EUROPEAN REGIONAL SCIENCE ASSOCIATION.

44th Congress. Porto. August, 25-29, 2004

Vuong and Wald tests. Simplicity vs. Complexity‡

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 specification of cross-sectional models is usually solved following a traditional procedure, highly supported by practitioners. In the first step, a simple model is proposed that will be subsequently improved with different elements if the evidence so advises. This procedure expedites the econometric solution and fits well into the Lagrange Multiplier approach, which contributes to explain its current popularity. However, there are other methods that could also be used, and some of them are considered in this paper. Specifically, we turn our attention to the Vuong test, developed in the context of the *Kullback-Leibler* information measure. This test represents an intermediate solution between the complexity inherent in the Wald test and the simplicity of the Lagrange Multiplier principle.

---

‡ **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.

1.- Introduction

In a recent paper, Florax, Folmer and Rey (2003, FFR in advance) review the strategies of building models employed in the area known as Spatial Econometrics. According to them, at present, a classical forward stepwise approach dominates, structured in three stages:

(i)- In the first place, a simple model is estimated, static and specified under ideal conditions.

(ii)- Secondly, a series of tests of spatial dependence are applied to the equation.

(iii)- If the null is rejected in any test, the corresponding adjustments will be applied: reformulating the equation, filtering the variables, incorporating elements of spatial dynamics, ...

This method could be called *Specific to General Modelling* in terms of Charemza and Deadman (1997), and is very popular among econometricians. Implicitly, the reliability of the procedure is tied to the number and scope of the misspecification tests solved during the process, which explains the wide-ranging set of such tests habitually reported. However, this kind of *expanding-through-remedies* strategy is not the only possibility for building models in a cross-sectional setting. As FFR indicate, we could adopt exactly the opposite approach, *General to Specific Modelling*, or, in their own words (p.557), a '*Hendry-like specification strategy'*. Between these two alternatives there are a million different possibilities of dealing with the specification process. It is unlikely that all of them would finish up with an identical model, which poses the question of the method as one of the most important for us.

The scope of our paper is more limited. We would like just to pick up on some of the questions proposed in the work of FFR, especially those associated with the dynamics of the equation. However, it should be stressed that the real background of the discussion is that of model specification. In the next section we summarise a series of results well established in the literature on econometric methods about the problem of how to compare models. The third section focuses on the proposal of Vuong (1989) and the special features caused by the spatial framework of our specifications. In the fourth section we solve a small Monte Carlo experiment devoted to the test of Vuong. The last section is reserved for the conclusions and final comments.

2.-Information measures and other techniques for model selection in a spatial setting.

Almost every introductory essay to the discipline states that spatial relationships are characterised by cross-sectional dependence and heterogeneity and that these features should guide our econometric searches. This may explain why the literature on spatial econometrics methods is so biased towards problems dealing with testing procedures and with the development of powerful estimation techniques in several non-standard situations. These questions are of greater importance, although we must recognise that there remain other areas in which the discussion has been brief. In particular, those related to the Methodology merit some further attention.

Obviously, Spatial Econometrics is not a world apart from mainstream Econometrics so that the methodological discussion registered in the field of Economics since Robbins (1935) is also of relevance here. There are details, arising from the nature of the data (be they spatial or temporal), that result in peculiarities that are necessary to acknowledge. For example, the objective of a cross-sectional model will rarely be that of prediction, as it commonly is in a temporal context. Furthermore, the structure of the relationships that dominate in the first case is multidirectional and accompanied by multiple external effects (spillover, contagion, etc.). In these circumstances it is very difficult to apply the statement of Judge et al. (1980): '*... many of the choice procedures* (between econometric models) *(...) recommend splitting the data into two sets; one set for model choice and estimation and the other set for assessing the empirical construct'*.

Practical experience in the field of Spatial Econometrics shows a clear preference for the use of simple measures, mainly associated with the goodness of fit of the model. Anselin (1988) warns against the indiscriminate use of the R-squared statistics in models with spatial structure. Instead, it seems preferable to employ techniques that make use of an explicit *Loss Function* (Aznar, 1989). Anselin and Griffith (1988) study the properties of Mallows' Cp under different spatial configurations and their conclusions are not very positive for this statistic. The Akaike Information Criterion, AIC, (Akaike, 1973) is also widely used in applied econometrics, including those with spatial or regional contents, in spite of being inconsistent[1]. There are other selection statistics whose behaviour is more robust such as, for example, the Schwarz (1978) Bayesian Criterion, SBC, or the Hannan-Quinn (1979) Criterion, HQC. In any case, the claim of Anselin (1988, p. 297) when he says that: '*these techniques can be related to important bodies of theory and methodology in regional science and geography. (...) and have been suggested at various points in time as the new and improved research directions to follow'*, still seems to be valid. This is shown by the work of FFR, which takes up again the question of searching for the best specification (or, at least, the more acceptable one).

The present paper insists along the same lines although our objectives are more modest. Our motivation is just to explore the potential of the procedure of model selection developed by Vuong (1989). The focus of his proposal is the *Kullback-Leibler* information measure, although with a slightly different point of view. The AIC, as well as some other selection criteria, uses this statistic to measure the distance of the model under consideration from the DGP, whereas in the case of Voung it is used to determine which model, of the two considered, is closest to the true DGP. That is, it is just a question of selecting the best model in a set of two (the main aspects of this method are described in Appendix A). It could be argued that this scenario is too narrow to be useful in applied work. However, there are some common situations in which the Voung test fits perfectly, as we will show immediately.

In the Introduction to their paper, FFR state that, in a spatial context, habitually we begin by specifying a simple model under ideal conditions

$$\left.\begin{aligned} y &= X\beta + u \\ u &\sim N(0, \sigma_u^2 I) \end{aligned}\right\} \qquad (2.1)$$

The model is estimated by LS and then checked for errors in the equation. Spatial dependence, as well as heterogeneity, is a central feature of the specification and for this reason we will use the whole set of misspecification tests, most of them of a

[1] In terms of Aznar (1989, p. 153):' *....the asymptotic probability of overfitting is non-zero, whilst the asymptotic probability of underfitting is zero. So these criteria have an asymptotic probability, which is non zero, of selecting a model larger than the true one'*.

Lagrangian type. These kind of tests are easy to solve because they need just the LS estimators, but suffer from a lack of specificity in their alternative hypotheses. It is well-known, for example, that the LMERR, whose null is independence in the series of residuals, also reacts when we forget to include a lag of the dependent variable in the right hand side of the equation. The same is true with the LMLAG which reacts to errors in the error equation. This is the reason for the Lagrange Multiplier tests presented in Anselin et al. (1996), which are robust to local misspecification errors although they have some other weaknesses (Mur and Trivez, 2003, or Mur and Lauridsen, 2004). In short, it is not an unusual possibility to be faced with a situation where *all* misspecification tests are significant at the same time. FFR indicate that in this case the efforts should be directed first towards the problem with the bigger, or more significant, statistic.

A totally different alternative consists of inverting the sequence of the discussion using, in a *Hendry-like* approach, the more general model as the starting point. This model may contain dynamic terms both in the residuals and in the main equation[2]:

$$\left.\begin{array}{c} y = \rho W y + X\beta + u \\ u = \theta W u + \varepsilon \\ \varepsilon \sim N\left[0, \sigma^2 I\right] \end{array}\right\} \qquad (2.2)$$

Obviously, the main problem with this model is solving its estimation. The ML algorithm is computer time-consuming, and the reward is the set of Wald tests (see Appendix B for some additional results). However, this framework seems more robust because, at least theoretically, the rejection or acceptance of the hypothesis checked will not be influenced by omitted factors, avoiding the problem of local misspecifications. We are aware that this last statement must be investigated and, eventually, corroborated.

Finally, there is a third possibility that occupies an intermediate position between the simplicity of Lagrange Multipliers and the complexity inherent to the Wald tests. We are referring to the Vuong approach, as it has been presented before. To begin with, it is important to stress that we are dealing with a particular decision problem: it involves choosing between a model with *substantive autocorrelation* and one with

---

[2] As well as heterogeneity, non-linearity, non-normality and other anomalies that we do not care now.

residual autocorrelation, once clear symptoms of misspecification have been obtained from the initial static relationship (Equation 2.1). The problem now is that of determining which of these two models is preferable. This is an exercise of discriminating between non-nested models, given that the static specification of (2.1) has been discarded. Both models (substantive and residual dependence) will be overlapped only if the static specification of (2.1) were maintained as a third, rival, alternative. However, this circumstance has been discarded as wrongly specified.

3.-The spatial dimension in the Vuong approach.

The proposal of Vuong is singular because it deals with only two rival models. This is clearly a restriction that limits the usefulness of the method for model selection in a more general context. However, in some specific cases, the approach may be very valuable. This is true when we are sure that a static model is wrongly specified because it has no dynamics, but we do not know exactly where to expand the model: including lags in the main equation or an autocorrelation structure in the error.

Another question is the tightness of the *iid* clause on which the results of Vuong rest explicitly. To circumvent the implication of this clause, it is sufficient to filter the variables of the model, using the eigenvectors of W. If this weighting matrix is binary and symmetric, it will admit the basic spectral decomposition:

$$W = Q\Lambda Q' \Leftrightarrow \begin{cases} \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_R \end{bmatrix} = diag(\lambda_1, \lambda_2, \dots, \lambda_R) \\ Q = \begin{bmatrix} q_{11} & q_{21} & \cdots & q_{R1} \\ q_{12} & q_{22} & \cdots & q_{R2} \\ \dots & \dots & \dots & \dots \\ q_{1R} & q_{2R} & \cdots & q_{RR} \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_R \end{bmatrix} \end{cases} \quad (3.1)$$

where $\{\lambda_r, r=1, 2, \dots, R\}$ are the eigenvalues of W and the eigenvectors are in the columns of matrix Q.

Supposing that there is substantive autocorrelation in the model, we can use the matrix Q to obtain that:

$$\left.\begin{array}{c} y=\rho W y+X\beta+u \\ u\sim N(0,\sigma_1^2 I) \end{array}\right\} \Rightarrow \left.\begin{array}{c} y^*=\rho\Lambda y^*+X^*\beta+u^* \\ u^*\sim N(0,\sigma_1^2 I) \end{array}\right\} \tag{3.2}$$

where y*=Q'y, Z*=Q'Z and u*=Q'u are the filtered series. It is important to underline that the final model of (3.2) does not have relationships of cross-sectional dependency. Furthermore, if there is residual autocorrelation in the original model, the filtering process results in:

$$\left.\begin{array}{c} y=X\beta+u \\ u=\rho W u+\varepsilon \\ \varepsilon\sim N(0,\sigma_2^2 I) \end{array}\right\} \Rightarrow \left.\begin{array}{c} y^*=X^*\beta+u^* \\ u^*=\Delta^{-1}\varepsilon^* \\ \varepsilon^*\sim N(0,\sigma_2^2 I) \end{array}\right\} \tag{3.3}$$

where $\varepsilon^*=Q'\varepsilon$. The filter leads to a heteroskedastic error term ($u^*\sim N(0,\sigma_2^2\Delta^{-2})$), which is independent in a cross-sectional setting.

The log-likelihood of model (3.2) is:

$$l(y^*\mid\varphi_1)=-\frac{R}{2}\ln 2\pi-\frac{R}{2}\ln\sigma_1^2-\sum_{r=1}^{R}\frac{\left[(1-\rho\lambda_r)y_r^*-x_r^{*'}\beta\right]^2}{2\sigma_1^2}+\sum_{r=1}^{R}\ln(1-\rho\lambda_r) \tag{3.4}$$

where $\varphi_1=\left[\beta;\rho;\sigma_1^2\right]$ and $x_r^*=\left[x_{1r}^*;x_{2r}^*;....;x_{2kr}^*\right]$. The log-likelihood of model (3.3) is:

$$l(y^*\mid\varphi_2)=-\frac{R}{2}\ln 2\pi-\frac{R}{2}\ln\sigma_2^2-\sum_{r=1}^{R}\frac{\left[(y_r^*-x_r^{*'}\beta)(1-\theta\lambda_r)\right]^2}{2\sigma_2^2}+\sum_{r=1}^{R}\ln(1-\theta\lambda_r) \tag{3.5}$$

with $\varphi_2=\left[\beta;\rho;\sigma_2^2\right]$. Combining these results, the likelihood ratio (expression A.3 in Appendix A) can be expressed as:

$$LR_R(\tilde\varphi_1;\tilde\varphi_2)=l(y^*\mid\tilde\varphi_1)-l(y^*\mid\tilde\varphi_2)=-\frac{R}{2}\ln\frac{\tilde\sigma_1^2}{\tilde\sigma_2^2}+\sum_{r=1}^{R}\ln\frac{1-\tilde\rho\lambda_r}{1-\theta\lambda_r} \tag{3.6}$$

whose convergence limit may be established in general terms:

$$\frac{LR_R(\tilde\varphi_1;\tilde\varphi_2)}{R}\xrightarrow{p}E^0\left[l(y^*\mid\varphi_1^*)\right]-E^0\left[l(y^*\mid\varphi_2^*)\right]=E^0\left[\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right] \tag{3.7}$$

That is a well-known result in Statistics that states that the sample average (the likelihood ratio) of a series of independent random variables is a consistent estimator of the first order moment. Furthermore, the log likelihood functions of (3.4) and of (3.5), evaluated in the parameters of the DGP (φ_1 y φ_2) or in the *pseudo-true convergence values* (φ_1^* and φ_2^*), admit a Central Limit Theorem (Davidson, 2000), which extends to the likelihood ratio itself:

$$\sqrt{R}\left[\frac{LR_R\left(\varphi_1^*;\varphi_2^*\right)}{R} - E^0\left(\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right)\right]\overset{D}{\rightarrow} N\left(0;\varpi^{*2}\right) \qquad (3.8)$$

with $\varpi^{*2} = V^0\left(\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right)$. The last result guarantees the applicability, in this context, of the second part of Theorem 3.3 of Voung (1989, p. 313): *'(ii) if* $f(-;\theta^*) \neq g(-;\gamma^*)$ *(...) then*

$$n^{-1/2}LR_n\left(\hat{\theta}_n;\hat{\gamma}_n\right) - n^{-1/2}E^0\left(\lg\frac{f(Y_t\mid Z_t;\theta^*)}{g(Y_t\mid Z_t;\gamma^*)}\right)\overset{D}{\rightarrow} N\left(0;\varpi^{*2}\right) \qquad (3.9)$$

To sum up, two rival models have been identified which are non-nested. The null and alternative hypotheses can be expressed as:

$$\left.\begin{array}{l} H_0 : E^0\left[\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right] = 0 \\ H_A : \begin{cases} H_1 : E^0\left[\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right] > 0 \\ H_2 : E^0\left[\lg\frac{L(y^*\mid\varphi_1^*)}{L(y^*\mid\varphi_2^*)}\right] < 0 \end{cases} \end{array}\right\} \qquad (3.10)$$

The test statistic is that of (3.7), which we will use knowing that:

(i) under $H_0 : R^{-1/2}LR_R\left[\tilde{\varphi}_1;\tilde{\varphi}_2\right]/\tilde{\omega}_R \overset{D}{\rightarrow} N(0;1)$

(ii) under $H_1 : R^{-1/2}LR_R\left[\tilde{\varphi}_1;\tilde{\varphi}_2\right]/\tilde{\omega}_R \overset{as}{\rightarrow} +\infty$

(iii) under $H_2 : R^{-1/2} LR_R\left[\tilde{\varphi}_1;\tilde{\varphi}_2\right]/\tilde{\omega}_R \overset{as}{\rightarrow} -\infty$

with:

$$\begin{aligned}&\tilde{\omega}_R^2 = \frac{1}{R}\sum_{r=1}^{R}\left(\lg\frac{L_r\left(y_r^*|\tilde{\varphi}_1\right)}{L_r\left(y_r^*|\tilde{\varphi}_2\right)}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\lg\frac{L_r\left(y_r^*|\tilde{\varphi}_1\right)}{L_r\left(y_r^*|\tilde{\varphi}_2\right)}\right)^2 \Rightarrow\\&\Rightarrow \tilde{\omega}_R^2 = \frac{1}{R}\sum_{r=1}^{R}\left(\lg\frac{L_r\left(y_r^*|\tilde{\varphi}_1\right)}{L_r\left(y_r^*|\tilde{\varphi}_2\right)}\right)^2 - \left(LR_R(\tilde{\varphi}_1,\tilde{\varphi}_2)\right)^2\end{aligned} \quad (3.11)$$

4.-The performance of the Vuong approach: A simulation experiment

In the previous sections, we have defended the usefulness of the Vuong test to help us to resolve the specification of a cross-sectional static econometric model with clear symptoms of misspecification. The results seem interesting, although they are obtained in an asymptotic context and under the iid clause. To avoid the rigidity of the latter, we have proposed to filter the data using the eigenvectors of the weighting matrix W. The filter neutralises the relationships of transversal dependence although it will maintain a heteroskedastic structure, whose consequences on the performance of the test are difficult to determine in advance. In these circumstances we consider it fully justified to resolve a Monte Carlo exercise to examine how Vuong's test really works in circumstances similar to those of a problem of real modelling. We now describe the fundamental outlines of the exercise that has been carried out.

We have taken as a point of reference a simple lineal model such as:

$$y_r = \alpha + \beta x_r + u_r; \quad r = 1,2,...,R \quad (4.1)$$

In our framework, Vuong's test presupposes the presence of spatial effects, whether they are due to the presence of dynamic terms in the main equation of the model or to the existence of an autocorrelation structure in the error equation. The first case, expressed in matrix notation, corresponds to

$$\left.\begin{aligned}&y = \rho W y + X\varphi + u\\&u \sim N(0;\sigma_u^2 I)\end{aligned}\right\} \quad (4.2)$$

where X is a matrix of order (Rx2) and φ the column vector of parameters $\varphi=[\alpha,\beta]'$, associated with that matrix. The error term is assumed to be a white noise. The equations that develop in the second case are the habitual ones:

$$\left.\begin{array}{l} y = X\varphi + u \\ u = \rho W u + \varepsilon \\ \varepsilon \sim N(0;\sigma_\varepsilon^2 I) \end{array}\right\} \qquad (4.3)$$

The idea of the experiment consists of generating samples using the equations of (4.2) and of (4.3), and employing different configurations of values in the parameters. Now we will see what happens with these data when they are adjusted to a static model like that of (4.1). Concretely, for each estimation we will examine six misspecification tests (as are described in Appendix C): two of residual autocorrelation (LM-ERR and LM-EL), two of substantive autocorrelation (LM-LAG and LM-LE), as well as the joint test SARMA and the inevitable Moran test, already used in many cases as a generic specification test.

Regardless of the situation described by this set of tests, the Vuong test has been resolved (it would really only have been necessary when the above tests produced a confusing picture with respect to the type of error that exists in the specification). This test requires the ML estimation of the model with substantive dependency of (4.2) and of the model with residual dependence of (4.3). After resolving these estimations, the corresponding series have been filtered using the eigenvectors of matrix W, as a previous step to obtaining the Vuong statistic of (3.6) as well as the variance of (3.11). The final test is that of (3.9).

Other elements that intervene in the simulation are the following:

- Three sample sizes have been employed with R=25, R=100 and R=120.
- In the first two cases we have used *regular grids* of 5x5 and 10x10, respectively. The contacts have been obtained in rook movements giving rise to a binary and symmetrical weight matrix, W^b. This matrix has finally been multiplied by a scalar, $W=k\,W^b$, to assure a wide range of variation (between –1 and 1 in most cases) for the parameter of spatial dependence.

- The matrix of size 120 comes from the NUTS II European regional system for 15 member countries. Firstly, a matrix W^b with typical elements was obtained: $\{w_{rs}^b = d_{rs}^{-1}; r,s = 1,2,...,R; r \neq s\}$, where d_{rs} is the distance between the centroids of regions r and s. Then this matrix was standardised, as in the previous case, to obtain the final weighting matrix W.

- The data of the white noise and of the regressor come from a normal distribution, with mean zero and variance one.

- Three pairs of values in the parameters α and β have been simulated. The first, with values α=10 and β=0.5, assures an R^2 statistic of around 0.2, given that there are no spatial effects in the regression. The values of the second are α=10 and β=2 and the R^2 statistic is about 0.8. In the third case, the scale has again been fixed at 10 while the slope has been raised to 5, for an expected R^2 of approximately 0.95.

- The parameter of spatial dependence, ρ, oscillates between –1 and 1, in most cases, with increments of 0.1.

- Each configuration has been repeated 1000 times.

The most relevant results of this exercise are presented in Figures 4.1 to 4.6. In the first two we have samples of 25 observations, of 100 in the next two and of 120 in the last two. The weighting matrixes that have intervened in the elaboration of Figures 4.5 and 4.6 have been based on the geographical distance, while those used in Figures 4.1 to 4.4 have been constructed under the criteria of contiguity and rook movements in a regular grid. Furthermore, in Figures 4.1, 4.3 and 4.5 we have used a DGP with residual autocorrelation, while in the others a DGP with a substantive autocorrelation structure.

(Figures 4.1 to 4.6, Pages 15 to 20)

In the graphs are shown, horizontally, the value of the spatial dependence coefficient (substantive or residual) while, vertically, are represented the percentage of rejections of the null hypothesis achieved with the test indicated in each case. In all the

Figures, the series of dots without connecting line indicate the average value obtained for the R^2 statistic.

The graphs on the left of each figure, under the title *Risk of Confusion*, try to describe the potential risk of finding ourselves in a situation of lacking of a clear decision because the misspecification tests detect problems in the model but cannot agree on their origins. The series of the continuous dotted line, called *Traditional*, shows the number of draws in which the traditional tests of spatial dependence (the I of Moran, LM-LAG, LM-ERR and SARMA) reject *simultaneously* their respective null hypothesis. The continuous line, with the heading *Robust LM(1,1)*, indicates the number of times the two robust Multipliers, LM-LE and LM-EL, *simultaneously* reject their corresponding null hypothesis. The series in the continuous line with slashes with the heading *Robust LM(0,0)*, describes the cases in which both robust tests accept, *simultaneously*, their corresponding null hypothesis. Lastly, the series in the dashed line reflects the number of times in which *Vuong*'s test is not capable of discriminating between the two processes of spatial dependence contemplated in our analysis

The graphs on the right of each Figure try to describe the ability of the two available instruments to select the most adequate model of dependence for the data simulated. The continuous line with slashes represents the series called *Vuong (1,0)*, which corresponds to the number of times in which the Vuong test selects the model of substantive dependence. The series called *Vuong (0,1)*, represented in a continuous line, indicates the number of times in which this test selects a model of residual dependence. With respect to the robust tests, the representations are similar. The series called *Robust (1,0)*, in dashed line with slashes, reflects the number of draws in which the LM-LE test has been significant but the LM-EL test has not; that is, the model with substantive dependence has been selected. Finally, the series called *Robust (0,1)* describes the number of times in which the configuration of these tests advises us to select the model with residual dependence.

The comments on the results obtained can be grouped around two fundamental questions:

- How important is the risk of confusion?
- How trustworthy are the available discrimination instruments?

On the first aspect it can be said that the risk of facing a battery of non-conclusive specification tests is higher when the simulated process incorporates a structure of substantive dependence. In Figures 4.2, 4.4 and 4.6 it can be observed that the probability of all the tests being significant increases rapidly with the value of the parameter of spatial dependence. At the extremes of the parametric space considered, the probability of the four traditional statistics being significant is practically one.

The robust tests present strong anomalies that seem to be related to the sample size. For example, in Figure 4.2 with 25 observations, the series *Robust (1,1)* hardly surpasses the threshold of 10% in the whole sample space. However, when a sample size of 100 observations is used in Figure 4.4, unexpected '*bubbles'* appear in intermediate zones of the sample space. This anomaly is more noticeable for positive values of the coefficient of spatial dependence[3], and it is also present in the range of negative values. This situation occurs only when we have simulated a process with substantive dependence because, in the case of residual dependence of Figures 4.1, 4.3 and 4.5, all the series show a regular shape. As said before, this strange behaviour of the robust Multipliers had already been detected in other circumstances (Mur and Trivez, 2003, or Mur and Lauridsen, 2004).

It is clear that the discriminatory capacity of these instruments, Vuong's test and the robust tests, is not sufficiently consistent. Vuong's test seems to work quite well when the process that has intervened includes only residual-type dependence. Its sensitivity is very appreciable so that, for samples of size 100 or higher, it obtains excellent results even for reduced levels of spatial dependence. In any case, it improves the power of the robust Multipliers by between 10 and 30 points.

The problems arise when the structure of dependence is substantive. Vuong's test tends to produce better results than the robust tests although diverse incidences must be noted. In the first place, when the size of the sample is small and the parameter of spatial dependence takes positive values, the relationship is produced in the opposite sense: the robust Multipliers work better than Vuong's test, as can be seen in Figure 4.2. Secondly, the size of the sample accentuates the discriminatory capacity of both

---

[3] Between 0.4 and 0.5 and with a low R^2, the percentage of inconclusive cases is nearly 90%.

instruments although it also provokes certain distortions. The '*bubbles'* of the robust tests to which we referred with reference to the *Risk of Confusion* are again produced in this case, but in the opposite sense: as unexpected cuts in the discriminatory behaviour of the Multipliers. What is happening now is that the LM-EL test tends to be significant, erroneously, in this specific zone of the parametric space damaging the correct working of its complementary test, the LM-LE.

Lastly, a positive aspect that should be underlined with respect to the performance of these instruments is that the probability of them causing us to take an erroneous decision is very reduced. In all the cases, both Vuong's test and the robust Lagrange Multipliers, oscillate between the indefinition and the correct identification of the process. The estimated probability of taking an erroneous decision (selecting, for example, a model with substantive dependence when effectively a model with residual dependence has intervened) is very reduced. When the parameter of spatial dependence takes low values, close to zero, the incorrect decisions are close to 50% in the case of Vuong and between 5% and 10% with the robust tests. However, as the parameter takes higher values, this probability becomes negligible.

5.- Conclusions

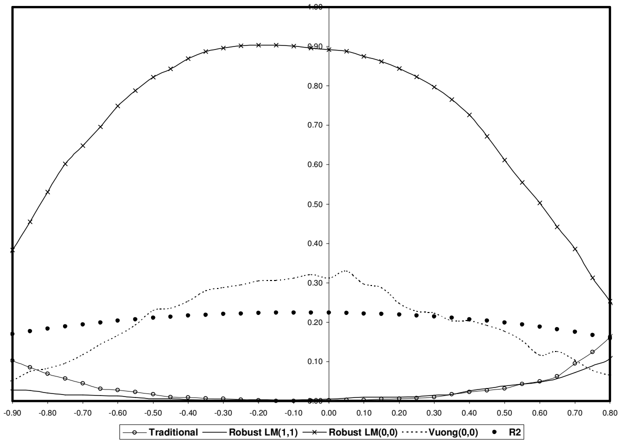
This paper has been undertaken as an exercise of reflection on a problem we consider of importance, the selection of the most adequate model for our data. Excepting the seminal work of Anselin (1988) and the more recent of Florax, Folmer and Rey (2003), there are not many specific references in the specialised literature in the field of Spatial Econometrics.

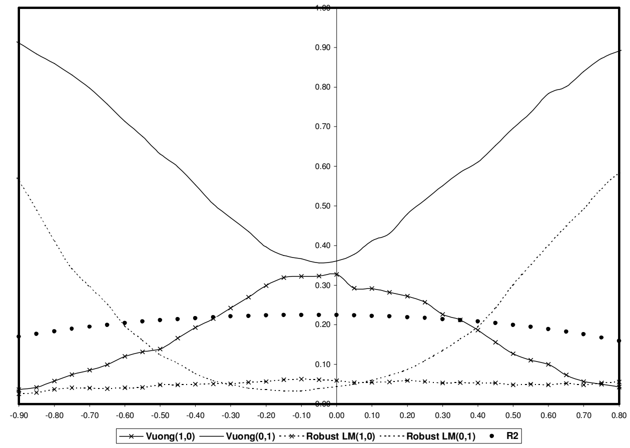
The content of the present paper has focused on the usefulness of Vuong's test. The requirements of this test are quite peculiar, given that only two rival alternatives are contemplated and it requires the iid clause. However, the proposal adapts well to a quite common situation in applied research, namely that of facing ourselves with a battery of misspecification tests that show symptoms of error in the model, but are unclear with respect to their motive. If, as habitually occurs, the analyst only contemplates a model with residual dependence or with substantive dependence as alternatives, the ideal circumstances for using Vuong's test have arisen.

The evidence obtained from the simulation exercise carried out in this paper points to the necessity of going deeper into the question. Vuong's test works reasonably well when the process that has intervened in the DGP includes a structure of residual dependence. In this case, it greatly improves the results of the robust Multipliers. Nevertheless, anomalies abound when the process that has intervened includes a mechanism of substantive dependence. It is not easy to find convincing arguments to explain these results, something that will form part of a future programme of research.

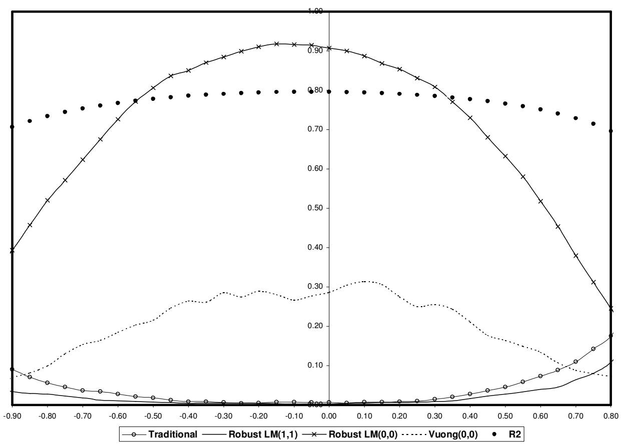
FIGURE 4.1: Residual Autocorrelation. R=25. Binary Normalised Matrix.

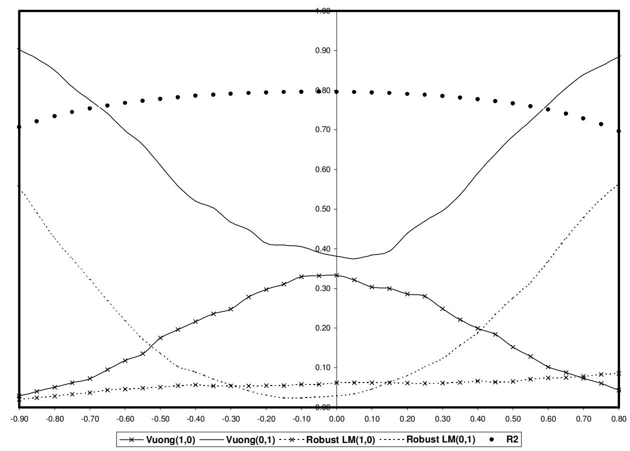
4.1a- Risk of Confusion LOW R^2 4.1b- Discriminating Performance



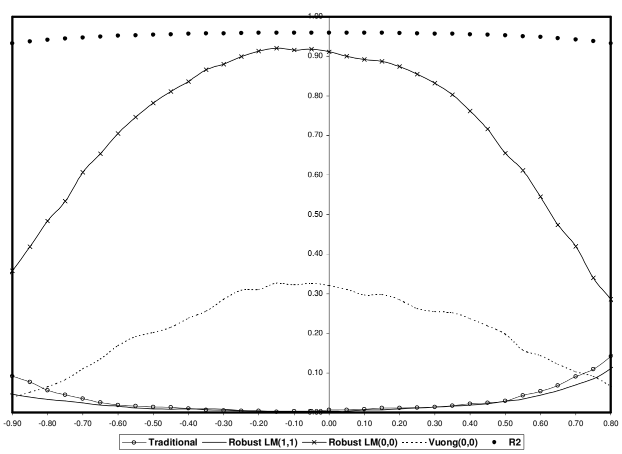


4.1c- Risk of Confusion MEDIUM R^2 4.1d- Discriminating Performance





4.1e- Risk of Confusion HIGH R^2 4.1f- Discriminating Performance



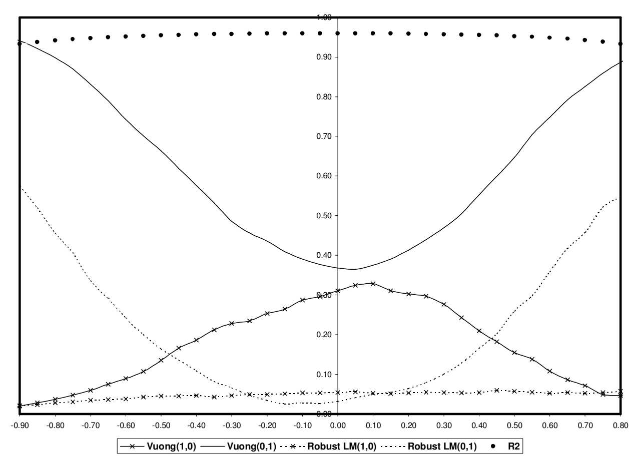
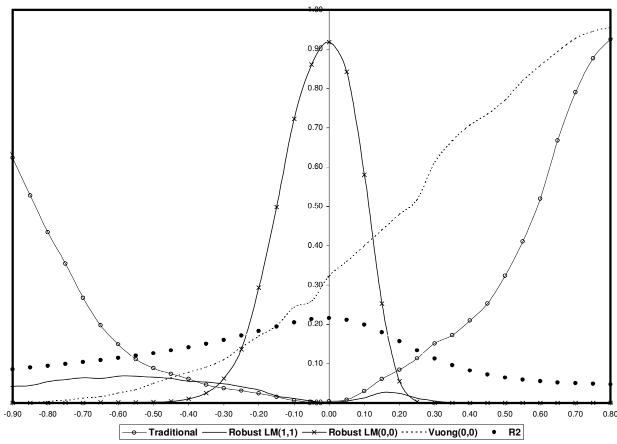
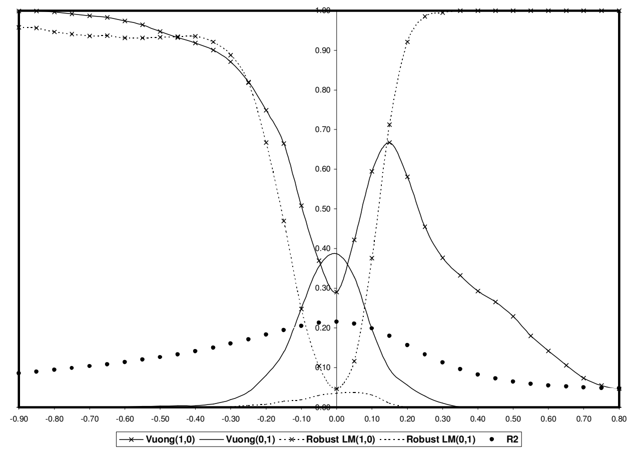


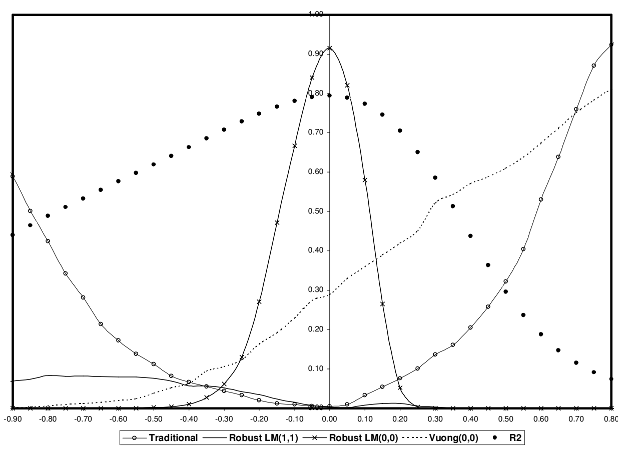
FIGURE 4.2: Substantive Autocorrelation. R=25. Binary Normalised Matrix.

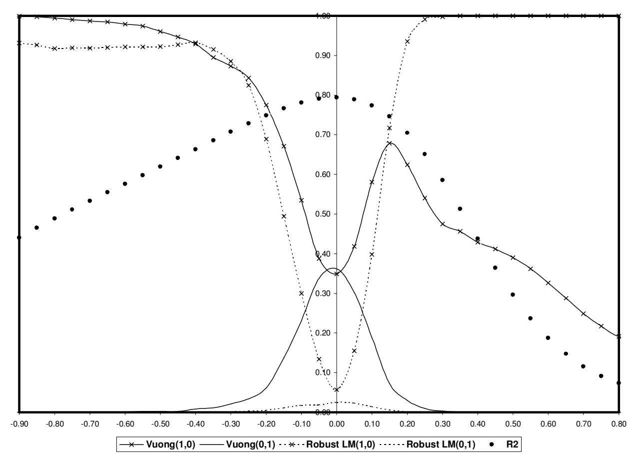
4.2a- Risk of Confusion LOW R^2 4.2b- Discriminating Performance



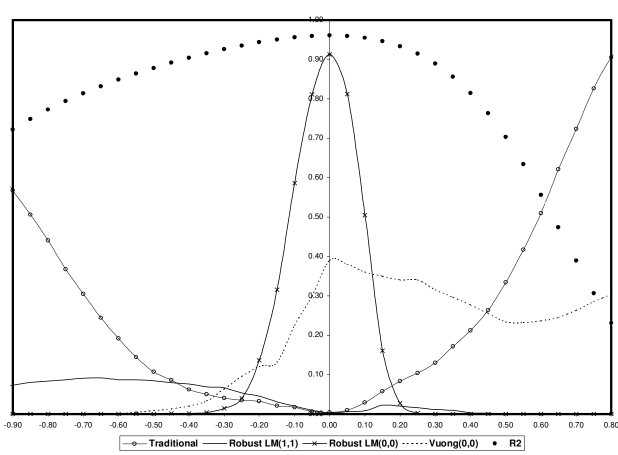


4.2c- Risk of Confusion MEDIUM R^2 4.2d- Discriminating Performance





4.2e- Risk of Confusion HIGH R^2 4.2f- Discriminating Performance



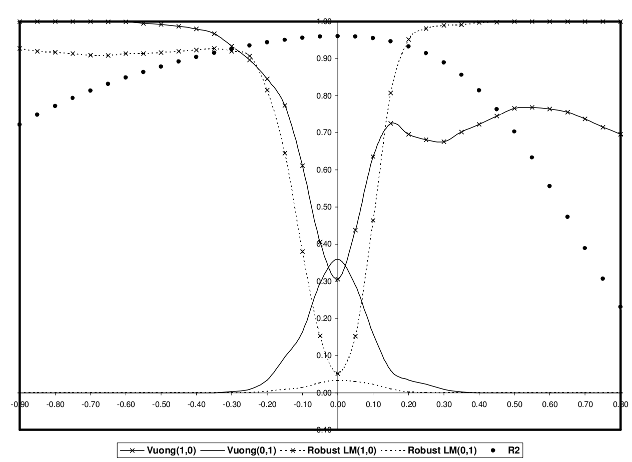
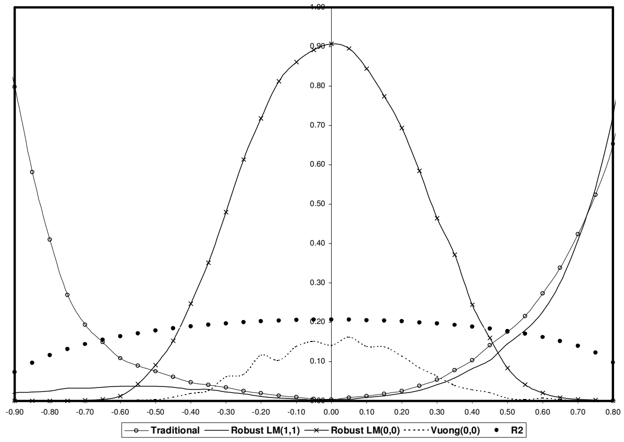
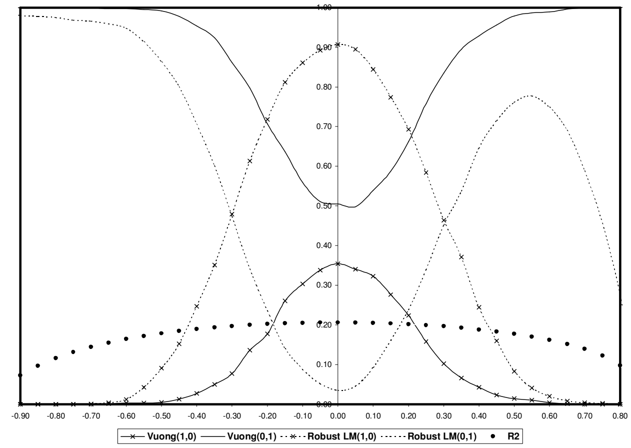


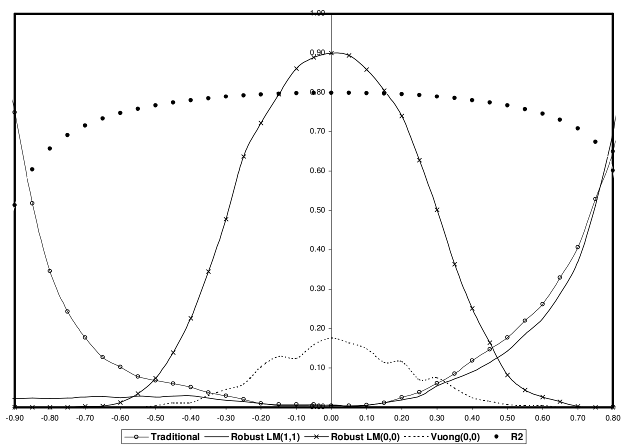
FIGURE 4.3: Residual Autocorrelation. R=100. Binary Normalised Matrix.

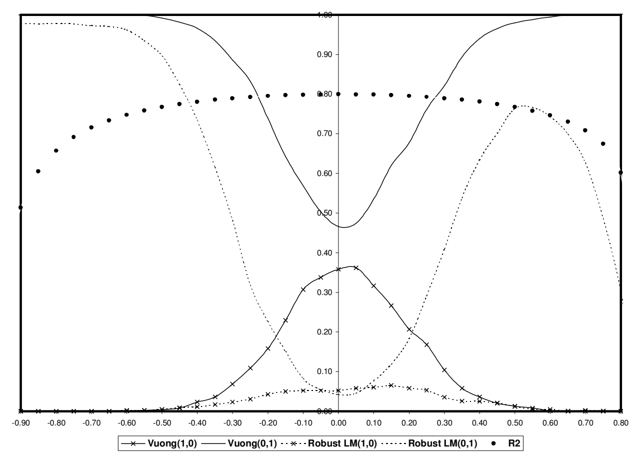
4.3a- Risk of Confusion   LOW R^2 4.3b- Discriminating Performance



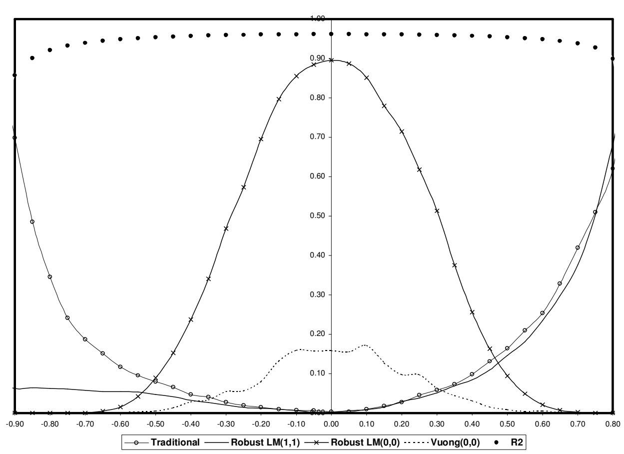


4.3c- Risk of Confusion   MEDIUM R^2 4.3d- Discriminating Performance





4.3e- Risk of Confusion   HIGH R^2 4.3f- Discriminating Performance



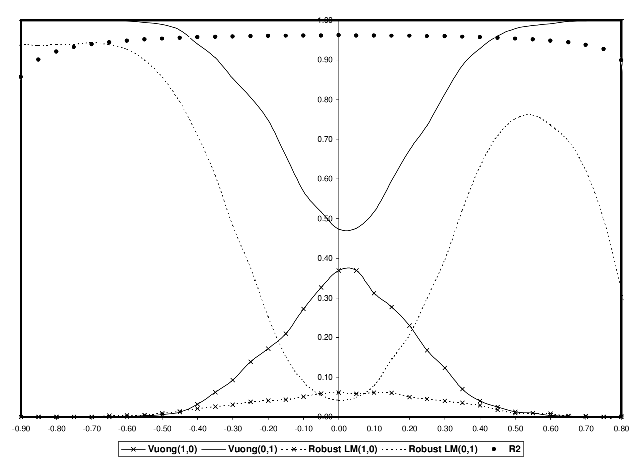
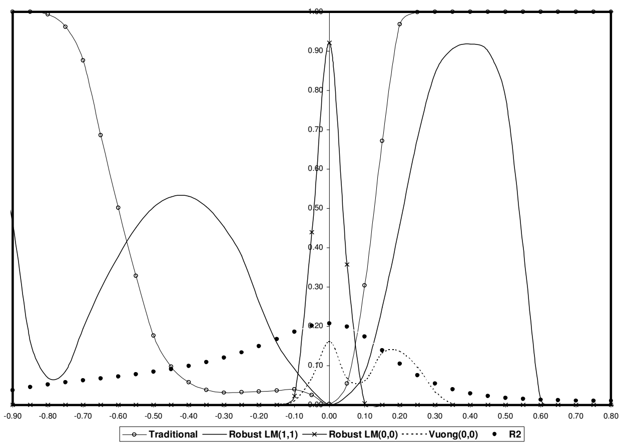
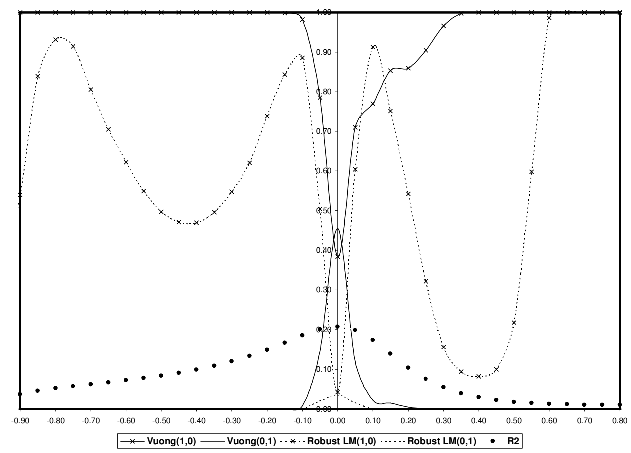


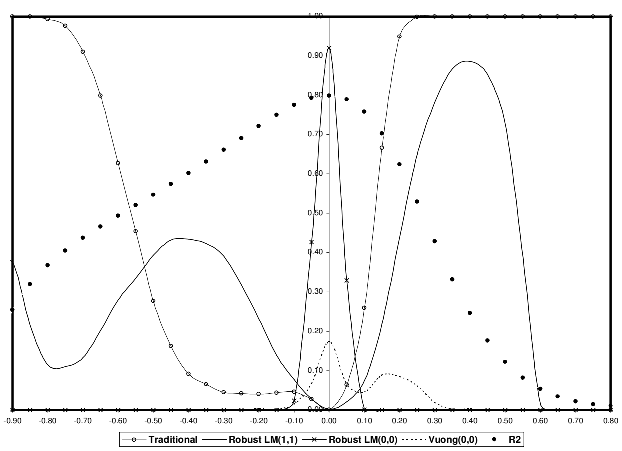
FIGURE 4.4: Substantive Autocorrelation. R=100. Binary Normalised Matrix.

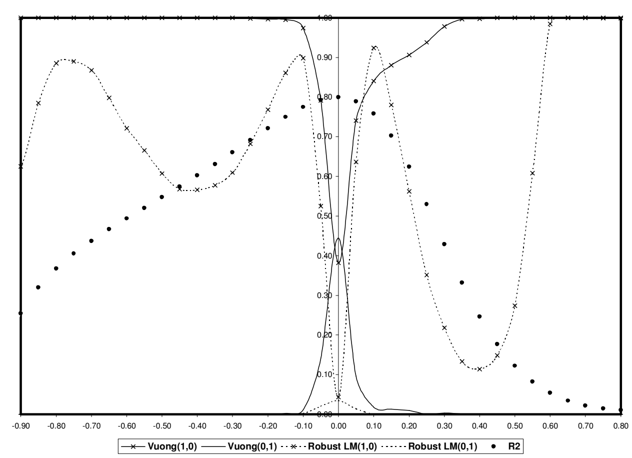
4.4a- Risk of Confusion **LOW R^2** **4.4b- Discriminating Performance**



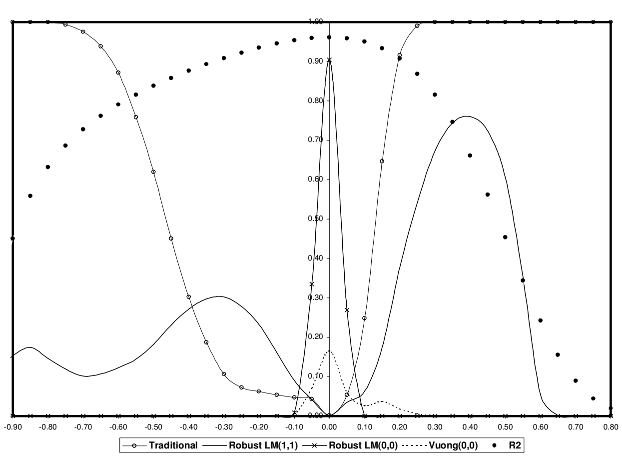


4.4c- Risk of Confusion **MEDIUM R^2** **4.4d- Discriminating Performance**





4.4e- Risk of Confusion **HIGH R^2** **4.4f- Discriminating Performance**



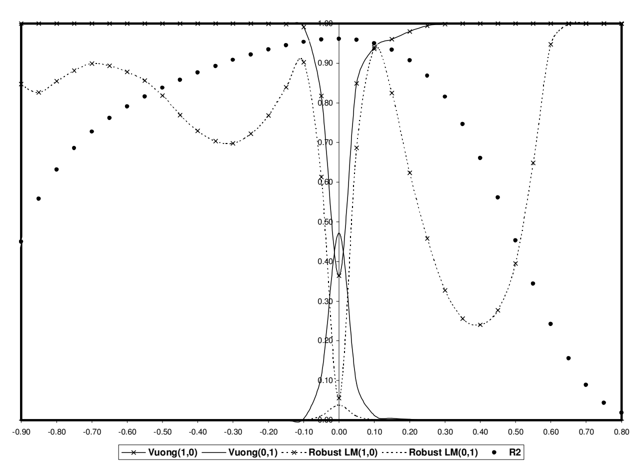
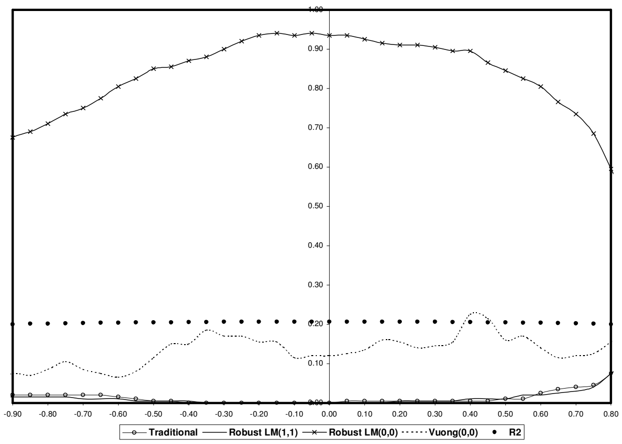
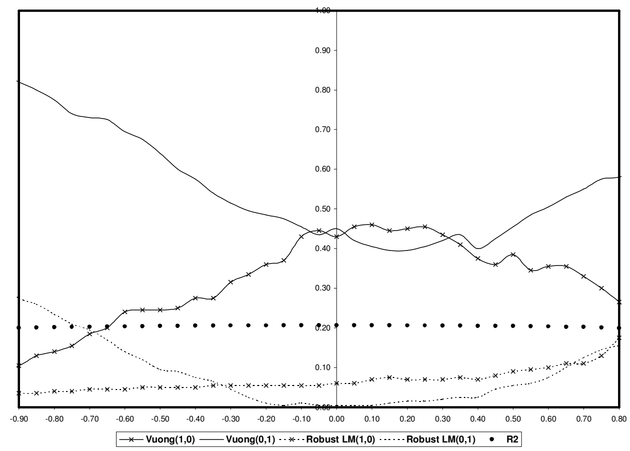


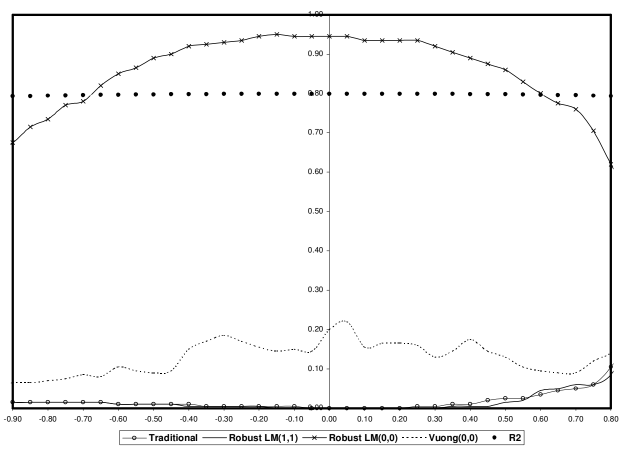
FIGURE 4.5: Residual Autocorrelation. R=120. Normalised Distance-based Matrix.

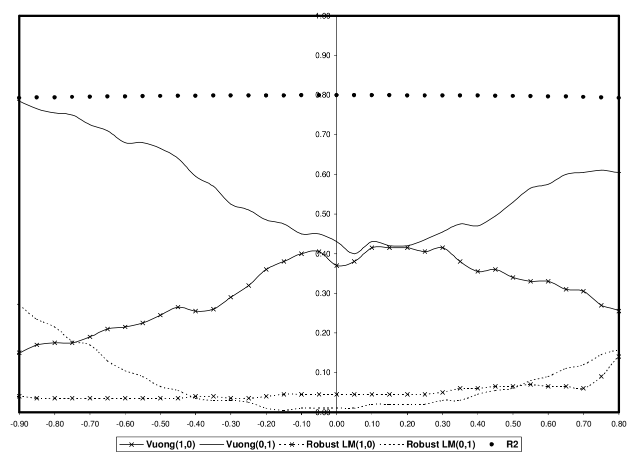
4.5a- Risk of Confusion **LOW R^2** **4.5b- Discriminating Performance**



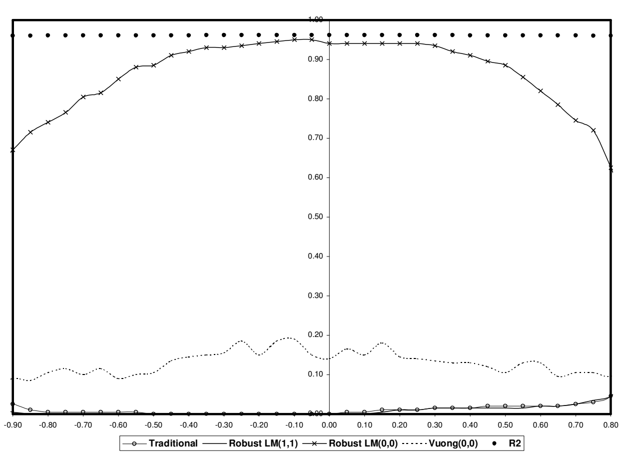


4.5c- Risk of Confusion **MEDIUM R^2** **4.5d- Discriminating Performance**





4.5e- Risk of Confusion **HIGH R^2** **4.5f- Discriminating Performance**



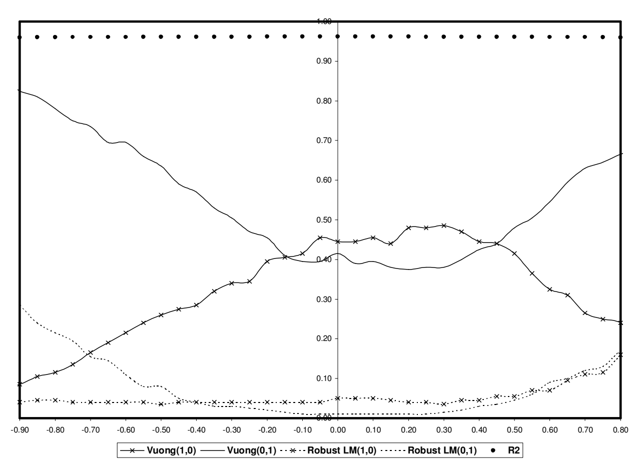
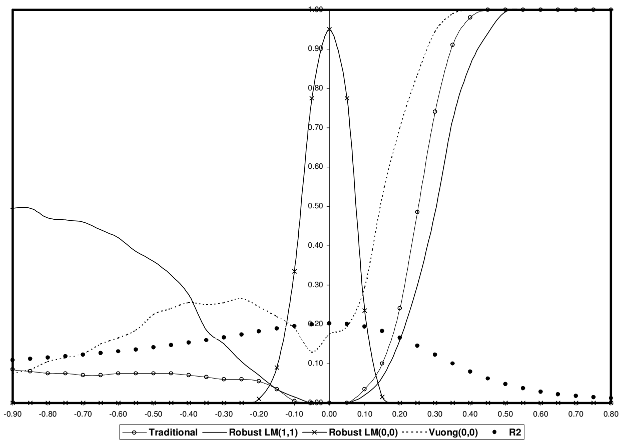
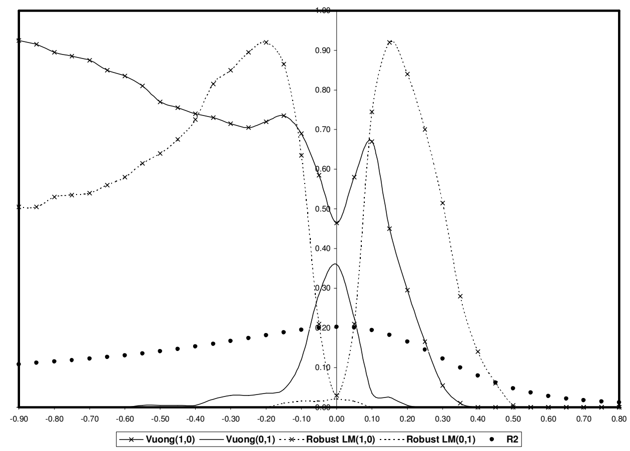


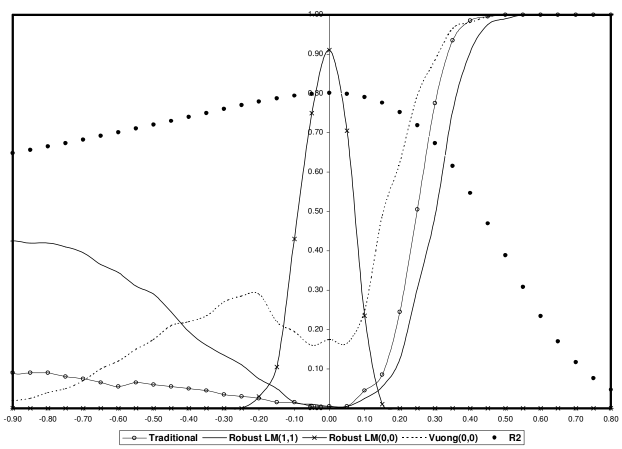
FIGURE 4.6: Substantive Autocorrelation. R=120. Normalised Distance-based Matrix.

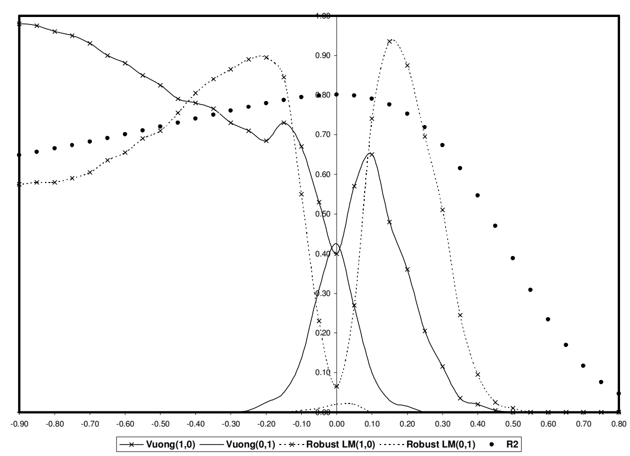
4.6a- Risk of Confusion LOW R^2 4.6b- Discriminating Performance



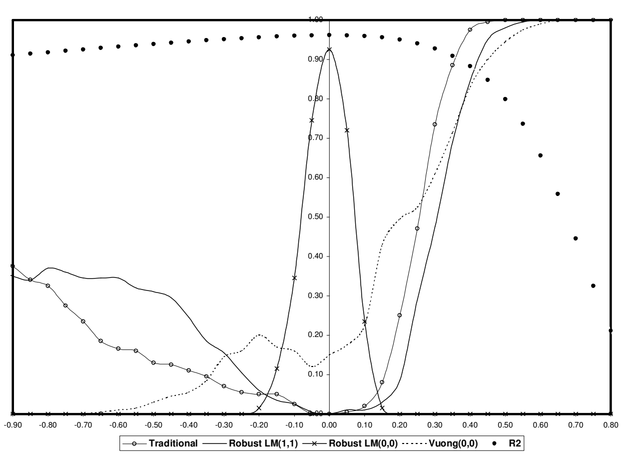


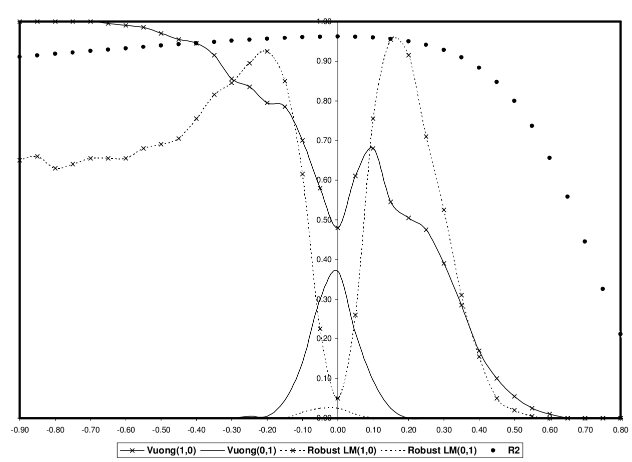
4.6c- Risk of Confusion MEDIUM R^2 4.6d- Discriminating Performance





4.6e- Risk of Confusion HIGH R^2 4.6f- Discriminating Performance





Appendix A: Basic aspects of Voung's approach

The strategy proposed by Vuong (1989) entails some complexity and it is not widely spread, so let us introduce its key elements. On one hand there is a random vector X_t of order (mx1), partitioned in X_t=[Y_t, Z_t] of orders n and k respectively (n+k=m). The true distribution function of vector X_t is H_X^0, although we are interested in the conditional of Y_t (endogenous) with respect to Z_t (explicative), $H_{Y|Z}^0$. Vectors X_t satisfy the *iid* clause and the distribution and density functions ($h_{Y|Z}^0$ and $h_{Y|Z}^0$), both conditional and unconditional, are well behaved.

The details of the true distribution function are unknown, so two possible rival families are contemplated: $F_\theta = \{F_{Y|Z}(\theta); \theta \in \Theta \subset \Re^p\}$ and $G_\gamma = \{G_{Y|Z}(\gamma); \gamma \in \Gamma \subset \Re^q\}$, which can be nested, overlapped or strictly independent. Again, the associated distribution and density functions are well behaved, so the matrices of second order moments exist:

$$\left.\begin{array}{ll} A_f(\theta) = E^0\left[\dfrac{\partial^2 \lg f(-;\theta)}{\partial\theta\partial\theta'}\right] & A_g(\gamma) = E^0\left[\dfrac{\partial^2 \lg g(-;\gamma)}{\partial\gamma\partial\gamma'}\right] \\ B_f(\theta) = E^0\left[\dfrac{\partial \lg f(-;\theta)}{\partial\theta}\dfrac{\partial \lg f(-;\theta)}{\partial\theta'}\right] & B_g(\gamma) = E^0\left[\dfrac{\partial \lg g(-;\gamma)}{\partial\gamma}\dfrac{\partial \lg g(-;\gamma)}{\partial\gamma'}\right] \\ B_{fg}(\theta) = E^0\left[\dfrac{\partial \lg f(-;\theta)}{\partial\theta}\dfrac{\partial \lg g(-;\gamma)}{\partial\gamma'}\right] & \end{array}\right\} \qquad \text{(A.1)}$$

where E^0 means '*expected value according to the true distribution function*' of X_t.

The maximum likelihood estimators of θ and γ will be consistent in the sense of White (1982). If the model is misspecified, these estimators will converge towards the *pseudo-true values* of Sawa (1978). Whatever the case, their asymptotic distribution will be standard:

$$\left.\begin{array}{ll} plim\, \tilde{\theta}_n = \theta^* & \sqrt{n}\left[\tilde{\theta}_n - \theta^*\right] \xrightarrow{D} N\left[0; A_f^{-1}(\theta^*)\, B_f(\theta^*)\, A_f^{-1}(\theta^*)\right] \\ plim\, \tilde{\gamma}_n = \gamma^* & \sqrt{n}\left[\tilde{\gamma}_n - \gamma^*\right] \xrightarrow{D} N\left[0; A_g^{-1}(\gamma^*)\, B_g(\gamma^*)\, A_g^{-1}(\gamma^*)\right] \end{array}\right\} \qquad \text{(A.2)}$$

with θ* and γ* being the convergence points, and n the sample size. This result allows us to estimate consistently the matrices of second order moments of (A.1) and the covariance matrices of (A.2). The difference between the log-likelihoods:

$$LR_n(\tilde{\theta}_n;\tilde{\gamma}_n) = L_f(\tilde{\theta}_n) - L_g(\tilde{\gamma}_n) = \sum_{t=1}^{n} \lg \frac{f(Y_t \mid Z_t;\tilde{\theta}_n)}{g(Y_t \mid Z_t;\tilde{\gamma}_n)} \qquad \text{(A.3)}$$

is a sum of n independent terms, so that, after standardising, we get results of *almost sure* convergence:

$$\frac{LR_n(\tilde{\theta}_n;\tilde{\gamma}_n)}{n} \xrightarrow{as} E^0\left[\lg \frac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right] \qquad \text{(A.4)}$$

and in distribution:

$$\sqrt{n}\left[\frac{LR_n(\tilde{\theta}_n;\tilde{\gamma}_n)}{n} - E^0\left[\lg \frac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right]\right] \xrightarrow{D} N(0;\omega^{*2}) \qquad \text{(A.5a)}$$

$$2LR_n(\tilde{\theta}_n;\tilde{\gamma}_n) \xrightarrow{D} M(p+q;\lambda^*) \qquad \text{(A.5b)}$$

The result of (A.5a) is obtained when the models are not equivalent, $f(Y_t \mid Z_t;\theta^*) \neq g(Y_t \mid Z_t;\gamma^*)$. In this expression, ω^{*2} is the variance of the quotient between the log-likelihoods, evaluated in the parameters of convergence:

$$\omega^{*2} = V^0\left[\lg \frac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right] = E^0\left[\left(\lg \frac{f(Y_t|Z_t;\theta^*)}{g(Y_t|Z_t;\gamma^*)}\right)^2\right] - \left(E^0\left[\lg \frac{f(Y_t|Z_t;\theta^*)}{g(Y_t|Z_t;\gamma^*)}\right]\right)^2 \qquad \text{(A.6)}$$

On the other hand, the result of (A.5b) corresponds to the case that both models are observationally equivalent, $f(Y_t \mid Z_t;\theta^*) = g(Y_t \mid Z_t;\gamma^*)$. The final statistic M(m, λ) is the sum of the squares of m normal variables, each of them weighted by the corresponding element of vector λ: $\mathrm{M}(\mathrm{m},\lambda) = \sum_{j=1}^{m} \lambda_j u_j^2; \quad u_j \sim \mathrm{iidN}(0,1)$. In particular, in (A.5b), λ* is the vector of (p+q) eigenvalues of matrix V:

$$V = \begin{bmatrix} B_f(\theta^*)A_f^{-1}(\theta^*) & -B_{fg}(\theta^*;\gamma^*)A_g^{-1}(\gamma^*) \\ -B_{gf}(\theta^*;\gamma^*)A_f^{-1}(\theta^*) & B_g(\gamma^*)A_g^{-1}(\gamma^*) \end{bmatrix} \qquad (A.7)$$

The second order moment of (A.6) can also be used to measure the distance between both families. If the models are equivalent, the variance will be zero, so:

$$\left.\begin{array}{l} H_0^{\varpi} : \omega^{*2} = 0 \\ H_A^{\varpi} : \omega^{*2} \neq 0 \end{array}\right\} \qquad (A.8)$$

The test statistic is the sampling variance of the series of contributions to the log-likelihoods of (A.3):

$$\tilde{\omega}_n^2 = \frac{1}{n}\Sigma_{t=1}^n\left(\lg\frac{f\left(Y_t|Z_t;\tilde{\theta}_n\right)}{g\left(Y_t|Z_t;\tilde{\gamma}_n\right)}\right)^2 - \left(\frac{1}{n}\Sigma_{t=1}^n\lg\frac{f\left(Y_t|Z_t;\tilde{\theta}_n\right)}{g\left(Y_t|Z_t;\tilde{\gamma}_n\right)}\right)^2 \qquad (A.9)$$

or in its centred version:

$$\hat{\omega}_n^2 = \tilde{\omega}_n^2 + \left(\frac{1}{n}LR_n(\tilde{\theta}_n;\tilde{\gamma}_n)\right)^2 \qquad (A.10)$$

Under the null hypothesis, both statistics have well-defined probability limits:

$$plim\,\tilde{\omega}_n^2 = \omega^{*2} \qquad (A.11a)$$

$$plim\,\hat{\omega}_n^2 = \omega^{*2} + E^0\left[\lg\frac{f\left(Y_t \mid Z_t;\theta^*\right)}{g\left(Y_t \mid Z_t;\gamma^*\right)}\right] \qquad (A.11b)$$

and their asymptotic distribution turns out to be that of (A.5b):

$$\left.\begin{array}{l} n\,\tilde{\omega}_n^2 \\ n\,\hat{\omega}_n^2 \end{array}\right\} \xrightarrow{D} M(p+q;\lambda^*) \qquad (A.12)$$

These results allow us to built a model selection strategy in which three cases are distinguished.

Non-nested models

This will happen when $F_\theta \cap G_\gamma = \varnothing$, which implies that the density functions will, necessarily, be different: $f(-;\theta^*) \neq g(-;\gamma^*)$. In this case, the likelihood ratio will lead to a simple test of the null hypothesis against a bilateral alternative:

$$\left.\begin{array}{l} H_0 : E^0\left[\lg \dfrac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right] = 0 \\ H_A : \left\{ \begin{array}{l} H_f : E^0\left[\lg \dfrac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right] > 0 \\ H_g : E^0\left[\lg \dfrac{f(Y_t \mid Z_t;\theta^*)}{g(Y_t \mid Z_t;\gamma^*)}\right] < 0 \end{array}\right. \end{array}\right\} \tag{A.13}$$

Theorem 5.1 of Vuong (1989, p. 318) establishes that: '.... *if F_θ and G_γ are strictly non-nested, then:*

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

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

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

(iv) properties (i) to (iii) hold if $\tilde{\omega}_n$ *is replaced by* $\hat{\omega}_n$ '.

Given that the two models are different, the acceptance of the null should be interpreted in the sense that the available evidence does not permit us to discriminate between both alternatives.

Overlapped models

In this case, both families will share some common distribution functions ($F_\theta \cap G_\gamma \neq \varnothing$), but they are not nested ($F_\theta \not\subset G_\gamma$ y $G_\gamma \not\subset F_\theta$). In these circumstances, both specifications may be equivalent so that this possibility should be contemplated at first. Once the equivalence has been discarded, we should proceed to discriminate between them. The test of the first stage (equivalence) is:

$$\left.\begin{array}{l} H_0 : f(-;\theta^*) = g(-;\gamma^*) \\ H_A : f(-;\theta^*) \neq g(-;\gamma^*) \end{array}\right\} \Leftrightarrow \left.\begin{array}{l} H_0^{\omega} : \omega^{*2} = 0 \\ H_A^{\omega} : \omega^{*2} \neq 0 \end{array}\right\} \tag{A.14}$$

for which the variance of (A.9) can be used. Theorem 6.1 of Vuong (1989, p. 321) indicates that:

‘*(i) under* H_0^{ω}*, for any* $x \geq 0$, $\Pr\left[n\,\tilde{\omega}_n^2 \leq x\right] - M_{p+q}(x;\tilde{\lambda}_n^2) \xrightarrow{as} 0$

(ii) under $H_A^{\omega}; n\,\tilde{\omega}_n^2 \xrightarrow{as} +\infty$

(iii) properties (i) and (ii) hold for $n\,\hat{\omega}_n^2$ ‘.

If the null hypothesis were rejected, we can conclude that both models are not observationally equivalent. Then, the LR test could be used to discriminate between them as in (A.13).

Nested models

In this case, one of the families is a particular case of the other ($G_\gamma \subset F_\theta$). When the ample model is well specified, the strategy based on the LR test of (A.3) leads to the usual case with a standard χ^2 distribution.

Appendix B: The Wald test in a general specification

The log-likelihood of model (2.2) is as usual:

$$\begin{array}{c} l(y/\varphi) = -\frac{R}{2}\ln 2\pi - \frac{R}{2}\ln\sigma^2 - \frac{[By - X\beta]' D' D[By - X\beta]}{2\sigma^2} + \ln|B| + \ln|D| \\ \ln|B| = |I - \rho W| = \sum_{r=1}^{R}(1 - \rho\lambda_r) \\ \ln|D| = |I - \theta W| = \sum_{r=1}^{R}(1 - \theta\lambda_r) \end{array} \tag{B.1}$$

where λ_r is the r-th eigenvalue of the weighting matrix W and φ the vector $\varphi = [\delta,\rho,\beta,\sigma^2]'$. The score is non-linear in parameters and its solution requires numerical methods:

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}\left[X'D'D(By - X\beta)\right] = \frac{1}{\sigma^2} X'D'\varepsilon \tag{B.2a}$$

$$\frac{\partial l}{\partial \rho} = \frac{1}{\sigma^2}\left[y'WD'D(By - X\beta)\right] - \sum_{r=1}^{R} \frac{\lambda_r}{1-\rho\lambda_r} = \frac{1}{\sigma^2}\left[y'WD'\varepsilon\right] - \sum_{r=1}^{R} \frac{\lambda_r}{1-\rho\lambda_r} \tag{B.2b}$$

$$\frac{\partial l}{\partial \theta} = \frac{1}{\sigma^2}\left[(By - X\beta)'WD(By - X\beta)\right] - \sum_{r=1}^{R} \frac{\lambda_r}{1-\theta\lambda_r} = \frac{1}{\sigma^2}\left[u'W\varepsilon\right] - \sum_{r=1}^{R} \frac{\lambda_r}{1-\theta\lambda_r} \tag{B.2c}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{R}{2\sigma^2} + \frac{\left[(By - X\beta)'D'D(By - X\beta)\right]}{2\sigma^4} = -\frac{R}{2\sigma^2} + \frac{\varepsilon'\varepsilon}{2\sigma^4} \tag{B.2d}$$

The ML estimation of β and σ^2 could be obtained conditionally on those of the parameters of spatial dependence, ρ and θ:

$$\tilde{\beta} = \left[X'\tilde{D}'\tilde{D}X\right]^{-1} X'\tilde{D}'\tilde{D}\tilde{B}y \tag{B.3a}$$

$$\tilde{\sigma}^2 = \frac{\left[\tilde{B}y - X\tilde{\beta}\right]'\tilde{D}'\tilde{D}\left[\tilde{B}y - X\tilde{\beta}\right]}{R} = \frac{\varepsilon'\varepsilon}{R} \tag{B.3b}$$

with $\tilde{D} = I - \tilde{\theta}W$ and $\tilde{B} = I - \tilde{\rho}W$. Then the procedure of Ord (1975) could be used.

The information matrix is of a general type:

$$I(\varphi) = -E\left[\frac{\partial^2 l}{\partial\varphi\partial\varphi'}\right] = \begin{bmatrix} \frac{X'D'DX}{\sigma^2} & \frac{X'D'DWB^{-1}X\beta}{\sigma^2} & 0 & 0 \\ \frac{\beta'X'B^{-1}WD'DX}{\sigma^2} & t(\rho) & 2\mathrm{tr}\,W^2B^{-1}D^{-1} & \frac{\mathrm{tr}WB^{-1}}{\sigma^2} \\ 0 & 2\mathrm{tr}\,W^2B^{-1}D^{-1} & t(\theta) & \frac{\mathrm{tr}WD^{-1}}{\sigma^2} \\ 0 & \frac{\mathrm{tr}WB^{-1}}{\sigma^2} & \frac{\mathrm{tr}WD^{-1}}{\sigma^2} & \frac{R}{2\sigma^4} \end{bmatrix} \tag{B.4}$$

where:

$$\left.\begin{array}{l}(\rho) = \dfrac{\beta' X' B^{-1} WD'DW\, B^{-1} X\beta}{\sigma^2} + \text{tr}\, D^{-2} B^{-1} WD'DW\, B^{-1} + \sum_{r=1}^{R}\left(\dfrac{\lambda_r}{1-\rho\lambda_r}\right)^2 \\ t(\theta) = \text{tr}\, D^{-1} W^2 D^{-1} + \sum_{r=1}^{R}\left(\dfrac{\lambda_r}{1-\theta\lambda_r}\right)^2\end{array}\right\} \quad (B.5)$$

The inverse of this matrix is the asymptotic covariance matrix of ML estimators. We will employ the following notation:

$$I(\varphi)^{-1} = \begin{bmatrix} V(\beta) & Cov(\delta,\alpha) \\ Cov(\alpha,\delta) & V(\alpha) \end{bmatrix} \quad (B.6)$$

being $\alpha = [\rho,\theta,\sigma^2]'$. The main results could be expressed as follows (the details are in Mur and Angulo, 2003):

$$V(\beta) = \sigma^2 [X'D'DX]^{-1} \left[I_{2k} + V(\rho) X'D'DW\, B^{-1} X\beta\beta' X' B^{-1} WD'DX \left(X'D'DX\right)^{-1}\right] \quad (B.7)$$

$$Cov(\delta,\alpha) = \begin{bmatrix} -V(\rho)\left(X'D'DX\right)^{-1} X'D'DW\, B^{-1} X\beta \\ -Cov(\rho,\theta)\left(X'D'DX\right)^{-1} X'D'DW\, B^{-1} X\beta \\ -Cov(\rho,\sigma^2)\left(X'D'DX\right)^{-1} X'D'DW\, B^{-1} X\beta \end{bmatrix} \quad (B.8)$$

$$V(\alpha) = \begin{bmatrix} V(\rho) & Cov(\rho,\theta) & Cov(\rho,\sigma^2) \\ Cov(\theta,\rho) & V(\theta) & Cov(\theta,\sigma^2) \\ Cov(\sigma^2,\rho) & Cov(\sigma^2,\theta) & V(\sigma^2) \end{bmatrix} \quad (B.9)$$

with:

$$V(\rho) = \left[\frac{e}{\sigma^2} + 2R\, \text{var}(m_\rho)(1 - corr^2(m_\rho, m_\theta))\right]^{-1} \quad (B.10a)$$

$$V(\theta) = \frac{1}{\text{var}(m_\theta)}\left[\frac{1}{2R} + V(\rho)\, \text{var}(m_\rho)\, corr^2(m_\rho, m_\theta)\right] \quad (B.10b)$$

$$V(\sigma^2) = \frac{2\sigma^4}{R}\left[1 + \mu^2(m_\theta)\right] + 4\sigma^4 V(\rho)\, \text{var}(m_\rho)\left[\mu(m_\rho) - \mu(m_\theta)\, corr^2(m_\rho, m_\theta)\right]^2 \quad (B.10c)$$

$$Cov(\rho,\theta) = -V(\rho)\left(\frac{\mathrm{var}(m_\rho)}{\mathrm{var}(m_\theta)}\right)^2 corr(m_\rho, m_\theta) \tag{B.10d}$$

$$Cov(\rho,\sigma^2) = -2\sigma^2 V(\rho)\,\mathrm{var}(m_\rho)\left[\mu(m_\rho) - \mu(m_\theta)corr(m_\rho, m_\theta)\right] \tag{B.10e}$$

$$Cov(\theta,\sigma^2) = \frac{\sigma^2}{\sqrt{\mathrm{var}(m_\theta)}}\left\{2V(\rho)\,\mathrm{var}(m_\rho)corr(m_\rho, m_\theta)\left[\mu(m_\rho) - \mu(m_\theta)corr(m_\rho, m_\theta)\right] - \frac{\mu(m_\theta)}{R}\right\} \tag{B.10f}$$

being $e = \beta' X' B^{-1} WD' M_{XD} DW B^{-1} X\beta$ and $M_{XD} = I - DX\left(X'D'DX\right)^{-1} X'D'$. In the above expressions, we have introduced the following auxiliary variables in order to facilitate these rather awkward results:

$$m_r(\rho) = \frac{\lambda_r}{1-\rho\lambda_r} \qquad \overline{m}(\rho) = \sum_{r=1}^{R} \frac{m_r(\rho)}{R} \qquad \mathrm{var}(m_\rho) = \frac{\sum_{r=1}^{R}\left[m_r(\rho) - \overline{m}(\rho)\right]^2}{R} \tag{B.11a}$$

$$m_r(\theta) = \frac{\lambda_r}{1-\theta\lambda_r} \qquad \overline{m}(\theta) = \sum_{r=1}^{R} \frac{m_r(\theta)}{R} \qquad \mathrm{var}(m_\theta) = \frac{\sum_{r=1}^{R}\left[m_r(\theta) - \overline{m}(\theta)\right]^2}{R} \tag{B.11b}$$

$$\mu(m_\rho) = \frac{\overline{m}(\rho)}{\sqrt{\mathrm{var}(m_\rho)}} \qquad \mu(m_\theta) = \frac{\overline{m}(\theta)}{\sqrt{\mathrm{var}(m_\theta)}} \tag{B.11c}$$

$$corr(m_\rho, m_\theta) = \frac{\sum_{r=1}^{R}\left[m_r(\rho) - \overline{m}(\rho)\right]\left[m_r(\theta) - \overline{m}(\theta)\right]}{\sqrt{\sum_{r=1}^{R}\left[m_r(\rho) - \overline{m}(\rho)\right]^2 \sum_{r=1}^{R}\left[m_r(\theta) - \overline{m}(\theta)\right]^2}} \tag{B.11d}$$

Using these results, it is straightforward to obtain the Wald tests for the simple null hypothesis relative to each one of the parameters of spatial dependence:

$$\left.\begin{array}{l} \left.\begin{array}{l} H_0: \rho = 0 \\ H_A: \rho \neq 0 \end{array}\right\} \Rightarrow W(\tilde{\rho}) = \dfrac{\tilde{\rho}^2}{V(\tilde{\rho})} \underset{as}{\sim} \chi^2(1) \\ \left.\begin{array}{l} H_0: \theta = 0 \\ H_A: \theta \neq 0 \end{array}\right\} \Rightarrow W(\tilde{\theta}) = \dfrac{\tilde{\theta}^2}{V(\tilde{\theta})} \underset{as}{\sim} \chi^2(1) \end{array}\right\} \tag{B.12}$$

At the same time, it may be of interest to obtain also a global test of spatial effects in the model, which implies a composite null:

$$\left.\begin{aligned} H_0 &: \theta = \rho = 0 \\ H_A &: \theta \neq 0 \vee \rho \neq 0 \end{aligned}\right\} \qquad \text{(B.13)}$$

$$\Rightarrow W(\tilde{\rho};\tilde{\theta}) = \frac{1}{1-r^2_{(\tilde{\rho};\tilde{\theta})}}\left[W(\tilde{\rho}) + W(\tilde{\theta}) - 2r^2_{(\tilde{\rho};\tilde{\theta})}\sqrt{W(\tilde{\rho})W(\tilde{\theta})}\right] \underset{as}{\sim} \chi^2(2)$$

where $r_{(\tilde{\rho};\tilde{\theta})}$ stands for the correlation coefficient between the ML estimator of ρ and θ[4]. If these estimators were orthogonal ($r_{(\tilde{\rho};\tilde{\theta})} = 0$), the join statistic of (B.13) would turn into the sum of the two single statistics of (B.12). On the other hand, when the parameters of spatial dependence take on similar values, the correlation between their corresponding ML estimators will tend to the unit ($r_{(\tilde{\rho};\tilde{\theta})} \to 1$). In this case, the statistic of (B.13) moves towards the difference between the unidirectional statistics of (B.12):

$$\underset{\rho\to\theta}{W(\tilde{\rho};\tilde{\theta})} \to \frac{1}{1-r^2_{(\tilde{\rho};\tilde{\theta})}}\left[t(\tilde{\rho}) - t(\tilde{\theta})\right]^2 \qquad \text{(B.14)}$$

with $t(\tilde{\rho}) = \sqrt{W(\tilde{\rho})}$ and $t(\tilde{\theta}) = \sqrt{W(\tilde{\theta})}$.

Appendix C: Misspecification tests employed.

The tests described here always refer to a static model, such as: y = Xβ + u. This model has been estimated by LS, where $\hat{\sigma}^2$ and $\hat{\beta}$ correspond to the LS estimations and $\hat{u}$ to the residual series. These tests are the following (see Florax and de Graaff, 2004, for the details):

Moran Test: $$I = \frac{R}{S_0}\frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}; \qquad S_0 = \sum_{r=1}^{R}\sum_{s=1}^{R} w_{rs} \qquad \text{(C.1)}$$

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

[4] That is $r_{(\tilde{\rho};\tilde{\theta})} = \dfrac{Cov(\tilde{\rho};\tilde{\theta})}{\sqrt{V(\tilde{\rho})V(\tilde{\theta})}}$.

LM-EL Test: $$\text{LM}-\text{EL}=\frac{\left(\dfrac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}-\dfrac{T_1}{R\hat{J}_{\rho\text{-}\beta}}\dfrac{\hat{u}'Wy}{\hat{\sigma}^2}\right)^2}{T_1-\dfrac{T_1^2}{R\hat{J}_{\rho\text{-}\beta}}} \quad (C.3)$$

LM-LAG Test: $$\text{LM}-\text{LAG}=\frac{1}{R\hat{J}_{\rho-\beta}}\left(\frac{\hat{u}'Wy}{\hat{\sigma}^2}\right)^2 \quad (C.5)$$

LM-LE Test: $$\text{LM}-\text{LE}=\frac{\left(\dfrac{\hat{u}'Wy}{\hat{\sigma}^2}-\dfrac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2}{R\hat{J}_{\rho\text{-}\beta}-T_1} \quad (C.6)$$

SARMA Test: $$\text{SARMA}=\frac{\left(\dfrac{\hat{u}'Wy}{\hat{\sigma}^2}-\dfrac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2}{R\hat{J}_{\rho\text{-}\beta}-T_1}+\frac{1}{T_1}\left(\frac{\hat{u}'W\hat{u}}{\hat{\sigma}^2}\right)^2 \quad (C.7)$$

Moreover, $R\hat{J}_{\rho-\beta}=T_1+(\hat{\beta}'X'WMWX\hat{\beta})/\hat{\sigma}^2$ and $M=[I-X(X'X)^{-1}X']$. As is well-known, the asymptotic distribution of the standardised Moran's I is an N(0,1); the four Lagrange Multipliers that follow, LM-ERR, LM-EL, LM-LAG and LM-LE have an asymptotic $\chi^2(1)$. The final SARMA statistic has a chi-square distribution with two degrees of freedom under the null of no spatial effects.

References

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In Petrow, B. and F. Csaki (eds.): *2nd International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiodo.

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

Anselin, L. and D. Griffith, 1988. Do Spatial Effects Really Matter in Regression Analysis?. *Papers of the Regional Science Association,* 65, 4-34.

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

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

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

Davidson, D. 2000. *Econometric Theory*. Oxford: Blackwell Publishers.

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.

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*. Berlin: Springer.

Hannan, E. and B. Quinn, 1979. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society, Series B,* 41, 190-195.

Judge, G, W. Griffiths, R. Hill and T. Lee, 1980. *The Theory and Practice of Econometrics*. Nueva York: John Wiley & Sons.

Mur, J. and A. Angulo, 2003. El test de Vuong y otros indicadores de dinámica espacial (I). Paper presented at the *XXIX Reunión de Estudios Regionales,* Santander.

Mur, J. and J. Lauridsen, 2004. Dirty Data in Cross-Sectional Regressions. Paper presented at the *3rd Spatial Econometric Workshop,* Strasburg.

Mur, J. and J. Trivez, 2003. Unit Roots and Deterministic Trends in Spatial Econometric Models. *International Regional Science Review.*

Ord, J., 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association,* 70, 120-126.

Robbins, L., 1935. *An Essay on the Nature and Significance of Economic Science*. London: McMillan.

Sawa, T., 1978. Information Criteria for Discriminating Among Alternative Regression Models, *Econometrica*, 46, 1273-1291.

Schwartz, C., 1978. Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461-464.

Vuong, Q., 1989. Likelihood Ratio-Tests for Model Selection and Non-Nested Hypotheses. *Econometrica,* 57, 307-333.

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