ABSTRACT

Testing for spatial group-wise heteroskedasticity in spatial autocorrelation regression models: Lagrange Multiplier scan tests

Authors and e-mails: Le Gallo, J. (julie.le-gallo@agrosupdijon.fr)[1a], López, F. (fernando.lopez@upct.es)[2b] and Chasco, C. (coro.chasco@uam.es)[3c]

Department: [1]CESAER, [2]Department of Quantitative Methods and Informatics, [3]Department of Applied Economics

University: [a] CESAER, Agrosup Dijon,
[b] Technical University of Cartagena
[c] Autonomous University of Madrid

Special Session: S10 - Procesos espaciales a nivel de punto y redes espaciales

Subject area: Métodos de análisis espacial, sistemas de información geográfica y big data

Abstract:

The aim of this paper is to develop a spatial group-wise heteroscedasticity (SGWH) test based on the scan methodology, specifically developed for spatial autocorrelation regression models (SAR and SEM): the "Scan-LM test". It is a scan test based on the Lagrange Multiplier (LM) principle instead of the usual log-likelihood ratio (LR). The main advantage of the LM approach, with respect to LR, lies in the comparative ease of implementation, since it is not necessary to obtain the maximum likelihood estimations for the alternative hypothesis. When rejecting the null hypothesis, this test identifies the shape and size of the spatial clusters with different residual variance, which is very useful for specification search of the regression model. Another important benefit of the Scan-LM test is that it does not need to specify a spatial weights matrix. An extensive Monte Carlo simulation confirms the existence of high degree of similarity between the Scan-LM test sizes and their respective nominal values. This test is also

robust in presence of no-normality and other forms of a-spatial heteroscedasticity. We finally propose an application on housing prices in the agglomeration of Madrid.

Keywords: Spatial Group-Wise Heteroskedasticity, Lagrange Multiplier test, Spatial Autocorrelation.

JEL codes: C21, C52, C63, R15

1. Introduction

In spatial econometric models, the behavior of tests based on the Lagrange multiplier (LM) principle to test for spatial autocorrelation in regression models has been largely studied since the original proposals by Burridge (1980) and Anselin (1988). In small samples, their behavior may be affected by deviations from the conditions under which they have been developed, such as deviation from normality or misspecification of the spatial weights matrix. This subject has lately received a lot of attention and several modifications of the LM tests have been proposed to guarantee their robustness in the presence of deviations from ideal conditions (Born and Breitung, 2011; Baltagi and Yang, 2013; Robinson and Rossi, 2014). Two aspects characterize this literature.

On the one hand, most LM tests for testing spatial autocorrelation are developed for a null hypothesis of independence of residuals of the classical regression models. Only a few proposals have been made for spatial regression models that already include substantive spatial autocorrelation as in the Spatial Autoregressive Model (SAR) or the Spatial Error Model (SEM). We label these tests marginal tests in order to differentiate them from the tests that apply to models without spatial autocorrelation. For instance, Anselin (1988) and Anselin et al. (1996) develop conditional LM tests for the presence of spatial error autocorrelation in a SAR model and the presence of spatial lag in a SEM model. Anselin (1988) also considers a more general model including heteroscedasticity. To the best of our knowledge, non Monte-Carlo exercise has been carried out to evaluate the performance of these in small samples, although, according to Guo et al. (2015), these tests presence problems both in size and power, but no evidence is provided for this claim.

On the other hand, the effects of heteroskedasticity on the asymptotic version of LM tests of spatial dependence are well-known. For example, Kelejian and Robinson (1998) propose a joint test for spatial autocorrelation and heteroskedasticity, where the variance is an unknown function of explanatory variables (conditional heteroskedasticity). However, in almost all cases, heteroskedasticity is analyzed when it does not have a specific functional form.

In this paper, we combine both these issues and consider the problem of heteroscedasticity in the context of marginal LM tests. More specifically, we focus on one particular form of heteroscedasticity, namely Spatial Group-Wise Heteroscedasticity (SGWH). Articles dealing with this issue are scarce. An exception is the paper by Anselin and Rey (1991) who conclude that 'heteroscedasticity [SGWH] invalidates the chi-square distribution for both LM tests, though slightly less for LM lag' (p. 124). Along the same line, Kelejian and Robinson (2004) warn about the pernicious problems that can be caused by SGWH on the specification of the model: "its effect (spatially autocorrelation heteroscedasticity) on MI and LM depends upon whether that correlation is positive or negative. If it is positive, our results imply that a researcher is more likely to conclude that the error terms are spatially correlated, when they are not; the reverse is true if it is negative' (p. 80). This statement leads them to propose a modification of Moran's I test that is robust to SGWH. In this sense, the inferential framework is extremely important, since, as shown by Born and Breitung (2011), the bootstrap version of these tests present problems. SGWH has also received attention in applied works focusing on ways to achieve the best model specification of the spatial regression model. When the presence of such a structure in the variance of the residuals of the regression models with spatial effects is suspected, observations with spatially differenciated behaviours are selected in an ad hoc way. For example Dall'erba and Le Gallo (2008) and Fischer and Stirböck (2004) assume the presence of SGWH in models that incorporate spatial effects. The identification of convergence clubs is done by using Getis and Ord's local spatial autocorrelation statistics.

All this literature addressing the question of heteroskedasticity, whether in groupwise form or not, always consider that the null hypothesis is the independence of the regression residuals. To the best of our knowledge, no paper considers the case of the null of spatial dependence. Therefore, this paper presents some results in this context, based on the idea that in most processes for cross-sectional data, it is very probable that both effects, spatial autocorrelation and spatially groupwise heteroskedasticity, are present. In order to consider this issue, we follow López et al. (2015) and Chasco et al. (2018) and propose tests for SGWH for residuals in SAR/SEM models based on scan methodology (Kulldorff et al., 2009; Kulldorf and Nagarwalla, 1995). Scan tests are useful in spatial econometric specification search as one of their by-product is the detection of spatial clusters of observations with similar, high or low, variance of residuals.

In this context, the aim of this paper is twofold. First, in order to fix ideas, the problems generated by SGWH in SAR/SEM models are analyzed using a small Monte-Carlo simulation in section 2. Second, in section 3, specific tests are developed to detect SGWH based on the scan methodology. The developed tests also give information on the size and form of SGWH without the need to provide information on spatial structure of the observations. This last question is important because the information provided by the test can be of help to perform a relevant specification search of the process. The behavior of the tests proposed is analysed with a Monte-Carlo simulation in section 4. Finally, in section 5, we illustrate the use of this methodology with an empirical application on housing prices in Madrid. Main conclusions appear in section 6.

2. Impact of spatial group-wise heteroskedasticity on marginal tests of spatial econometrics

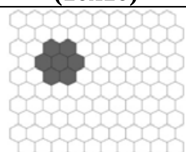
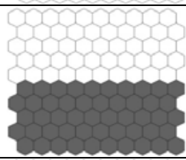
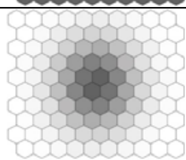
The marginal LM tests of spatial dependent assess for the presence of SAR or SEM structures in the error term. In this section, we perform a small Monte Carlo simulation to analyze the impact of SGWH in the residuals of SAR and SEM models.

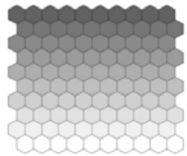
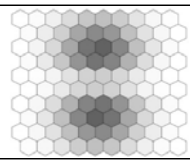
The design of the Monte-Carlo simulation is as follows:

1. The data generating process (DGP) are of SEM and SAR types with SGWH. Formally:

$$SEM: \quad y = X\beta + u \;\; u = \lambda W u + \varepsilon \;\; \varepsilon \sim N(0,\Omega)$$
$$SAR: \quad y = \rho W y + X\beta + \varepsilon \;\; \varepsilon \sim N(0,\Omega)$$

where $\Omega = \text{diag}(w_{ii})$ is a diagonal matrix where each element of the main diagonal w_{ii} depends on the location of observation i following patterns of SGWH determined by the function $w_{ii} = h(x,y)$, where x and y are the centroids of observation i:

DGP	Pattern of SGWH	Regular lattice (10x10)
SGWH1	$h(x,y)$=0.9 if (x,y) is inside the circle* $h(x,y)$=0.1 if (x,y) is outside the circle The circle has the same size, 11 cells, for all sample sizes	
SGWH2	$h(x,y)$=0.7 if (x,y) is inside the circle * $h(x,y)$=0.3 if (x,y) is outside the circle The circle has the same size, 11 cells, for all sample sizes	
SGWH3	$h(x,y)=\begin{cases} 0.1 & if \quad y < Me_y \\ 0.9 & if \quad y \leq Me_y \end{cases}$	
SGWH4	$h(x,y)=\begin{cases} 0.3 & if \quad y < Me_y \\ 0.7 & if \quad y \leq Me_y \end{cases}$	
SGWH5	$h(x,y) = e^{-0.05\,[(x-Me_x)^2+(y-Me_y)^2]}$	

DGP	Pattern of SGWH	Regular lattice (10x10)
SGWH6	$h(x,y) = y/\max(y)$	
SGWH7	$h(x,y) = 0.5\lvert y - Me_y \rvert e^{-0.05\,[(x-Me_x)^2+(y-Me_y)^2]}$	

Each observation is identified by the two coordinates of its centroid (*x*,*y*). Me_x is the median of all coordinates *x*. Me_y is the median of all coordinates *y*.

2. We consider regular hexagonal lattices of sample sizes *R* = 49, 100, 225, 400 and 900.
3. The weight matrix *W* is based on the rook contiguity criterion.
4. The levels of spatial autocorrelation are set to $\rho = 0.5$ and $\lambda = 0.5$

Table 1a shows the behavior of both tests.

Table 1a: Sensibility of the two marginal LM tests to the presence of SGWH: LM^*_{ERR} (conditional test of spatial error autocorrelation in a SAR model) and LM^*_{LAG} (conditional test of a spatial lag in a SEM model)

DGP: $y = X\beta + u \;\; u = \lambda Wu + \varepsilon \;\; \varepsilon \sim N(0,\Omega)$

% of times where LM^*_{LAG} in SEM rejects *H0*

R	H0	SGWH1	SGWH2	SGWH3	SGWH4	SGWH5	SGWH6	SGWH7
49	0.052	0.014	0.046	0.038	0.052	0.055	0.061	0.067
100	0.040	0.049	0.054	0.040	0.046	0.036	0.031	0.046
225	0.040	0.052	0.034	0.046	0.048	0.034	0.077	0.032
400	0.039	0.047	0.045	0.049	0.048	0.052	0.065	0.048
900	0.036	0.052	0.045	0.061	0.053	0.047	0.068	0.064

DGP: $y = \rho Wy + X\beta + \varepsilon \;\; \varepsilon \sim N(0,\Omega)$

% of times where LM^*_{ERR} in SAR rejects H0

R	H0	SGWH1	SGWH2	SGWH3	SGWH4	SGWH5	SGWH6	SGWH7
49	0.048	0.060	0.050	0.095	0.063	0.068	0.085	0.072
100	0.051	0.105	0.050	0.100	0.062	0.056	0.082	0.064
225	0.057	0.111	0.045	0.082	0.078	0.116	0.100	0.089
400	0.044	0.099	0.059	0.105	0.055	0.181	0.069	0.157
900	0.049	0.094	0.062	0.098	0.066	0.429	0.082	0.362

H0→ homoscedastic models

The main results obtained in this small simulation exercise show that:

(i) The presence of SGWH does not affect the size of the test LM^{*}_{LAG} in SEM models. Both structures (spatial autocorrelation and SGWH) are mixed in the model residuals and the marginal test does not identify the presence of substantive spatial autocorrelation.

(ii) The presence of SGWH in SAR models induces an oversize of the LM^{*}_{ERR} test. In this case, the SGWH structure introduced in the residuals is a confounding effect for the marginal test identifying the presence of spatial autocorrelation when there is actually a SGWH structure.

Based on these results, it is necessary, at least in the case of SAR-type specifications, to identify the presence of SGWH before testing if there is a SARAR structure using the marginal tests. When identifying a SGWH structure, both marginal tests can be adapted to the presence of heteroscedasticity (**Anselin 1988**).

3. Scan-LM test

In this section, we develop the SGWH tests for SAR and SEM models. We specify, as a general model, a typical regression model with spatial effects with SGWH:

$$\mathrm{Ay} = \mathrm{X}\beta + \mathrm{u};\quad \mathrm{Bu} = \varepsilon;\ \ \varepsilon \sim \mathrm{N}(0,\Omega) \qquad (1)$$

Where A=I-ρW; B=I-λW with W the spatial weights matrix and $\Omega = diag(\omega_1,...,\omega_R)$.

The alternative hypothesis under which this test is developed is defined by two levels of residual variability. The Ω elements are defined as follows:

$$\omega_i = \begin{cases} \sigma^2 & si \quad i \notin Z \\ h_\sigma \sigma^2 & si \quad i \in Z \end{cases} \qquad (2)$$

where Z is a subset of connected regions (spatial clusters) and $h_\sigma > 0$ is a parameter which allows the introduction of different levels of SGWH.

The objective of this paper is testing the following hypothesis:

$$\begin{aligned} &H_0 : h_\sigma = 1 \quad \forall Z \in \Theta \\ &H_A : \exists Z \in \Theta \ \ where \ \ h_\sigma \neq 1 \ \ (or \ \ h_\sigma > 1;\ or\ h_\sigma < 1;\ or\ h_\sigma > 0) \end{aligned} \qquad (3)$$

being Θ the set of all possible connected regions which could be considered in the study area. Typically in practice, this set Θ is reduced to only circular and/or elliptic regions, though it is also possible to work with spatial clusters of flexible shapes (**Tango, 2005**)

3.1 The LR approach of the Scan test: Computational problems

Scan methods are based on the log-likelihood ratio approach. That is, authors compute the following statistic:

$$LR^{SGWH} = \sup_{Z \in \Theta} \left\{ l(\theta)\big|_{H_A} - l(\theta)\big|_{H_0} \right\} \tag{4}$$

for l(θ) the log-likelihood function, which depends of the parameters θ=($\beta,\rho,\lambda,\sigma,h$) and Z a set of connected regions.

The log-likelihood function has the following general form:

$$l(\theta) = \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \ln|A| + \ln|B| - \frac{1}{2}(Ay - X\beta)' \, \Omega^{-1} (Ay - X\beta) \tag{5}$$

In order to compute the supremum value, it is necessary to maximize this function under H_A as many times as the number of the considered windows, #(Θ). For example, in the case of a sample size N, considering only circular windows, Z, with a maximum number of observations $\#(Z)<[N/2]$, it would be necessary to calculate the log-likelihood function $[N^2/2]$ times (with [] denoting the integer part).

The statistical significance of this statistic must be obtained by a permutational resampling method. That is, it could only be obtained by the maximization of the log-likelihood under the alternative hypothesis as many times as necessary bootstrap samples (typically, $B = 999$).

Since the log-likelihood is a highly non-linear function, the maximization process must be done using some of the known algorithm (e.g. Newton-Raphson). Therefore, this is a high-performance computation procedure, since it must compute $[B*N^2/2]$ times the H_A, what makes all this process quite impractical, particularly for Monte Carlo simulations.

3.2 The LM approach of the SGWH Scan test

The scan tests can also be evaluated by an LM procedure. Hence, it will not be necessary to consider any alternative hypothesis or to set specific assumptions about the h_σ parameter. Consequently, this procedure greatly facilitates calculation of the statistic p-value.

The LM version of the scan test is developed in the same way as the LR one. Firstly, the LM test is computed for a determined *Z* window (circles, ellipses, flexible shapes…). Secondly, we select the highest value of all the LM tests. Thirdly, we evaluate the statistical significance of this value by a permutational resampling procedure.

Although the LM test does not require the model estimation under the alternative hypothesis, getting a more concrete expression of this test will allow a more efficient computation. Here we present the LM Scan test for a spatial autoregressive (SAR) model and spatial error model (SEM), which are the most well-known spatial models.

1) SGWH LM Scan test for a SAR model

The SAR model can be specified as follows:

$$\begin{aligned} & y = \rho WY + X\beta + \varepsilon \\ & \varepsilon \sim N(0, \Omega) \end{aligned} \tag{6}$$

The general expression of the LM test for these model residuals is the following[1]:

$$LM_{SAR}^{SGWH} = \left[g(\theta)|_{H_0} \right]' \left[I(\theta)|_{H_0} \right]^{-1} \left[g(\theta)|_{H_0} \right] \sim \chi_1^2 \tag{7}$$

where $g(\theta)$ is the log-likelihood gradient and $I(\theta)$ is the information matrix. In order to be computationally efficient with this test, it is essential to find a more concrete expression that allows detecting among terms depending (or not depending) from the window *Z*. After some calculations, the final expression for the SGWH LM Scan test for the SAR model is the following:

$$LM - Scan_{SAR}^{SGWH} = \sup_{Z \in \Theta} \frac{g(h_\sigma)^2 a^{44}}{\det(I(\theta)|_{H_0})} \tag{8}$$

where the determinant of the information matrix results in a sum of terms depending (or not) of the *Z* windows:

$$\begin{aligned} \det(I(\theta)|_{H_0}) = \left(\frac{1}{\sigma^2}\right)^{k+3} & \left[\frac{\sigma^2 R_Z}{2} \det(A_{33}) - \frac{R_Z}{2} \left(\frac{R_Z}{2} \det(A_{22}) - trWA^{-1} \det(X'X) tr\, H_{h_\sigma} WA^{-1} \right) \cdot \right. \\ & \left. + tr\, H_{h_\sigma} WA^{-1} \det(X'X) \left(\frac{R_Z}{2} trWA^{-1} - \frac{R}{2\sigma^2} tr\, H_{h_\sigma} WA^{-1} \right) \right] \end{aligned} \tag{9}$$

---

[1] The complete presentation of the test is shown in the Annexes I.1 and I.2.

This test can be modified as a function of a determined alternative hypothesis. For example, if we want to test the following hypothesis:

$$\begin{array}{l} H_0 : h_\sigma = 1 \quad \forall Z \in \Theta \\ H_A : \exists Z \in \Theta \quad donde \quad h_\sigma > 1 \end{array} \tag{10}$$

we should define this test as follows:

$$LM - Scan_{SAR}^{SGWH} = \sup_{Z \in \Theta} \frac{g(h)^2 a^{44}}{\det(I(\theta)|_{H_0})} I(\sigma_Z^2 > \sigma_{\bar{Z}}^2) \tag{11}$$

for $I()$ an indicator function with value one if the inequality is true and zero otherwise.

2) SGWH LM Scan test for a SEM model

The SEM model can be specified as follows:

$$\begin{array}{l} \mathrm{y} = \mathrm{X}\beta + \mathrm{u} \\ \mathrm{u} = \lambda \mathrm{u} + \varepsilon \\ \varepsilon \sim \mathrm{N}(0, \Omega) \end{array} \tag{12}$$

The complete development of the LM scan test for SGWH in SEM models is shown in the Annex I.2. The formal expression of the SGWH LM Scan test in the SEM models is the following:

$$LM - Scan_{SEM}^{SGWH} = \sup_{Z \in \Theta} \frac{g(h)^2 b^{44}}{\det(I(\theta)|_{H_0})} \tag{13}$$

where the determinant of the information matrix can be expressed as a sum of terms which depends (or not) of the Z windows:

$$\begin{array}{l} \det(I(\theta)|_{H_0}) = \left(\dfrac{1}{\sigma^2}\right)^{k+3} \det\left(X'B'BX\right) \cdot \\ \cdot \left[I_{22} \dfrac{NN_Z}{4} + N_Z trWB^{-1} trH_h WB^{-1} - \dfrac{N}{2\sigma^2}\left(trH_h WB^{-1}\right)^2 - \dfrac{\sigma^2 N_Z}{2}\left(trWB^{-1}\right)^2 - \dfrac{N_Z^2}{4} I_{22} \right. \\ I_{22} = \sigma^2 trWB^{-1}WB^{-1} + \sigma^2 trB^{-1}W'WB^{-1} \end{array} \tag{14}$$

3.3 Statistical inference

The theoretical distribution of the scan test under the null hypothesis is unknown. For this reason, it is necessary to employ resampling techniques to evaluate the significance of this test. The resampling method calculates the p_B-value comparing the outcome of the test for the real dataset with the empirical distribution of B spatial reordered samples. The process is presented here:

1. The $LM-Scan_{SAR}^{SGWH}$ test is calculated for the original dataset $\{y_i, X_i\}_{i \in S}$, for S a set of spatial locations determined by their corresponding latitude and longitude coordinates.
2. The set of observations is reordered through a permutation process, assigning randomly observations to locations. Hence, the sets $\{y_i^r, X_i^r\}_{i \in S}$ are obtained, where r is the replica index.
3. This same permutation is also done on the rows and columns of the spatial weights matrix (W^r), so as the structure of the Z window elements is broken, though the spatial units keep their original spatial connectivity among themselves.
4. The $LM-Scan(r_i)_{SAR}^{SGWH}$ test is calculated for the permuted sample $\{y_i^r, X_i^r\}_{i \in S}$ with the corresponding permuted spatial weights matrix W^r.
5. Steps 2 to 4 are repeated (B–1) times to obtain a total number of B values of this test $\{LM(r_i)_{SAR}^{SGWH}\}_{i=1}^{B-1}$.
6. The pseudo p-value of this statistic is computed as follows:

$$p_B - value = \frac{1}{B}\sum_{i=1}^{B-1} I(Scan - LM_{SAR}^{SGWH} > Scan - LM(r_i)_{SAR}^{SGWH}) \qquad (15)$$

where $I(\bullet)$ is an indicator function which assigns the value one when the inequality is true and zero otherwise.

7. The null hypothesis is rejected when the p_B-$value < \alpha$ for an α nominal value.

3.4 Secondary clusters

The Scan-LM tests are based on statistics obtained under the alternative hypothesis of a single cluster (with a known form and size). If the test rejects the null hypothesis and identifies a significant cluster, a natural question would be to ask if there is another cluster, not overlapped with the most likely cluster, the variance of which are significantly different from the rest. These clusters are the so-called secondary ones. There are several ways to assign p-values to secondary clusters, see the paper of Zhang et al. (2010) where several alternatives are presented. The standard approach consists of ordering all the elements of Θ according to the likelihood ratio from highest to lowest. The most likely cluster (hereinafter MLC) will be the one that takes the maximum value. It will be assigned the p-value corresponding to the percentile that corresponds to the likelihood ratio in the distribution obtained by permutational resampling. The first secondary cluster will be the one that takes the maximum value within those elements of Θ that do not overlap with the MLC, assigning the p-value corresponding to the percentile that occupies the distribution obtained by permutational resampling. The procedure continues until no non-overlapping clusters are found statistically significant at a level (1-α)%. Zhang et al. (2010) show that this procedure yields conservative *p*-values. Therefore, they suggest an iterative method that consists of eliminating from the sample the observations included in the most probable cluster and re-obtaining the value of the statistic with this sub-sample once the cluster has been eliminated and all the statistically significant secondary clusters identified by this iterative process. Zhang et al. (2010) confirm that this procedure offers more power to identify secondary clusters. This will therefore be the method used in this paper.

1. Small-sample properties: a Monte-Carlo simulation

In this section, we evaluate the behavior of the tests developed in the previous section by means of a Monte-Carlo simulation. Its design is as follows:

i. The model is a simple regression model with one regressor (x_i) generated by a uniform distribution U(0,1) as following:

$$y_i = 2 + 3x_i + \varepsilon_i \tag{15}$$

Where the values of the coefficients in the model guarantee, in the absence of spatial effects, a value of R^2 close to 0.8.

ii. With respect to the error term;, we consider five different distributions:
   1. DGP1, the residuals (ε) follow a standard normal N(0,1) distribution;
   2. DGP2, the residuals present a pattern of random heteroskedasticity generated along a uniform distribution ε_i = N(0,u_i) where u_i is a realisation of a U(0,1);
   3. DGP3, a SARAR process with ρ = 0,5 and λ = 0,5

4. DGP4, a SEM process with $\lambda = 0{,}5$ (resp. SAR with $\rho = 0{,}5$)
5. DGP5, a SARAR process with $\rho = 0{,}2$ and $\lambda = 0{,}8$ (resp. $\rho = 0{,}8$ and $\lambda = 0{,}2$);

iii. We consider regular lattices of dimensions (7x7), (10x10), (15x15) y (20x20) that yield sample sizes of R = 49, 100, 225 and 400 respectively.

iv. Each combination is repeated 1000 times. The number of permutations in each repetition to obtain the p_B-*value* is set to B = 999.

v. The set Θ is formed by all circular windows[2] centered on each observation and with a maximum size of less that 50% of the total sample.

4.1. Size

Tables 2a and 2b show the sizes of the $\mathrm{ScanLM}_{\mathrm{SAR}}^{\mathrm{SGWH}}$ y $\mathrm{ScanLM}_{\mathrm{SEM}}^{\mathrm{SGWH}}$ tests for the various DGPs.

Table 2a: Size of the $\mathrm{ScanLM}_{\mathrm{SAR}}^{\mathrm{SGWH}}$. % of p_B-values< 0.05

	DGP1	DGP2	DGP3	DGP4	DGP5
7x7	0.049	0.046	0.064	0.036	0.041
10x10	0.047	0.052	0.061	0.037	0.051
15x15	0.043	0.041	0.064	0.056	0.048
20x20	0.054	0.059	0.049	0.053	0.049

Table 2b: Size of the $\mathrm{ScanLM}_{\mathrm{SEM}}^{\mathrm{SGWH}}$. % of p_B-values< 0.05

	DGP1	DGP2	DGP3	DGP4	DGP5
7x7	0.047	0.046	0.758	0.050	0.044
10x10	0.047	0.071	0.052	0.055	0.041
15x15	0.043	0.043	0.054	0.041	0.051
20x20	0.052	0.043	0.048	0.047	0.049

* 999 replications

The results show that overall the size of the test is correct in all situations. It is not affected by random heteroscedascity, nor by the presence of SARAR structures in the residuals.

[2] The power of the test is not affected by the type of window used (circular, elliptic of flexible), what changes is the capacity of the test to identify the true form of the cluster.

4.2. Power of the *Scan-LM* tests

In order to evaluate the power of the statistics, we consider the same patterns of SGWH presented in section 2. The results are displayed in Tables 2a and 2B.

Table 2a: Power of the $\text{ScanLM}_{\text{SAR}}^{\text{SGWH}}$. % of p_B-values< 0.05

	SGWH1	SGWH2	SGWH3	SGWH4	SGWH5	SGWH6	SGWH7
7x7	0.139	0.561	0.130	0.338	0.270	0.331	0.321
10x10	0.145	0.702	0.201	0.590	0.330	0.443	0.544
15x15	0.153	0.824	0.272	0.867	0.456	0.579	0.678
20x20	0.138	0.813	0.313	0.936	0.652	0.811	0.910

Table 2a: Power of the $\text{ScanLM}_{\text{SEM}}^{\text{SGWH}}$. % of p_B-values< 0.05

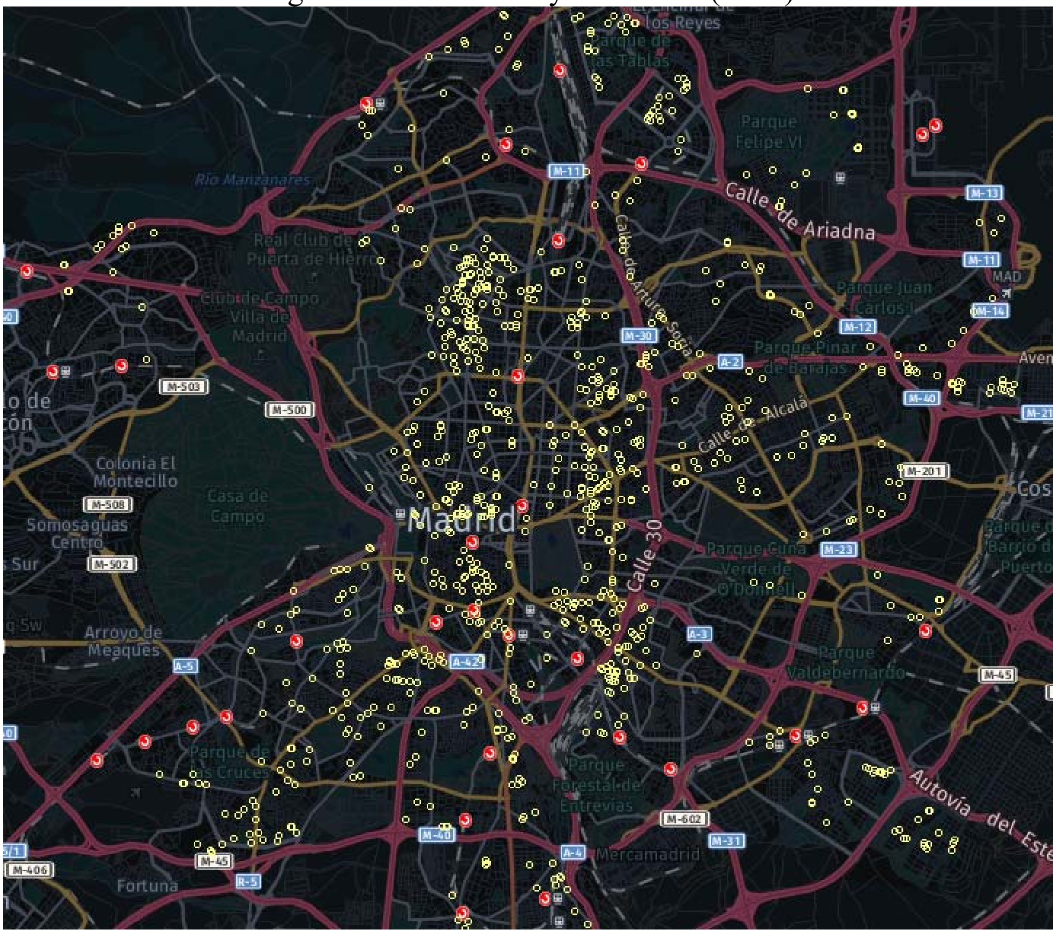
	SGWH1	SGWH2	SGWH3	SGWH4	SGWH5	SGWH6	SGWH7
7x7	0.120	0.534	0.127	0.296	0.254	0.364	0.401
10x10	0.133	0.673	0.184	0.575	0.443	0.551	0.621
15x15	0.159	0.811	0.261	0.861	0.601	0.759	0.769
20x20	0.144	0.809	0.298	0.940	0.784	0.882	0.927

The results show the high power of the test even in sample samples. The power is also directly related to the difference between variances inside and outside the cluster.

5. Application on attic prices in the city of Madrid

In this section, the behavior of the $Scan_\sigma$ test as a useful guide to specification search is illustrated with an application on a hedonic model for attic prices in the city of Madrid. Madid is a Spanish municipality formed of 21 districts (divided in turn into 128 neighborhoods and 2,420 census tracts). Particularly, we focus on attics, which are considered as the star performers in the Spanish real estate market in either the coast or the city: terraces, views and altitude give a value added to these properties. Urban citizens living in inland regions like Madrid show an overwhelming desire to buy a residence where they can enjoy of (semi-)fresh air and disconnect from the concrete jungle (Sánchez-Martín 2015, Gantes 2017). So it is interesting to evaluate the real price of certain amenities in this submarket inside Madrid. Our records refer to January 2015 and were drawn from an on-line real estate database, 'idealista.com' since, due to confidentiality constraints, it is almost impossible to obtain housing prices microdata from Spanish official institutions. The asking price has then been used as a proxy for the selling price as usual in many other cases (Cheshire and Sheppard 1998, Orford 2000, Chasco and Le Gallo, 2013). In total, 886 attic prices were finally recorded after the corresponding consolidation and geocoding processes.

Fig. 1: Attics in the city of Madrid (2015)



As a benchmark model, we first specify a standard hedonic house price model with a broad set of explanatory variables: twelve are attribute variables and fifteen are accessibility measures, since they are frequently advertised by real estate agents and often capitalized in housing prices. Additionally, in order to proxy all the micro-geographic determinants that buyers and sellers can observe, but are hidden for the econometricians, we also include the Earth coordinates of latitude and longitude plus 133 dummy variables corresponding to the 21 districts and 112 neighborhoods with attics, which can be considered as contextual variables (e.g. Anselin and Lozano-Gracia 2008).[3] Table 4 contains a listing of all variables together with their definitions.

Variables used in the model

Variable	*Description*	*Source*	*Units*
ly	Sale price	Idealista	Euros (logs)
FLOOR	House floors	Idealista	number
LM2B	Surface	Idealista	m^2 (logs)
bedr	Bedrooms	Idealista	number
REFORM	Needs renovation	Idealista	0-1
NEW	New	Idealista	0-1
NOLIFT	Building without elevator	Idealista	0-1
INTERIOR	All the rooms are facing an inner courtyard	Idealista	0-1
EXTERIOR	All the rooms are facing outdoor public areas	Idealista	0-1
gara	Garage space	Idealista	0-1
STORE	Storage room	Idealista	0-1
TERRACE	Attic with terrace	Idealista	0-1
VIEWS	Nice views	Idealista	0-1
Variables of accessibility			
DISCEN	Distance to the business center ("Nuevos Minist.")	Self-elaboration	Km
DISSOL	Distance to the historical center ("Puerta del Sol")	Self-elaboration	Km
PERIPH	Peripheral districts (outside Central Almond)	Self-elaboration	0-1
dm30	Distance to closest entrance to the M-30	Self-elaboration	Km
dm40	Distance to closest entrance to the M-40	Self-elaboration	Km
dmet	Distance to the closest metro station	Self elab./C.Madrid	Km
MINCERCA	Distance to the closest train station	Self elab./C.Madrid	Km
dhub	Distance to closest intermodal transport hub	Self-elaboration	Km
dair	Distance to the international airport	Self elab./C.Madrid	Km
MINPARK	Distance to the closest green area (over 1 Ha.)	Self elab./C.Madrid	Km
dhip	Distance to closest hypermarket	Self elab./C.Madrid	Km
dsho	Distance to closest shopping center	Self elab./C.Madrid	Km
dkil	Distance to closest 'category killer' center	Self elab./C.Madrid	Km
dser	Distance to service providers' outlets (retailing, hotels and restaurants)		Km
MINNIGHT	Distance to the nearest leisure night outlet (restaurants, bars, casinos, cinemas, theaters…)	Self.elab./SABI	Km
Geographical characteristics			
D1-21	District 1 to 21	C. Madrid	-
N1-112	Neighborhood 1 to 112	C. Madrid	-
xcoo	Longitude coordinate	Self.elab./GIS	Km (UTM)
ycoo	Latitude coordinate	Self.elab./GIS	Km (UTM)
xyco	Longitude × Latitude coordinate	Self-elaboration	Km^2

We first specify **three standard hedonic house models**, expressed in semi-log form, to compute the SGWH can test for OLS residuals (Chasco et al. 2018).

1. A standard hedonic house model for structural attributes:

[3] Districts are official administrative units defined by the Spanish National Statistics Office (INE) and neighborhoods, which are nested in the districts, are officious divisions recognized by the city council (http://www.munimadrid.es). Neighborhoods are characterized by certain homogeneity in terms of population density, infrastructure, historical and socioeconomic features.

$$lpri_i = \beta_0 + \sum_{s=1}^{S} \alpha_s x_{si} + u_i \qquad (14)$$

where *lpri*$_i$ is the log of price of transaction *i*; *S* is the number of property structural attributes, x_{si} and u_i is a well-behaved error term.

2. A standard hedonic house model for structural attributes and accessibility variables:

$$lpri_i = \beta_0 + \sum_{s=1}^{S} \alpha_s x_{si} + \sum_{c=1}^{C} \gamma_c x_{ci} + u_i \qquad (14)$$

where *C* is the number of accessibility variables x_{ci}.

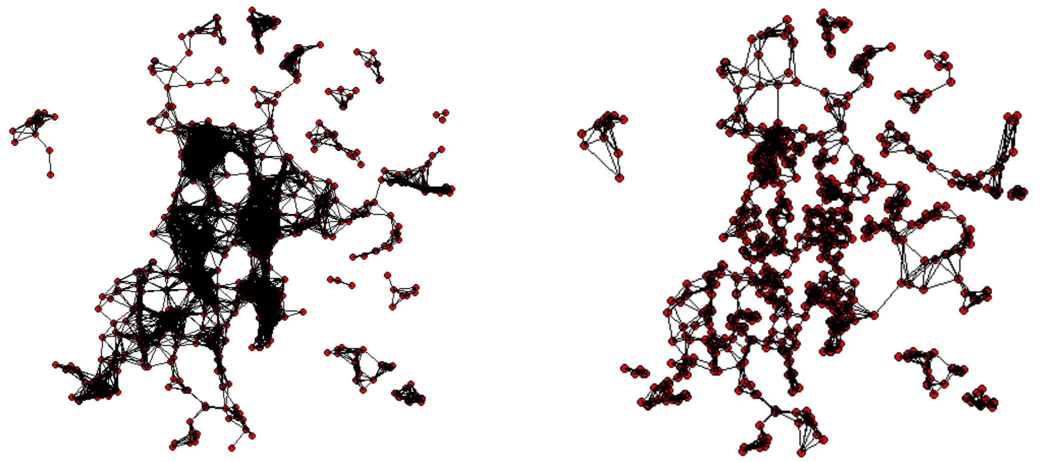
3. A standard hedonic house model for structural attributes, accessibility variables and geographical contextual variables :

$$lpri_i = \beta_0 + \sum_{s=1}^{S} \alpha_s x_{si} + \sum_{c=1}^{C} \gamma_c x_{ci} + \sum_{g=1}^{G} \delta_g x_{gi} + u_i \qquad (14)$$

where *G* is the number of geographic contextual variables x_{gi}.

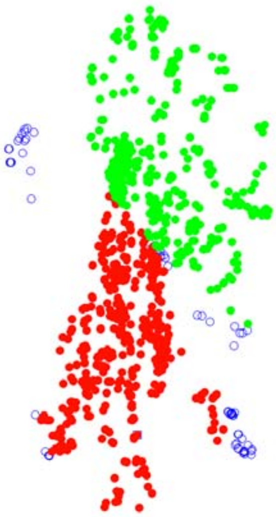
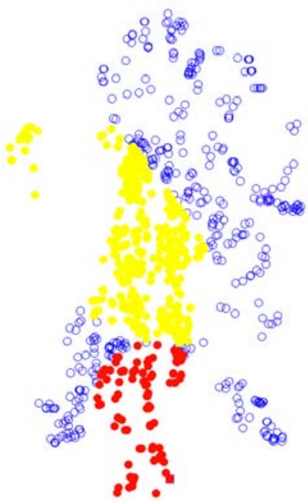
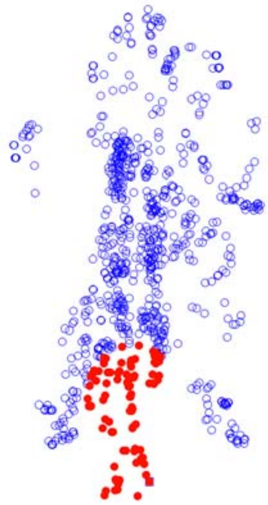
In order to test for spatial autocorrelation and specify spatial models, we select **spatial weights matrices** connecting nearby attics, which are more or less located in a same neighborhood:

Minimum distance (1.06 km.) | **Six nearest neighbors (average)**



Since the number of the regressors is so high, we opt to use a non-parametric regression technique called "multivariate adaptive regression splines" (MARS) as introduced by Friedman (1991).

The application of the SHGW scan test offer the following clusters:

OLS-1 Only structural variables	OLS-2 Structural+accessibility variables	OLS-3 Struatural+accessibility+ contextual variables
		

STILL PENDING…

6. Conclusion

Spatial autocorrelation and spatial heteroskedasticity are of fundamental importance in spatial econometric modeling and and the current state of the literature still lacks tests oriented towards the detection of spatial heteroskedasticity.

This paper has developed tests of SGWH on the residuals of the two most common spatial regression models (SAR/SEM). The scan methodology, applied on tests based on the Lagrange multiplier principle, not only allows testing the null hypothesis of homoscedasticity but also, in the case of rejecting the null hypothesis, suggests the cluster(s) with unequal variance. This information can be used, either to identify the causes that generate this heteroskedasticity and to improve the specification of the model or to introduce heteroskedastic specification in spatial models. This has been illustrated on read housing prices data for the housing market in Madrid.

Appendix. A graphical example of the spatial bootstrap procedure

Figures 3a and 3b show graphically the resampling method shown in section 3. In Figure 3a, the scan test identifies a spatial cluster of 9 observations (in black) with a different behavior. The lines indicate the connectivity structure defined by the spatial weights matrix W. The scanning process consider a total amount of #(Θ) number of *Z* dummy variables (windows) capturing all the possible clusters of observations. The LM_{α} statistic will get the maximum value for the dummy variable with these nine observations. In order to evaluate the statistical significance of this test, we permute randomly these observations (sampling without replacement), the matrix of explanatory variables *X*, as well as the spatial weight matrix.

Figure 3b shows the spatially reshuffled observations y_p, so as the lines show the new configuration of permuted spatial weight matrix, W_p. Note that the rows and columns of the *W* matrix are also permuted in the same way, keeping the initial neighborhood structure. For this permutation, a new scanning process is produced using all the *Θ* predefined windows, though in this case, we can hardly find a spatial cluster and the values of the LM_{α} test will be, in general, less than the obtained with the original configuration. Note that the estimation of the SAR model does not experience any variation because we have used the same permutation for the observations of the dependent variable (*y*), the regressors matrix (*X*) and the spatial weight matrix (*W*).

Figure 3: Resampling by permutation in a spatial setting

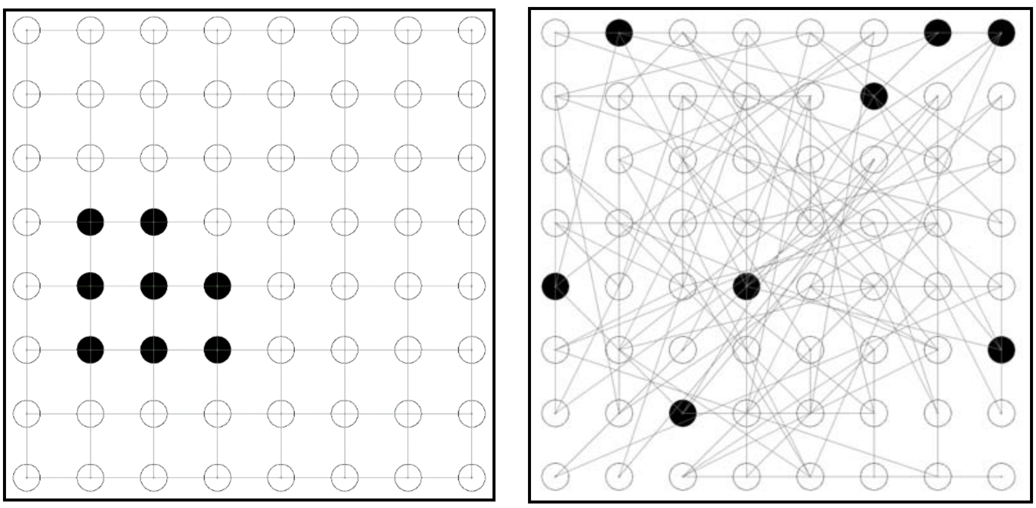


Figure 3a Figure 3b

Numerically, this problem is equivalent to comparing the maximum value of the LM_α test with connected windows with the maximum value of the LM_α test for the B sets of disconnected windows.

References

Abrams, A.M., Kleinman, K., Kulldorff, M. (2010): Gumbel based p-value approximations for spatial scan statistics. *International Journal of Health Geographics*, 9(1), 61.

Anselin, L. (1988): Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20(1), 1-17.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J. (1996): Simple diagnostic test for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77-104.

Anselin, L., Rey, S. (1991): Properties of test for spatial dependence in linear regression models. *Geographical Analysis*, 23(2), 112-131.

Baltagi, B.H., Yang, Z. (2013): Standardized LM test for spatial error dependence in linear or panel regressions. *The Econometrics Journal*, 16(1), 103-134.

Burridge, P. (1980): On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 42(1), 107-108.

Born, B., Breitung, J. (2011): Simple regression-based test for spatial dependence. *The Econometrics Journal*, 14(2), 330-342.

Chasco, C., Le Gallo J., López, F.A. (2018): A scan test for spatial groupwise heteroscedasticity in cross-sectional models with an application on house prices in Madrid. *Regional Science and Urban Economics*, 68, 226-238.

Dall'Erba, S., Le Gallo, J. (2008): Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis. *Papers in Regional Science*, 87(2), 219-244.

Fischer, M.M., Stirböck, C. (2006): Pan-European regional income growth and club-convergence. Insights from a spatial econometric perspective. *Annals of Regional Science*, 40, 693-721.

Friedman J. H. (1991) Multivariate Adaptive Regression Splines, *The Annals of Statistics* 19(1), 1–141.

Gantes, Y. (2017), Objetivo: ático en Las Salesas, el barrio de los bohemios con dinero de Madrid, *elEconomista.es*, 14/07/2017.

Guo, P., Lihu, L., Zhengming, Q. (2015): Robust test for spatial error model: considering changes of spatial layouts and distribution misspecification. *Communications in Statistics-Simulation and Computation*, 44(2), 402-416.

Kelejian, H.H., Robinson, D.P. (2004): The influence of spatially correlated heteroskedasticity on test for spatial correlation. In Anselin L., Florax R.J.G.M., Rey S.J. (Eds.), *Advances in Spatial Econometrics* (pp. 79-97). Springer Berlin Heidelberg.

Kulldorff, M., Huang, L., Konty, K. (2009): A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geography*, 8, 58–73.

Kulldorff, M., Nagarwalla, N. (1995): Spatial disease clusters: detection and inference. *Statistics in Medicine*, 14, 799–810.

López, F.A., Chasco, C., Le Gallo, J. (2015): Exploring scan methods to test spatial structure with an application to housing prices in Madrid. *Papers in Regional Science*, 94, 317-356.

Robinson, P.M., Rossi, F. (2014): Improved Lagrange multiplier test in spatial autoregressions. *The Econometrics Journal*, 17(1), 139-164.

Sánchez-Martín, A. (2015) Los áticos, las estrellas del mercado, *El Mundo*, 22/07/2015.

Tango, T., Takahashi, K. (2005): A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4(1), 11.

Zhang, Z., Assunção, R., Kulldorff, M. (2010): Spatial scan statistics adjusted for multiple clusters. *Journal of Probability and* Statistics, 2010.

ANNEX I.1:

Scan-SAR

In order to obtain the LM test, we evaluate the gradient of the log-likelihood function:

$$\frac{\partial l}{\partial \beta} = (Ay - X\beta)'\Omega^{-1}X$$

$$\frac{\partial l}{\partial \rho} = -\mathrm{tr}\,A^{-1}W + (Ay - X\beta)'\Omega^{-1}Wy$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}\mathrm{tr}\,\Omega^{-1}H_{\sigma^2} + \frac{1}{2}(Ay - X\beta)'\Omega^{-2}H_{\sigma^2}(Ay - X\beta)$$

$$\frac{\partial l}{\partial h} = -\frac{1}{2}\mathrm{tr}\,\Omega^{-1}H_h + \frac{1}{2}(Ay - X\beta)'\Omega^{-2}H_h(Ay - X\beta)$$

where $H_h = H_h = \frac{\partial \Omega}{\partial h}$

Under the null

$$g(\theta)\,|_{H_0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\frac{R_Z}{2} + \frac{1}{2}\frac{u'_Z u_Z}{\sigma^2} \end{bmatrix}$$

Where $u = Ay - X\beta$. By u_Z, we denote the elements of the vector u that belong to zone Z. In order to obtain the elements of the information matrix $I = -E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right]$, it is necessary to derive the following elements under the null:

$$I_{\beta\beta} = X'\Omega^{-1}X$$

$$I_{\beta\rho} = X'\Omega^{-1}WA^{-1}X\beta$$

$$I_{\beta\sigma^2} = 0$$

$$I_{\beta h} = 0$$

$$I_{\rho\rho} = \operatorname{tr} WA^{-1}WA^{-1} + \operatorname{tr}\Omega A'^{-1}W'\Omega^{-1}WA^{-1} + \beta'X'A'^{-1}W'\Omega^{-1}WA^{-1}X\beta$$

$$I_{\rho\sigma^2} = \operatorname{tr}\Omega^{-1}H_{\sigma^2}WA^{-1}$$

$$I_{\rho h} = \operatorname{tr}\Omega^{-1}H_{h}WA^{-1}$$

$$I_{\sigma^2\sigma^2} = \frac{1}{2}\operatorname{tr}\Omega^{-2}H_{\sigma^2}H_{\sigma^2}$$

$$I_{\sigma^2 h} = \frac{1}{2}\operatorname{tr}\Omega^{-2}H_{\sigma^2}H_{h}$$

$$I_{hh} = \frac{1}{2}\operatorname{tr}\Omega^{-2}H_{h}H_{h}$$

The information matrix under the null has the following expression:

$$I(\theta)|_{H_0} = \frac{1}{\sigma^2}\begin{bmatrix} X'X & X'WA^{-1}X\beta & 0 & 0 \\ -- & \sigma^2\operatorname{tr} WA^{-1}WA^{-1} + \sigma^2\operatorname{tr} H + \beta'X'HX\beta & \operatorname{tr} WA^{-1} & \operatorname{tr} H_h WA^{-1} \\ -- & -- & \frac{1}{2}\frac{R}{\sigma^2} & \frac{R_Z}{2} \\ -- & -- & -- & \frac{1}{2}\sigma^2 R_Z \end{bmatrix}$$

$$H = A'^{-1}W'WA^{-1}$$

Finally, the LM test has the following general expression:

$$LM(Z)_{SAR}^{SGWH} = \left[g(\theta)|_{H_0}\right]'\left[I(\theta)|_{H_0}\right]^{-1}\left[g(\theta)|_{H_0}\right]$$

Which, after some computations, can be expressed as:

$$LM_{SAR}^{SGWH} = \frac{g(h)^2 a^{44}}{\det(I(\theta)|_{H_0})}$$

The expression of the determinant of the information matrix can be written in compact form as:

$$\det(I(\theta)|_{H_0}) = \left(\frac{1}{\sigma^2}\right)^{k+3}\left[\frac{\sigma^2 R_Z}{2}\det(A_{33}) - \frac{R_Z}{2}\left(\frac{R_Z}{2}\det(A_{22}) - \mathrm{tr}WA^{-1}\det(X'X)\mathrm{tr}\,H_h WA^{-1}\right) + \right.$$

$$\left. +\mathrm{tr}\,H_h WA^{-1}\det(X'X)\left(\frac{R_Z}{2}\mathrm{tr}WA^{-1} - \frac{R}{2\sigma^2}\mathrm{tr}\,H_h WA^{-1}\right)\right]$$

where A_{22} is the following matrix that does not depend upon Z:

$$A_{22} = \begin{bmatrix} X'X & X'WA^{-1}X\beta \\ -- & \sigma^2 \mathrm{tr}\,WA^{-1}WA^{-1} + \sigma^2 \mathrm{tr}\,H + \beta'X'HX\beta \end{bmatrix}$$

where

$$A_{33} = \begin{bmatrix} X'X & X'WA^{-1}X\beta & 0 \\ -- & \sigma^2 \mathrm{tr}\,WA^{-1}WA^{-1} + \sigma^2 \mathrm{tr}\,H + \beta'X'HX\beta & \mathrm{tr}\,WA^{-1} \\ -- & -- & \frac{1}{2}\frac{R}{\sigma^2} \end{bmatrix}$$

ANNEX I.2:

Scan-SEM

In order to obtain the LM test, we evaluate the gradient of the log-likelihood function:

$$\frac{\partial l}{\partial \beta} = (y - X\beta)'B'\Omega^{-1}BX$$

$$\frac{\partial l}{\partial \lambda} = -\mathrm{tr}\,B^{-1}W + (y - X\beta)'B'\Omega^{-1}W(y - X\beta)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}\mathrm{tr}\,\Omega^{-1}H_{\sigma^2} + \frac{1}{2}(y - X\beta)'B'\Omega^{-2}H_{\sigma^2}B(y - X\beta)$$

$$\frac{\partial l}{\partial h} = -\frac{1}{2}\mathrm{tr}\,\Omega^{-1}H_h + \frac{1}{2}(y - X\beta)'B'\Omega^{-2}H_h B(y - X\beta)$$

The gradient under the null hypothesis is:

$$g(\theta)|_{H_0} = \begin{bmatrix} g(\beta)|_{H_0} \\ g(\lambda)|_{H_0} \\ g(\sigma^2)|_{H_0} \\ g(h)|_{H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\frac{R_Z}{2} + \frac{1}{2}\frac{v'_Z v_Z}{\sigma^2} \end{bmatrix}$$

where $u = y - X\beta$; $v = Bu$; by v_Z, we denote the elements of the vector v that belong to zone Z.

In order to obtain the elements of the information matrix $I = -E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right]$, it is necessary to derive the following elements under the null:

$$I_{\beta\beta} = X'B'\Omega^{-1}BX$$
$$I_{\beta\lambda} = 0$$
$$I_{\beta\sigma^2} = 0$$
$$I_{\beta h} = 0$$
$$I_{\lambda\lambda} = \mathrm{tr}\,WB^{-1}WB^{-1} + \mathrm{tr}\,\Omega B'^{-1}W'\Omega^{-1}WB^{-1}$$
$$I_{\lambda\sigma^2} = \mathrm{tr}\,\Omega^{-1}H_{\sigma^2}WB^{-1}$$
$$I_{\lambda h} = \mathrm{tr}\,\Omega^{-1}H_h WB^{-1}$$
$$I_{\sigma^2\sigma^2} = \frac{1}{2}\mathrm{tr}\,\Omega^{-2}H_{\sigma^2}H_{\sigma^2}$$
$$I_{\sigma^2 h} = \frac{1}{2}\mathrm{tr}\,\Omega^{-2}H_{\sigma^2}H_h$$
$$I_{hh} = \frac{1}{2}\mathrm{tr}\,\Omega^{-2}H_h H_h$$

The information matrix under the null has the following expression:

$$I(\theta)|_{H_0} = \frac{1}{\sigma^2}\begin{bmatrix} X'B'BX & 0 & 0 & 0 \\ -- & \sigma^2\mathrm{tr}\,WB^{-1}WB^{-1} + \sigma^2\mathrm{tr}\,B'^{-1}W'WB^{-1} & \mathrm{tr}\,WB^{-1} & \mathrm{tr}\,H_h WB^{-1} \\ -- & -- & \frac{1}{2}\frac{R}{\sigma^2} & \frac{R_Z}{2} \\ -- & -- & -- & \frac{1}{2}\sigma^2 R_Z \end{bmatrix}$$

After some calculations, we obtain:

$$LM_{SEM}^{SGWH} = \frac{g(h)^2 b^{44}}{\det(I(\theta)|_{H_0})} \sim \chi_1^2$$

where

$$b^{44}(\text{no depende de } Z) = \left(\frac{1}{\sigma^2}\right)^{k+2}\det\left(\begin{bmatrix} X'X & 0 & 0 \\ & \sigma^2\mathrm{tr}\,WB^{-1}WB^{-1} + \sigma^2\mathrm{tr}\,B'^{-1}W'WB^{-1} & \mathrm{tr}\,WB^{-1} \\ & & \frac{1}{2}\frac{R}{\sigma^2} \end{bmatrix}\right) =$$

$$= \left(\frac{1}{\sigma^2}\right)^{k+2}\det(X'X)\left(\frac{1}{2}\frac{R}{\sigma^2}(\sigma^2\mathrm{tr}\,WB^{-1}BA^{-1} + \sigma^2\mathrm{tr}\,B'^{-1}W'WB^{-1}) - \left(\mathrm{tr}\,WB^{-1}\right)^2\right)$$

The determinant of the information matrix can be expressed as a linear combination of terms that depend of Z:

$$\det(\mathbf{I}(\theta)|_{\mathbf{H}_0}) = \left(\frac{1}{\sigma^2}\right)^{\mathbf{k}+3} \det(\mathbf{X'B'BX})\left[\mathbf{I}_{22}\frac{\mathbf{RR_Z}}{4} + \mathbf{R_Z trWB^{-1}tr\,H_hWB^{-1}} - \frac{\mathbf{R}}{2\sigma^2}\left(\mathbf{tr\,H_hWB^{-1}}\right)^2 - \right.$$

$$\left. -\frac{\sigma^2\mathbf{R_Z}}{2}\left(\mathbf{trWB^{-1}}\right)^2 - \frac{\mathbf{R_z^2}}{4}\mathbf{I}_{22}\right]$$

$$\mathbf{I}_{22} = \sigma^2\mathbf{tr\,WB^{-1}WB^{-1}} + \sigma^2\mathbf{tr\,B'^{-1}\,W'WB^{-1}}$$

ANNEX I.3:

LM^*_{ERR} in SAR

We derive the form of the marginal test of spatial error autocorrelation in SAR models: LM^*_{ERR}

The gradient:

$$g(\theta)|_{H_0} = \begin{bmatrix} g(\beta)|_{H_0} \\ g(\rho)|_{H_0} \\ g(\lambda)|_{H_0} \\ g(\sigma^2)|_{H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ -\mathrm{tr}\,A^{-1}W + \frac{u'Wy}{\sigma^2} \\ \frac{u'Wu}{\sigma^2} \\ -\frac{R}{2\sigma^2} + \frac{1}{2}\frac{u'u}{\sigma^4} \end{bmatrix}$$

and the matrix of information under the null:

$$I(\theta)|_{H_0} = \frac{1}{\sigma^2}\begin{bmatrix} X'X & X'WA^{-1}X\beta & 0 & 0 \\ -- & \sigma^2\mathrm{tr}\,WA^{-1}WA^{-1} + \sigma^2\mathrm{tr}\,H + \beta'X'HX\beta & \sigma^2\mathrm{tr}\,W'WA^{-1} + \sigma^2\mathrm{tr}(WWA^{-1}) & \mathrm{tr}\,WA^{-1} \\ -- & -- & \sigma^2\mathrm{tr}WW + \sigma^2\mathrm{tr}W'W & 0 \\ -- & -- & -- & \frac{R}{2} \end{bmatrix}$$

$$H = A'^{-1}W'WA^{-1}$$

The test LM^*_{ERR} is:

$$LM_{\mathrm{quedaSEM}} = \left[g(\theta)|_{H_0}\right]'\left[I(\theta)|_{H_0}\right]^{-1}\left[g(\theta)|_{H_0}\right]$$