In this paper we present another proposal that may be useful when the sample size is not very large, given that it implies using the eigenvectors and eigenvalues of the contiguity matrix W.

If the series is an SAR, its DGP is given by:

$$y = \mu l + \delta W y + u \Rightarrow y = [I - \delta W]^{-1}(\mu l + u) = B^{-1}(\mu l + u) \qquad (4.1)$$

where l is a vector of ones of (Rx1) order, μ the factor of scale, δ the parameter of autocorrelation and u a white noise vector $N(0,\sigma^2 I)$. If it is an SMA, the associated DGP will be:

$$y = \mu l + [I - \delta W]u = \mu l + Bu \quad (4.2)$$

In both cases, the series y and u are referenced to the canonical base ($\vec{e}$), although we can also use other bases such as that composed by the eigenvectors of B ($\vec{q}$), coincident with those of W and independent of the parameters of the process. If we order these vectors in the columns of matrix Q, the mapping matrix from one base to the other will be Q' which, applied to (4.2) allows us to write:

$$Q'y = \mu Q'l + Q'Q[I - \delta\Lambda]Q'u \Rightarrow y^* = \mu q^* + \Delta u^* = \mu q^* + v^* \quad (4.3)$$

where y*=Q'y and u*=Q'u are the co-ordinates of the original vectors y and u in the new base $\vec{q}$ and q*=Q'l. Δ is the matrix of characteristic roots of B which depends on the corresponding matrix of roots of W (Λ). The transformed noise vector maintains the characteristics of the original: $u^* \sim N(0,\sigma^2 I)$. Given that it is not observable, its incidence is resumed in the random term v*, different to the previous one in that it is heteroskedastic: $v^* \sim N(0,\sigma^2\Delta^2)$.

If we resolve a similar transformation in the SAR process of (4.1) we obtain:

$$Q'y = Q'Q[I - \delta\Lambda]^{-1}(\mu Q'l + Q'u) \Rightarrow y^* = \Delta^{-1}(\mu q^* + u^*) = \Delta^{-1}\mu q^* + w^* \quad (4.4)$$

with $w^* = \Delta^{-1}u^* \sim N(0,\sigma^2\Delta^{-2})$. Lastly, when the series is composed of a factor of scale and a white noise without spatial structure (δ=0), the filter above leads to:

$$Q'y = \mu Q'l + Q'u \Rightarrow y^* = \mu q^* + u^* \quad (4.5)$$

The last three equations (4.3), (4.4) and (4.5) allow us to design a strategy of analysis of spatial series based on the following considerations:

(I)- Given that the matrix of eigenvectors Q is not singular and known (it is directly associated with W), the filtering of the original series should not have any effect on the quality or the quantity of information contained in the sample.

(II)- If a factor of scale intervenes in the DGP of the original series, a term q*=Q'l should appear in the DGP of the filtered series.

(III)- When there is any spatial structure (of SAR or SMA type) in the DGP of the original series, the error term linked to the filtered series will be heteroskedastic. Also, the heteroskedastic function will respond exclusively to the series of eigenvalues of the contiguity matrix.

(IV)- The systematic part of the equation that describes the DGP of the filtered series will be linear in the variable q* (proportional in accordance with 4.3) in the SMA case, and non-linear in the SAR case (equation 4.4).

This strategy can be made operative using, as a testing equation:

$$y^* = \alpha + \beta q^* + n^* \qquad (4.6)$$

in which it will be true that $\alpha=0$, $\beta=\mu$ and $n^*=u^*$ when the original series is a white noise. If the DGP is of SMA type, it will be verified that $\alpha=0$, $\beta=\mu$ and $n^*=v^*$. Lastly, when the DGP is of SAR type, the expansion of (4.4) leads to:

$$y^* = \mu q^* + (\mu\delta)\left[\Lambda q^*\right] + (\mu\delta^2)\left[\Lambda^2 q^*\right] + \ldots + w^* \qquad (4.7)$$

so that, in (4.6), $\alpha=0$, $\beta=\mu$ and $w^* = v^* + \mu \sum_{j=1}^{\infty} \delta^j \left[\Lambda^j q^*\right]$. The LS estimation of (4.6) will produce unbiased and consistent estimators of α and β when the DGP is of the first or second type (white noise with scale or SMA), but they will be biased and inconsistent in the SAR case. This bias can be corrected, at least partially, using a testing equation such as:

$$y^* = \alpha + \beta_1 q_1^* + \beta_2 q_2^* + \ldots + \beta_p q_p^* + w^* \qquad (4.8)$$

with $q_j^* = \Lambda^j q^*; j = 0,1,\ldots,p$. For a sufficiently high value of p we can expect that the impact of the specification error (which will still exist in the SAR case) on the LS estimation will be moderate. In any case, the relevant aspect is that if heteroskedasticity is detected in the error term of (4.8), associated explicitly with the structure of the contiguity matrix, this will be an unequivocal sign of spatial autocorrelation in the original series.

Among the various possibilities that exist, the Goldfeld-Quandt test appears to bring together the principal requirements:

- It is easy to obtain.

- Its distribution is known for all sample sizes.
- Spatial structure can be introduced explicitly in the test process.

In respect to the latter, it must be taken into account that in the disturbance of the SMA process of (4.2) it will be true that:

$$V\left[v_r^*\right]=\sigma^2(1-\delta\lambda_r)^2 \Rightarrow \frac{\partial V\left[v_r^*\right]}{\partial \lambda_r}=-2\delta(1-\delta\lambda_r)\begin{Bmatrix}<0 \Leftrightarrow \delta>0\\ >0 \Leftrightarrow \delta<0\end{Bmatrix}\forall r \qquad (4.9)$$

while in SAR type processes:

$$V\left[w_r^*\right]=\frac{\sigma^2}{(1-\delta\lambda_r)^2} \Rightarrow \frac{\partial V\left[w_r^*\right]}{\partial \lambda_r}=\frac{2\delta\sigma^2}{(1-\delta\lambda_r)^3}\begin{Bmatrix}>0 \Leftrightarrow \delta>0\\ \\ <0 \Leftrightarrow \delta<0\end{Bmatrix}\forall r \qquad (4.10)$$

In both cases, for a concrete value of δ, the variance evolves systematically with λ_r. This is all the information that we need for resolving the Goldfeld-Quandt test. The values of y* will be ordered highest to lowest (or lowest to highest) with λ_r; the central c observations will be excluded (our experience corroborates the normal practice of excluding a third of the sample); the equation (4.8) will be estimated by LS in the first and last subsamples and the test statistic will be obtained as:

$$GQ=\frac{SR_{MAX}}{SR_{MIN}} \sim F_{\left(\frac{R-c}{2}-k;\frac{R-c}{2}-k\right)} \qquad (4.11)$$

where SR_{MAX} and SR_{MIN} are the highest and lowest residual sums respectively and k=p+1.

In the context in which we have set the discussion, another attractive alternative is the Breusch-Pagan (BP) LM test. In this case, the concrete functional form associated with the heteroskedastic variance is unknown (it could be 4.9 or 4.10), although its arguments are identified (the eigenvalues of the contiguity matrix W). Other possibilities (the tests of Szroeter, 1978 and White, 1982) have been considered but without much success.

One result derived from the above discussion is that, if heteroskedastic relationships have been detected in the filtered series, this information can be used to try to identify the DGP of the series. In accordance with (4.3), if the series is of SMA type, the existence of a scale different to zero in the DGP will result in $\beta \neq 0$. Also, if the variance of the error term n* responds to the sequence of values $\{\lambda_r, r=1, 2, \ldots, R\}$, the conclusion is that the original series presents a structure of spatial correlation. On the other hand, if the DGP is of SAR type, the equation of reference is (4.4) expanded in (4.7). That is to say that what differentiates both types of process is that the autoregressive requires a wider structure of regressors (q^*, Λq^*, $\Lambda^2 q^*$, …..) than that of the moving average (only q^*).

This reasoning can be developed in different ways. In the first place, (4.3) and (4.4) are different functional forms, which could give rise to a model selection strategy based on examining the suitability of each functional form or on some other more specific criterion. Another possibility could be to use (4.8) as a nesting equation of both processes to contrast the restrictions that lead to an SMA structure ($\beta_2 = \beta_3 = \ldots = \beta_p = 0$) and to an SAR structure (which will be non-linear). Another less rigorous option, although of simpler resolution, could be to accept an SMA structure initially when signs of autocorrelation have been found. The adoption of SAR structures instead of SMA ones would be carried out only when there is strong evidence in their favour. This strategy could be carried out by means of a simple testing equation:

$$y^* = \alpha + \beta_1 q_1^* + \beta_2 q_2^* + w^* \tag{4.12}$$

estimated by LS. The acceptance of $\beta_2=0$ implies the adoption of an SMA structure while its rejection leads to the adoption of an SAR structure. The test statistic could be the t-ratio associated with this parameter with the peculiarity that, given that the error term of that equation is heteroskedastic, a consistent estimation of the covariance matrix of the LS estimators must be used. In this sense, the proposal of White (1982) is very useful because it generates consistent estimations of the matrix even under certain subspecifications in the model (White, 1980), among which we can include errors in the functional form as in the case of (4.12) in relation to the SAR structure presented in (4.4).

5. - A small simulation

In the third section we have commented that Moran's I is sensitive to the scale of the series when this has been generated by an SAR type process. This limitation becomes a problem when it occurs in zones of the parametric space in which the test has zero power. In the fourth section some solutions have been put forward: taking logarithms on the series before obtaining the autocorrelation test or resorting to one of the heteroskedasticity tests on the filtered series. The usefulness of these proposals will be checked below by means of a small Monte Carlo exercise whose most relevant characteristics are the following:

- For the moment we have replicated just one sample size (R=74).
- We have used only one contiguity matrix of binary type and of order (74x74) corresponding to the system of European regions (NUTS II level) of 12 member states.
- Series of random numbers of order (74x1) have been obtained from a N(0,1) distribution, which have later been transformed in SAR or SMA processes using (4.1) or (4.2). In each case 100 replications have been resolved.
- Different factors of scale have been used in order to analyze their impact on each test. The replicated values were μ =0, 1, 10, 100 and 1000.
- The stability interval associated with the contiguity matrix is (-0.31, 0.17).

The most relevant results appear in Figures 5.1 to 5.6 in terms of the percentage of rejections of the null hypothesis of independence. Figures 5.1 and 5.2 refer to the GQ test. In the first an SMA has been used as the DGP and in the second an SAR. Figures 5.3 and 5.4 show the performance of the BP test, In the last two figures, 5.5 and 5.6, Moran's I has been used on the logarithms of the original series (to guarantee their existence only scales 10, 100 and 1000 have been replicated.

The testing equation used both for GQ and for BP has been 4.8, fixing p as 3. This value seems to mantain a certain equilibrium between the over-specification that exists when the series has been generated by an SMA (which results in a loss of power of the tests) and the sub-specification characteristic of the SAR case (of which the result is estimated power curves that seem anomalous). Lastly, the heteroskedastic hypothesis of the BP test has been specified using λ_r, λ_r^2 y λ_r^3 as regressors, with the aim of buffering the effects of the error in functional form that exists in the SAR case.

The results collected in these figures allow us to highlight some provisional conclusions:

- As was foreseeable, there is no scale effect either in the GQ test or in the BP test when they have been applied to SMA series. However, signs are still appreciable in the SAR case, more clearly with the BP test. Their incidence can be diluted by increasing the order of p in the testing equation (4.8), at the cost of a progressive worsening of the power of both tests.

- The GQ test tends to overestimate the size of the test while BP tends to underestimate it. In the first case, the percentage of rejections observed for a zero value of the parameter of autocorrelation is systematically above 5%, in a range comprised between 6% and 10%. The size estimated in the BP test is closer to the theoretical significance level of 5%, although with a tendency to fluctuate between 3% and 4%. We believe this is due to a lack of precision in the obtention of the eigenvectors of the contiguity matrix.
- The estimated power for the GQ test is clearly superior to that of the BP test in all cases.
- The logarithmic transformation of the series does not prevent the scale effect characteristic of Moran’s I in SAR processes. The appearance of Figures 5.6 and 3.4 is similar (the same is true for 3.3 and 5.5 in the SMA case), with a slight reduction in the range of zero power (now it is –0.06 to 0.06). However, other problems arise such as the increase in the size of the test as the scale of the process grows (it is 12% with a scale of 1000), or some misleading covariances in the numerator of the test which lead to the change of the sign of the sampling Moran’s I.

Figure 5.1: The Goldfeld-Quandt test. SMA case.

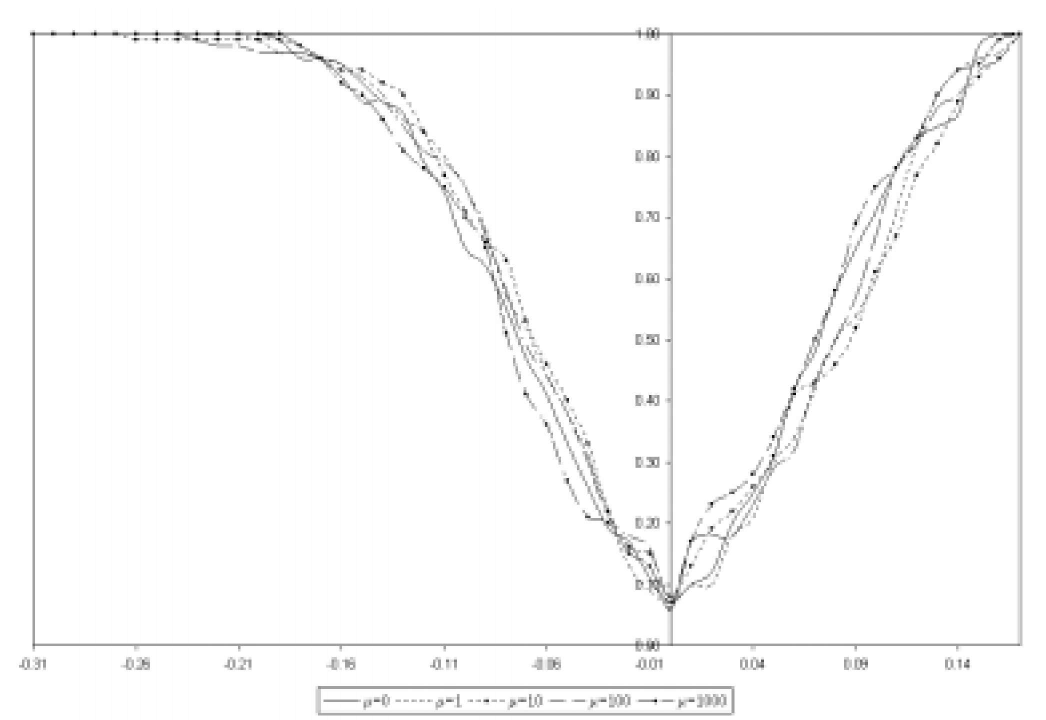


Figure 5.2: The Goldfeld-Quandt test. SAR case.

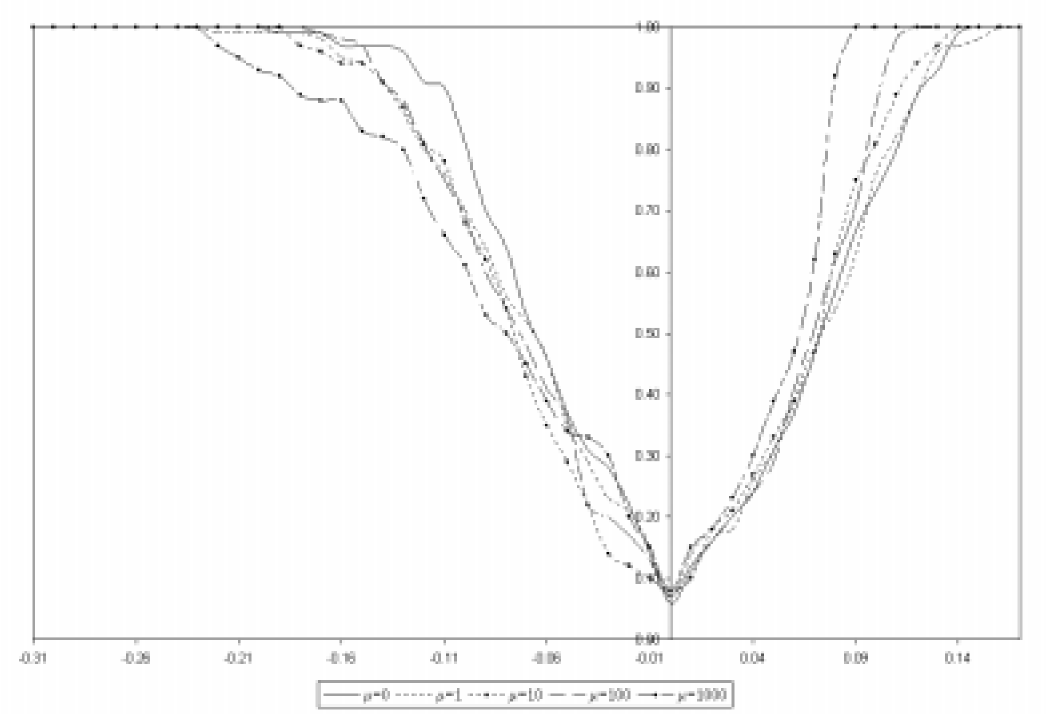


Figure 5.3: The Breusch-Pagan test. SMA case.

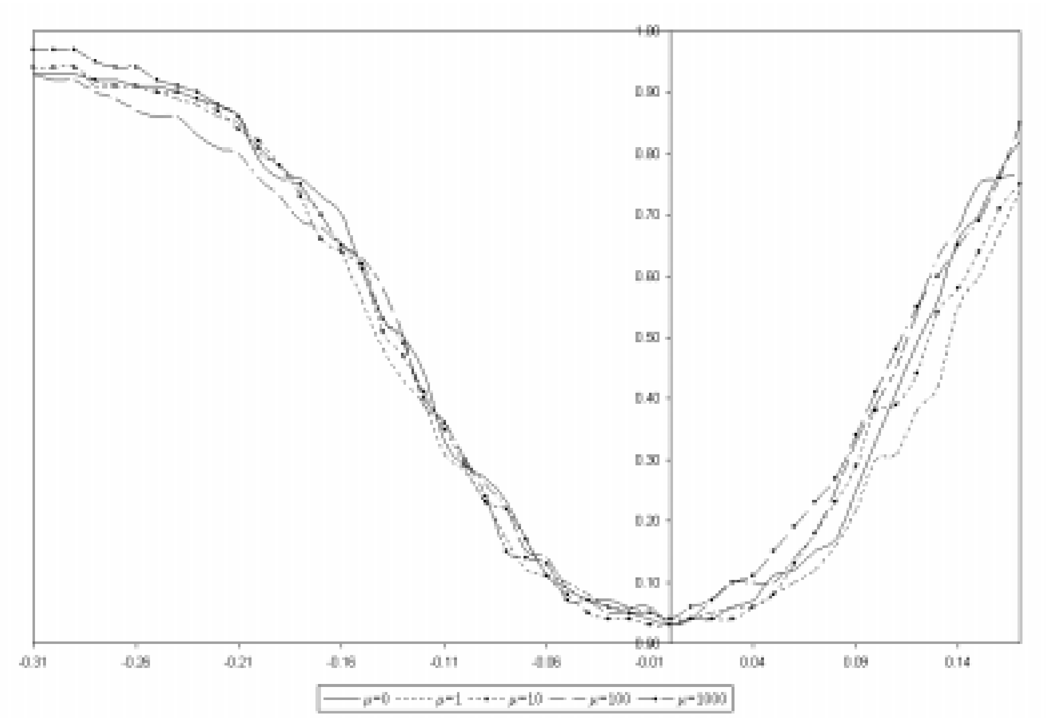


Figure 5.4: The Breusch-Pagan test. SAR case.

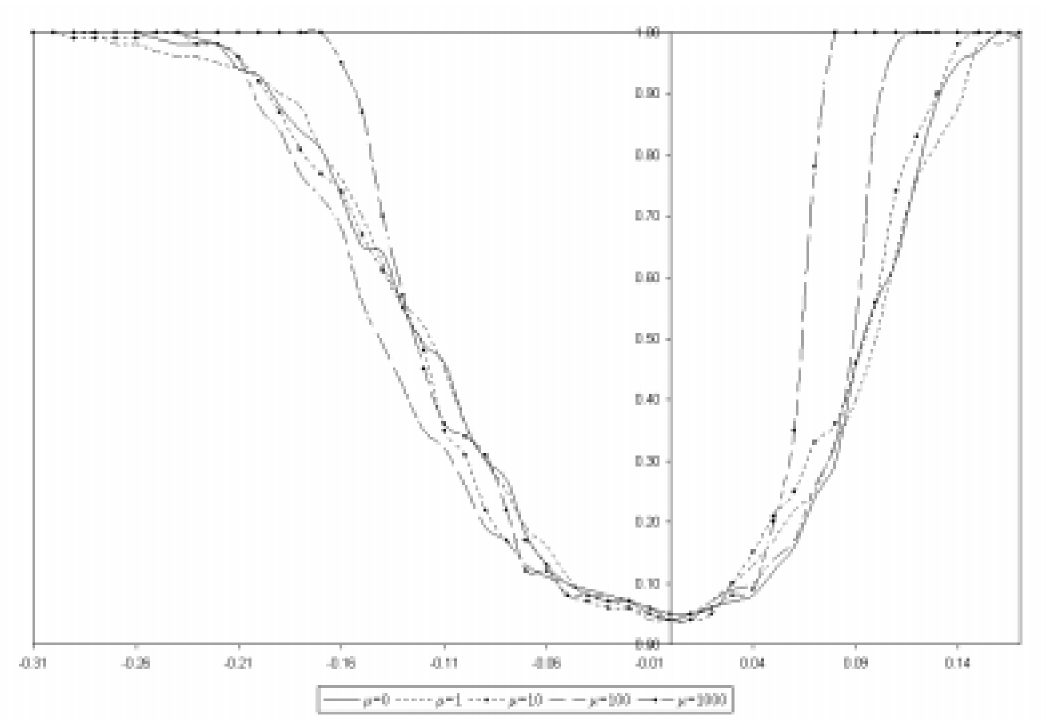


Figure 5.5: Moran's I test. Logarithmic Transformation. SMA case.

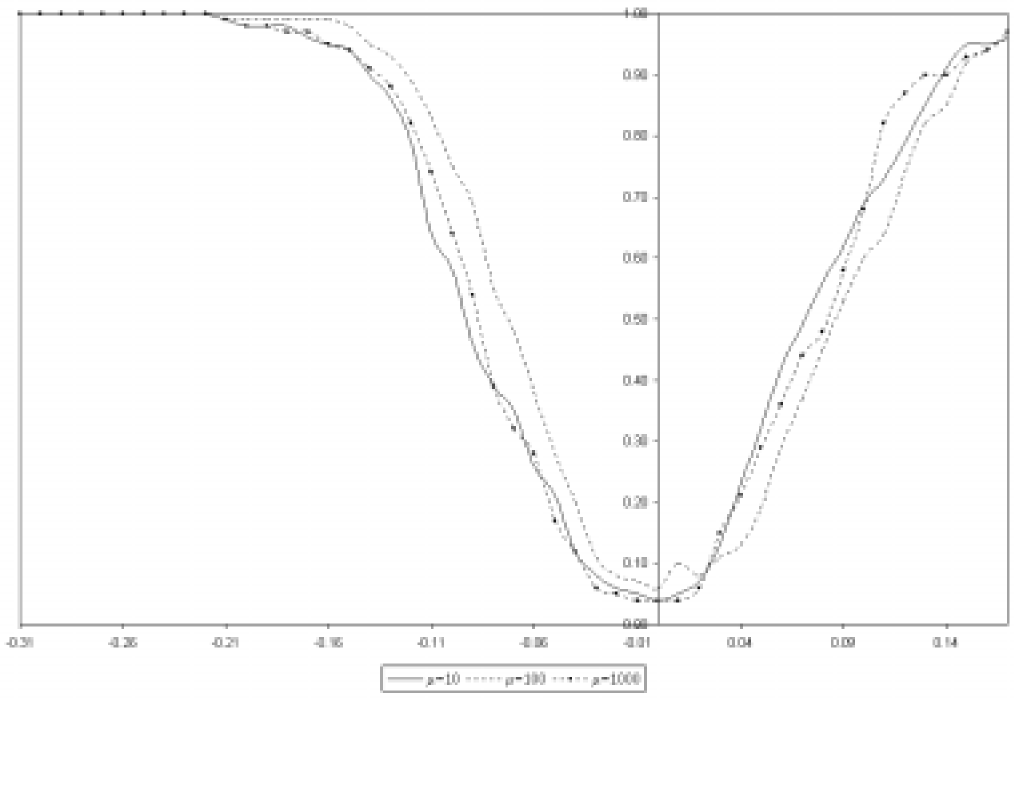
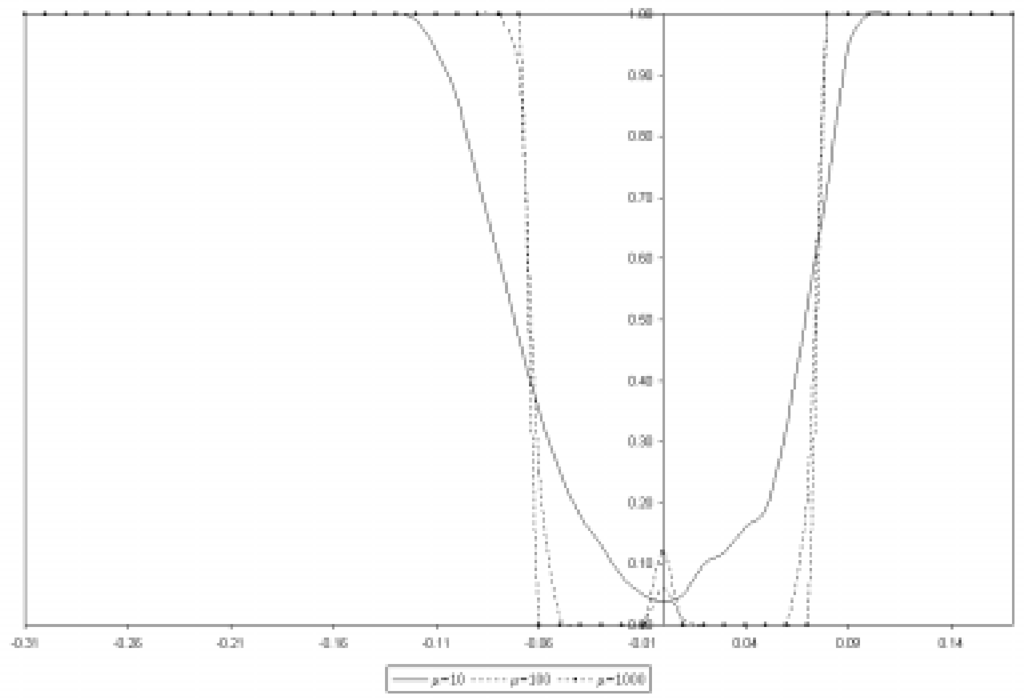


Figure 5.6: Moran's I test. Logarithmic Transformation. SAR case.



6. - Conclusions

Moran's I is an efficient test for detecting relationships of cross-sectional dependence in spatial series. However, its behavior is sensitive to the scale of the process in series of SAR type. When the coefficient of variation of the series is high, the power of the test is zero for a not unimportant range of values of the coefficient of autocorrelation.

In this paper we have noted some solutions and discarded others. Among the latter the redimensioning of the series before resolving Moran's I stands out. In the list of proposals, the GQ and BP tests offer certain guarantees though the solution doesn't seem to be definitive. It is necessary to elaborate a more structured and consistent analysis framework where the impact of scale in SAR processes can be absorbed. It is also necessary to extend the cases analyzed in order to contemplate different types of scales and of samples sizes. Lastly, another aspect to consider is what happens, with regard to the same question of the factor of scale, with the set of LM tests used in the specification of the spatial dynamics in a causal econometric model.

Acknowledgements

This research has been financed by the CICYT-FEDER 2FD97-1004-C03-03 project for whose collaboration we are sincerely thankful.

References

Anselin, L. and S. Rey (1991): Properties of Tests of Spatial Dependence in Linear Regression Models. *Geographical Analysis*, 23, 112-131.

Arbia, G. (1989): *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Dordrecht: Kluwer.

Breusch, T. and A. Pagan (1979): A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47, 1287-1294.

Burridge, P. (1980): On the Cliff-Ord Test for Spatial Correlation. *Journal of the Royal Statistical Society B*, 42, 107-108.

Cliff, A. and K. Ord (1972): Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*, 7, 267-284.

Cliff, A. and K. Ord (1981): *Spatial Processes. Models and Applications*. London: Pion.

Dacey, M. (1965): A Review of Measures of Contiguity fir Two and k-colour Maps, in *Spatial Analyses: A Reader in Statistical Geography* (pp. 479-495). Ed. B. Berry and D. Marble. Englewood Cliff: Prentice Hall.

De Jong, P. C. Sprenger and F. van Veen (1984): On Extreme Values of Moran's I and Geary's c. *Geographical Analysis*, 16, 17-24.

Godfrey, L. (1988): *Misspecification Tests in Econometrics. The Lagrange Multiplier Principle and other Approaches*. Cambridge: Cambridge University Press.

Goldfeld, S. and R. Quandt (1964): Some Tests for Homoscedasticity. *Journal of the American Statistical Association*, 60, 539-547

King, M. (1981): A Small-Sample Property of the Cliff-Ord Test for Spatial Correlation. *Journal of the Royal Statistical Society B*, 43, 263-264.

Moran, P. (1950): Notes on Continuous Stochastic Phenomena. *Biometrika*, 37, 178-181.

Sen, A. (1976): Large Sample Size Distribution of Statistics Used in Testing for Spatial Correlation. *Geographical Analysis*, 9, 175-184.

Sen, A. (1990): Distribution of Spatial Correlation Statistics, in *Proceeding from the Symposium Spatial Statistics: Past, Present and Future* (pp. 257-272). Ed. D. Griffith. Ann Arbor, Institute of Mathematical.

Szroeter, J. (1978): A Class of Parametric Tests for Heteroskedasticity in Linear Econometric Models. *Econometrica*, 46, 1311-1328.

Tiefelsdorf, M. and B. Boots (1995): The Exact Distribution of Moran's I. *Environment and Planning A*, 27, 985-999.

Yule, U. and M. Kendall (1950): An Introduction to the Theory of Statistics. London: Charles Griffin.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-25.