Tests for Spatial Lag Dependence Based on Method of Moments Estimation

by

Luz A. Saavedra*
Department of Economics
University of South Florida
4202 East Fowler Ave. BSN 3111
Tampa, FL 33620-5524
lsaavedr@coba.usf.edu

January 2000

Abstract

In this paper, we formulate GMM versions of the Wald, the Likelihood ratio and the Lagrange multiplier test statistics for spatial lag dependence in the spatial lag model with autocorrelated errors. The tests are based on the GMM estimator suggested by Kelejian and Robinson (1993) and on the work of Newey and West (1987). We also analyze their finite-sample properties in a Monte Carlo Experiment. The tests are computationally simple and asymptotically equivalent to their maximum likelihood counterpart. The Monte Carlo results indicate that the small sample distribution of the tests is well approximated by their asymptotic chi-squared distribution for relatively small sample sizes, but for small values of the spatial error parameter. In addition, the empirical power approaches one for low values of the spatial lag parameter under reasonable small sample sizes. However, we find variations across different spatial matrices for given sample size and given values of the spatial error parameter.

* This paper is drawn from dissertation work at the University of Illinois at Urbana-Champaign. I thank Omar Arias and Anil Bera for useful comments. In addition, I thank Roger Koenker for allowing me to run the Monte Carlo experiments in Ragnar.econ.uiuc.edu

Tests for Spatial Lag Dependence Based on Method of Moments Estimation

by

Luz A. Saavedra

1. Introduction

The main characteristic of spatial econometrics is that it allows for the presence of spatial dependence in cross-sectional data. Spatial dependence can arise from spatial correlation between non-observable explanatory variables (disturbance terms) or from spillover e®ects that determine the behavioral structure of the model. An example of the ¯rst case occurs when the disturbance in one observation, say in spatial unit $i;$ is correlated with the disturbance in other observations, say spatial units $j:$ In this case the errors follow a spatial autoregressive process. The second case occurs when the values of the dependent variable for a given observation depend on the values it takes for other observations, creating models with a spatially lagged dependent variable. The resulting speci¯cations are known as spatial error and spatial lag models, respectively. In both models, a weight matrix plays the role of a \spatial lag operator," which shifts the error term and the dependent variable in space, resembling a lag or forward operator in the time series context. In addition, the presence of both spatial e®ects, yields a spatial lag model with a spatial autoregressive disturbance.[1]

Because no consistent estimates of the parameters of these models can be obtained from ordinary least squares estimation, the maximum likelihood approach has become the most common method of estimation and speci¯cation testing. Alternatively, instrumental variables (IV) estimation has sometimes been implemented, mainly for the estimation of the spatial lag model. However, these methods of estimation have been rarely used in the estimation of the spatial lag model with a spatial autoregressive disturbance, denoted hereafter the \full spatial model." Exceptions are Case et al. (1993), and Kelejian and Robinson, (1993). In part, this may be due to the di±culties that arise in the operational implementation of maximum likelihood estimation of this spatial model, and to a potential problem of identi¯cation of the spatial parameters that have been documented in the spatial econometrics literature (see Anselin (1988a), Anselin and Bera (1996), and Anselin et al. (1996)).[2]

Speci¯cation testing in spatial econometrics has received considerable attention in this literature. Since Cli® and Ord's (1972) extension of Moran's (1950) test for spatial autocorrelation to regression residuals, there has been extensive work focused on the derivation of di®erent speci¯cation tests. Examples are Burridge's (1980) test for spatial error autocorrelation, Anselin's (1988b) tests for spatial lag dependence and for the joint hypothesis of spatial lag and error dependence (for other available tests see Anselin and Florax (1995)).

The need for available estimators and test statistics that do not depend on stringent distributional and functional form assumptions has generated an increasing interest in the development of robust methods of estimation and speci¯cation testing in econometrics.[3] This interest has also captured the attention of spatial econometricians, who have tried to develop estimation methods which do not require distributional assumptions, and which provide practitioners with methods of estimation that are less computationally challenging than maximum likelihood. Examples of suggested methods of estimation that do not need strong parametric assumptions are the general method of moments (GMM) estimator for the spatial error components model of Kelejian and Robinson (1993), the generalized method of moments estimator for a spatial error model of Kelejian and Prucha (1996), and more recently, the generalized spatial two state least squares estimator for a spatial autoregressive model with autoregressive disturbances of Kelejian and Prucha (1997).

This literature has also focused on the derivation of test statistics that are robust to a variety of misspeci¯cations. Examples are the robust Lagrange multiplier tests for spatial lag dependence and for error dependence by Anselin et al. (1996), and tests of spatial dependence that are robust to heteroscedasticity of an unspeci¯ed form, by Anselin (1988b) (see also Anselin (1990)). An interesting feature of these tests is that they can be constructed based on least squares residuals.

Test statistics that do not require the speci¯cation of the distribution of the disturbances that are widely used in spatial econometrics are Cli® and Ord's (1972) version of Moran's (1950) test for spatial dependence and Kelejian and Robinson's (1992) test for error dependence. Another robust method, which is used in applied econometrics and that has been suggested for inference in spatial models, is the bootstrap method.[4]

The development of estimators that do not require stringent distributional assumptions offers opportunities for the development of test procedures for spatial econometric models within this framework. The availability of these tests procedures is of potential interest to practitioners.

In this paper, we formulate GMM versions of the Wald, the Likelihood ratio and the Lagrange multiplier test statistics (denoted hereafter $\text{Wald}_G; \text{LR}_G;$ and $\text{LM}_G;$ respectively) for spatial lag dependence in the spatial lag model with autocorrelated errors (ie. full spatial model). The tests are based on the generalized method of moments (GMM) estimator suggested by Kelejian and Robinson (1993), on the work of Kelejian and Prucha (1996) and on the efficient GMM tests formulated by Newey and West (1987). As Newey and West show these tests are computationally simple, and as their maximum likelihood counterparts, they are large sample tests that are asymptotically equivalent and distributed as chi-squared random variables. In addition, under certain conditions, they are also numerically equivalent, giving to practitioners some freedom in the choice of the test statistic (see Newey and West (1987)).

We investigate the finite sample performance of the $\text{LM}_G;$ the $\text{LR}_G;$ and the Wald_G and compare them with the robust test for spatial lag dependence of Anselin et al. (1996), in a Monte Carlo simulation.[5]

The Monte Carlo experiments indicate that the small sample distribution of the tests is well approximated by their asymptotic chi-squared distribution for relatively small sample sizes, but for small values of the spatial error parameter. In addition, the empirical power approaches one for low values of the spatial lag parameter under reasonable small sample sizes. However, we find variations across different spatial matrices for given sample size and given values of the spatial error parameter.

The second part of the paper outlines Kelejian and Robinson's (1993) GMM estimators, and Kelejian and Prucha's (1996) consistent estimator of the error parameters. Section 3 presents the GMM analogs of the Wald, likelihood ratio, and Lagrange multiplier test statistics. Section 4 describes and presents the results from the Monte Carlo experiment, and section 5 concludes.

2. GMM estimators

We will consider test statistics of the null hypothesis of no spatial lag dependece in a full spatial model and formulate GMM counterparts of the well known Wald, Lagrange multiplier, and Likelihood ratio test statistics for this hypothesis. As in the maximum likelihood framework, the W_G requires the unconstrained GMM estimator for its computation, the LR_G requires both the unconstrained and constrained GMM estimators, while the LM_G uses the gradient of the GMM criterion function, evaluated at the restricted parameter estimator. The restricted and unrestricted GMM estimators are based on the work of Kelejian and Robinson (1993) and Kelejian and Prucha (1996).

2.1 Kelejian and Robinson's (1993) Unconstrained GMM Estimator

Kelejian and Robinson (1993) consider the following spatial model:

$$y = \rho My + X\beta + u; \tag{1}$$

$$u = W\eta + \psi; \tag{2}$$

where y is a nx1 vector of observations on the spatial dependent variable, with its ith observation corresponding to the ith region; ρ is the spatial lag parameter, which is assumed to be less than one in absolute value; X is a nxk matrix of observations on the exogenous variables; β is a kx1 parameter vector; u is the regression disturbance vector, η and ψ are nx1 vectors of random shocks; and M and W are nxn non-stochastic weight matrices, which are known a priori. Unlike W, M has zeros in its main diagonal. Furthermore, the nonzero elements of the ith row of M and W define the set of "neighbors" and the set of error terms corresponding to the ith region, respectively.

Note that in this model, the error is composed of two stochastic shocks that are generated within each region. The component ψ is specific to each region, and the component η is common to different regions which are determined by W: These error vectors are assumed to be independent of X and independent of each other. In addition, their ith elements η_i and ψ_i are iid. with zero means and variance σ_η^2 and σ_ψ^2; respectively. Therefore, $E(\eta) = E(\psi) = 0$;

$E(\eta\eta') = \sigma^2_\epsilon I_n$; and $E(\psi\psi') = \sigma^2_\psi I_n$: The error covariance matrix is given by

$$E(uu') = \Sigma_u = \sigma^2_\epsilon WW' + \sigma^2_\psi I_n; \tag{3}$$

In the formulation of the GMM test statistics we will not consider (2). Instead, we will assume the following autoregressive error process

$$u = \lambda Wu + \psi; \tag{2a}$$

where λ a scalar spatial autocorrelation parameter, which is assumed to be less than one in absolute value, and ψ is as above. In addition, we note that in this formulation of the model W has diagonal elements equal to zero. The difference between this and Kelejian and Robinson's (1993) model is that we consider a disturbance that follows a spatial autoregressive process, while they consider a disturbance with two error components. Both spatial models are important in applied work. See for example, Anselin (1980), Brueckner (1997), Brueckner and Saavedra (1998), Case (1991), Case et al (1993), Cliff and Ord (1981), and Figlio and Reid (1998) for applications of both models.

As in Kelejian and Robinson (1993), we also assume that u is independent of X; and that $(I - \rho W)$ and $(I - \lambda W)$ are non-singular for all $|\rho| < 1$; and $|\lambda| < 1$: In addition, we assume that the weight matrices are the same, i.e. $M = W$: Rewriting equations (1) and (2a), the model we consider can be written as

$$y = (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}u; \tag{1'}$$

$$u = (I - \lambda W)^{-1}\psi: \tag{2a'}$$

The previous assumptions imply that the disturbance term has zero expectation, i.e. $E(u) = 0$; and variance-covariance matrix is given by

$$E(uu') = \Sigma_u = \sigma^2_\psi(I - \lambda W)^{-1}(I - \lambda W')^{-1}: \tag{4}$$

The assumptions also imply that $E(y/X) = (I - \rho W)^{-1}X\beta$ and $E(Wy\,u' \mid X) = [\sigma_{\varepsilon}^2 W(I - \rho W)^{-1}(I - \lambda W)^{-1}(I - \lambda W')^{-1}] \neq 0$. This last expression indicates that the parameters of equation (1) cannot be consistently estimated by ordinary least squares.[6]

Going back to Kelejian and Robinson's (1993) setup, let $Z = (Wy, X)$ and $\gamma' = (\rho, \beta')$. Let $H = (X, WX^*)$ denote a matrix of instruments of $\mathrm{rank}(H) = k + k^*$, where X^* is a submatrix of X consisting of $k^* < k$ of its columns.[7] Also, rewrite (1) as

$$y = Z\gamma + u. \tag{5}$$

Their efficient GMM estimator is obtained by minimizing a criterion function that is constructed from an orthogonality condition between the set of instruments H and the disturbance term u. This orthogonality condition is given by

$$E[H'u] = 0, \tag{6}$$

which corresponds to the population moment condition. The sample counterpart of (6) can be written as

$$F_n(\gamma) = n^{-1}H'u = 0, \tag{6'}$$

which is a $(k + k^*)$ by 1 vector of error sample averages, with variance-covariance matrix given by $\Omega = n^{-2}H'\Sigma_u H$.

Their GMM estimator chooses $\hat{\gamma}_{GMM}$ so that $F_n(\gamma)$ is close to zero. Thus, $\hat{\gamma}_{GMM}$ minimizes the criterion function given by

$$Q_n(\gamma) = u'H[H'\Sigma_u H]^{-1}H'u, \tag{7}$$

which corresponds to the general form of a GMM criterion function

$$Q_n(\gamma) = F_n(\gamma)'\Omega^{-1}F_n(\gamma), \tag{7'}$$

where Γ has been chosen to be the inverse of the covariance matrix Ω, ie., $\Gamma = \Omega^{-1}$. Due to this choice of Γ, their proposed GMM estimator is optimal (i.e., efficient) within the considered class of GMM estimators. Note that a consistent estimator of Ω^{-1} is needed to obtain $\hat{\gamma}_{GMM}$. According to (3), a consistent estimate of Ω can be obtained from consistent estimates of σ^2 and σ^2_{ψ}. After a consistent estimator of Ω is obtained and substituted in the criterion function (7), the minimization of this function with respect to γ results in

$$\hat{\gamma}_{GMM} = (Z'\hat{D}Z)^{-1}Z'\hat{D}y, \qquad (8)$$

where $\hat{D} = H(H'\hat{\Sigma}_u H)^{-1}H$. As is shown in their paper, $\hat{\gamma}_{GMM}$ is asymptotically distributed as

$$\sqrt{n}(\hat{\gamma}_{GMM} - \gamma) \to N(0, G'\Omega^{-1}G), \qquad (9)$$

where $G = E\Delta_{\gamma}F_n(\gamma)$, and Δ_{γ} denotes the derivative with respect to γ. From (5) and (6), $G = H'Z$. Thus, the estimated asymptotic covariance matrix of $\hat{\gamma}_{GMM}$ is given by

$$V(\hat{\gamma}_{GMM}) = [Z'H(H'\hat{\Sigma}_u H)^{-1}H'Z]^{-1} \qquad (10)$$

In short, Kelejian and Robinson (1993) proposed GMM estimator is obtained by using first a set of instruments H in a two stage least squares procedure to obtain a consistent preliminary estimate of γ. Then, this estimate is in turn substituted in (5) to obtain estimates of u, which are then used in a least squares regression to obtain consistent estimates $\hat{\sigma}^2$ and $\hat{\sigma}^2_{\psi}$. Finally, γ_{GMM} is obtained from (8). The following steps summarize the estimation process:

(i) Regress Wy on H, and obtain fitted values $\widehat{Wy} = H(H'H)^{-1}H'Wy$.

(ii) Obtain a preliminary consistent estimate of γ regressing y on $\hat{Z} = (\widehat{Wy}, X)$.

(iii) Use the preliminary estimate $\hat{\gamma}$ in (5) to obtain estimates $\hat{u}$ of u.

(iv) To obtain consistent estimates of σ^2 and σ^2_{ψ}, regress $\hat{u}^2$ on the diagonal elements of WW' as suggested by (3). Use these estimates to obtain $\hat{\Sigma}_u$.

(v) Finally, substitute $\hat{\Sigma}_u$ in (7) and obtain $\hat{\gamma}_{GMM}$ as indicated by equation (8).[8]

Because we consider the model given by equations (1') and (2a'), we need consistent estimators of λ and σ^2_ε to implement our version of (8). Note that in our case, $\Omega = n^{-2}H'\Sigma_u H$, where Σ_u is given by (4). According to (4), a consistent estimate of Σ_u can be obtained by using consistent estimates of σ^2_ε and λ. Observe that the difference between our model and Kelejian and Robinson's (1993) model is that ours contains the parameter λ, which is not present in theirs, and ours does not contain the element σ^2, which is present in theirs. Overall, this implies that step (iv) above needs to be modified. To estimate these error parameters, we use Kelejian and Prucha's (1996) suggested generalized moments estimator, which we describe next.

2.2. Kelejian and Prucha's (1996) Generalized Moments Estimator

Kelejian and Prucha (1996) consider a generalized moments estimator for the parameters of a linear model with the error term following a spatial autoregressive process like the one described by equation (2a), or equivalently, by (2a'). Using our notation, we briefly describe their estimator.

Let $\bar{u} = Wu$, $\bar{\bar{u}} = WWu$, and $\bar{\varepsilon} = W\varepsilon$. Recall that ε_i is iid.$(0, \sigma^2_\varepsilon)$, and that the diagonal elements of the weight matrix W are equal to zero, i.e., $w_{ii} = 0$. These assumptions imply that:

(i) $E[\frac{1}{n}\varepsilon'\varepsilon] = \sigma^2_\varepsilon$;

(ii) $E[\frac{1}{n}\bar{\varepsilon}'\bar{\varepsilon}] = \sigma^2_\varepsilon n^{-1}Tr(W'W)$; and

(iii) $E[\frac{1}{n}\bar{\varepsilon}'\varepsilon] = 0$;

where $Tr(.)$ denotes the trace operator. In addition, note from equation (2a) that $u - \lambda\bar{u} = \varepsilon$, and that $\bar{u} - \lambda\bar{\bar{u}} = \bar{\varepsilon}$. Substituting these in (i), (ii), and (iii) we get the following system of equations:

$$\begin{bmatrix} 2n^{-1}E(u'\bar{u}) & -n^{-1}E(\bar{u}'\bar{u}) & 1 \\ 2n^{-1}E(\bar{\bar{u}}'\bar{u}) & -n^{-1}E(\bar{\bar{u}}'\bar{\bar{u}}) & n^{-1}Tr(W'W) \\ n^{-1}E(u'\bar{\bar{u}} + \bar{u}'\bar{u}) & -n^{-1}E(\bar{u}'\bar{\bar{u}}) & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda^2 \\ \sigma^2_\varepsilon \end{bmatrix} - \begin{bmatrix} n^{-1}E(u'u) \\ n^{-1}E(\bar{u}'\bar{u}) \\ n^{-1}E(u'\bar{u}) \end{bmatrix} = 0 \qquad (11)$$

Letting $\alpha = (\lambda, \lambda^2, \sigma^2_\varepsilon)$, and $\gamma_n = n^{-1}(E(u'u), E(\bar{u}'\bar{u}), E(u, \bar{u}))'$, this system of three equations can be written as:

$$\Gamma_n \alpha - \gamma_n = 0, \qquad (12)$$

where Γ is equal to the first bracketed expression in the left-hand side of (11). Assuming that $|\hat{u}_i - u_i| \to 0$ the sample analog of (11) is given by

$$J_n\theta' - j_n = \nu_n(\rho, \sigma_\varepsilon^2); \tag{13}$$

where ν_n can be seen as a 3 x 1 vector of residuals as suggested by these authors, and

$$J_n = \begin{bmatrix} 2n^{-1}\hat{u}'\hat{\bar{u}} & -n^{-1}\hat{\bar{u}}'\hat{\bar{u}} & 1 \\ 2n^{-1}\hat{\bar{\bar{u}}}'\hat{\bar{u}} & -n^{-1}\hat{\bar{\bar{u}}}'\hat{\bar{\bar{u}}} & n^{-1}Tr(W'W) \\ n^{-1}(\hat{u}'\hat{\bar{\bar{u}}} + \hat{\bar{u}}'\hat{\bar{u}}) & -n^{-1}\hat{\bar{u}}'\hat{\bar{\bar{u}}} & 0 \end{bmatrix}; \quad j_n = \begin{bmatrix} n^{-1}\hat{u}'\hat{u} \\ n^{-1}\hat{\bar{u}}'\hat{\bar{u}} \\ n^{-1}\hat{u}'\hat{\bar{u}} \end{bmatrix}:$$

Kelejian and Prucha (1996) suggest the generalized moments estimator for λ and σ_ε^2 as the non-linear least squares estimator

$$(\hat{\lambda}_{NLS;n};\ \hat{\sigma}^2_{\varepsilon NLS;n}) = \operatorname{argmin}\left[\nu_n(\lambda;\sigma_\varepsilon^2)'\,\nu_n(\lambda;\sigma_\varepsilon^2) : \lambda \in [-a;a];\ \sigma_\varepsilon^2 \in [0;b]\right];$$

where $a \geq 1$ and $b < \infty$: They show that $(\hat{\lambda}_{NLS;n};\ \hat{\sigma}^2_{\varepsilon NLS;n}) \to (\lambda;\ \sigma_\varepsilon^2)$ in probability, as $n \to \infty$:[9]

We use these generalized moments estimators to obtain estimates of λ and σ_ε^2, and consequently, of Σ_u as indicated in (4). Therefore, in the derivation of the test statistics, the unconstrained GMM estimator of γ is obtained as described by steps (i) through (v) in section 2.1, but changing (iv) by the Kelejian and Prucha's (1996) GMM estimator, and modifying (v) as is obvious.[10]

2.3. The constrained GMM estimator

A constrained GMM estimator can be defined by optimizing the GMM criterion function subject to the null hypothesis of the form $H_o : a(\gamma) = 0$; where $a(\gamma)$ is a r x 1 vector. The restricted estimator of γ; denoted $\tilde{\gamma}_{GMM}$ will be the solution to

$$Min_\gamma Q_n(\gamma) \text{ subject to } a(\gamma) = 0 \tag{14}$$

We are interested on testing the restriction $\rho = 0$; so the restricted model is

$$y = X\beta + u; \tag{15}$$

$$u = \lambda W u + \psi; \tag{16}$$

which corresponds to a linear model with spatial error autocorrelation. The same Γ matrix in used in both restricted and unrestricted minimization problems.[11] The solution to (14) for this specific case gives

$$\hat{\gamma}_{GMM} = (X'\hat{D}X)^{-1}X'\hat{D}y \tag{17}$$

where D is as before.

3. GMM Test Statistics

Newey and West (1987) devised a counterpart for the GMM estimator of the Wald, Lagrange multiplier and Likelihood ratio test statistics from the maximum likelihood estimation context. The model used in the derivation of the tests implies an orthogonality condition, that in their notation, is given by

$$E[g(z_t; b_o)] = 0; \tag{20}$$

where b_o is a q x 1 vector of parameters, z_t is a p x 1 data vector, and $g(z; b)$ is a r x 1 vector of functions of the data and parameters. Under the assumption that (20) is correct, the sample moment $g_T(b) = \sum_{t=1}^{T} g(z_t; b)/T$ is close to zero when evaluated at $b = b_o$: Solving $\min_b g_T(b) W_T g_T(b)$; where W_T is a positive semi-definite matrix, gives a GMM estimator $\hat{b}_T$ of b_o: They consider a general null hypothesis of the form

$$H_o : a(b) = 0; \tag{21}$$

where $a(b)$ is a sx1 vector. Because the null hypothesis implies a set of restrictions on the parameters, a second "restricted" GMM estimator $\tilde{b}_T$ is obtained by solving the same criterion function, ie., $\min_b g_T(b) W_T g_T(b)$; but subject to (21). The test statistics that use these two estimators are

$$W_G = a(\hat{b})'[\hat{A}\,\Omega^{-1}\hat{A}']^{-1}a(\hat{b}) \tag{22}$$

$$LR_G = [Q_T(\tilde{b}) - Q_T(\hat{b})] \tag{23}$$

$$LM_G = [g_T(\tilde{b})'\,\Omega^{-1}\tilde{G}_T](\tilde{G}_T'\,\Omega^{-1}\tilde{G}_T)^{-1}[\tilde{G}_T'\,\Omega^{-1}g_T(\tilde{b})]; \tag{24}$$

where $A = \partial a(b_o)/\partial b$; $\hat{A} = \partial a(\hat{b}_T)/\partial b$; $g_T(b_o)$ denotes the sample moment conditions, b_o is the vector of parameters, $\hat{G}_T = \partial g_T(\hat{b}_T/\partial b)$; $\tilde{G}_T = \partial g_T(\tilde{b}_T)/\partial b$; $Q_T(b)$ denotes the GMM criterion function, Ω^{-1} is the inverse of the variance-covariance matrix of the moment conditions, and hat and tilde refer to the unrestricted and restricted cases, respectively.

These test statistics have a similar interpretation to their maximum likelihood counterpart. Thus, the W_G can be interpreted as measuring the distance of the unrestricted estimates from its null value. The LM_G as measuring the distance of the gradient of the moment condition evaluated at the constrained estimates from zero, and the LR_G as measuring the distance between the criterion function evaluated at the constrained and unconstrained estimates (see Newey and McFadden (1994) for a graphical representation of this interpretation).

Newey and West (1987) show that these test statistics are asymptotically equivalent and distributed in large samples as a chi-squared random variable with degrees of freedom equal to the number of components of $a(b)$: This result assumes that the test statistics are based on the optimal GMM estimator, for which $\Gamma = \Omega^{-1}$: As noted by these authors, the choice of Γ other than Ω^{-1} implies that a test statistic based on the difference of the criterion functions at the restricted and unrestricted estimates is not asymptotically chi-squared.[12]

3.1 GMM Tests of Spatial Lag Dependence

Kelejian and Robinson's (1993) GMM estimator implies an orthogonality condition as in (20), which in our notation, is written as:

$$H'u = H'y - H'Z\gamma: \tag{25}$$

The GMM test statistics to test the null hypothesis of $H_o : \rho = 0$ based on the unrestricted and restricted GMM estimators that minimize the squared distance of the sample moment conditions in (25), are then given by

$$W_G = \frac{\hat{\rho}^2}{Var\,\hat{\rho}} \tag{26}$$

$$LR_G = \tilde{u}'H[H'\hat{\Sigma}_u H]^{-1}H'\tilde{u} - \hat{u}'H[H'\hat{\Sigma}_u H]^{-1}H'\hat{u}; \tag{27}$$

$$LM_G = \frac{[\tilde{u}'H(H'\hat{\Sigma}_u H)^{-1}H'WY]^2}{[J_\rho]} \tag{28}$$

where the notation is as above,

$$J_\rho = Y'W'H[H'\hat{\Sigma}_u H]^{-1}H'WY - Y'W'H[H'\hat{\Sigma}_u H]^{-1}H'X[J_{-\rho}]^{-1}X'H[H'\hat{\Sigma}_u H]^{-1}H'WY;$$

and $J_{-\rho} = X'H[H'\hat{\Sigma}_u H]^{-1}H'X$. These test statistics are distributed in large samples as a chi-squared random variable with one degree of freedom.[13]

As noted before, the computation of the $Wald_G$ statistic requires the unconstrained GMM estimator of ρ. The statistic evaluates how close is $\hat{\rho}$ to its constrained value of zero. The LM_G requires the constrained estimator of ρ; which corresponds to the estimator in (17). This statistic evaluates how close to zero is the gradient of the criterion function evaluated at $\tilde{\rho}$. The LR_G statistic requires the constrained and unconstrained estimator of ρ. Similarly, this statistic looks at how the values of the GMM criterion function differ when evaluated at the unrestricted and restricted estimators.

Note that all three statistics use the same estimate $\hat{\Gamma}$ of Γ. In addition, observe that the orthogonality conditions are linear in the parameters γ; which implies that $\hat{G}_n = G_n = H'Z$. This means that the derivatives of the moment conditions with respect to the parameter do not depend on the parameters. Proposition 3 in Newey and West (1987) establishes that in a case like this, the LM_G and LR_G test statistics are numerically equivalent, i.e. $LM_G = LR_G$. Therefore, in our case the GMM Lagrange multiplier and the GMM Likelihood ratio test statistics are also numerically equivalent. This numerical equivalence also holds for the $Wald_{GMM}$ when the constraint is also linear (this is their Proposition 4), but does not hold when the test statistics use a di®erent estimator of the asymptotic covariance matrix of the moment conditions. In our case we use the same estimator, and therefore all three tests are numerically equivalent.

4 Small Sample Performance of GMM Tests for Spatial Dependence

We carried out a small Monte Carlo simulation to analyze the ¯nite sample properties of the test statistics presented in the previous section. In this simulation exercise, we compare the ¯nite-sample performance of these tests with an alternative test of spatial lag dependence, which is based on maximum likelihood estimation. The test statistic is the $LM^{¤}_{½}$ robust test for spatial lag dependence of Anselin et al. (1996). This test statistic is easily computable from least squares residuals of the restricted model, and is robust to local misspeci¯cation of spatial error dependence. When the weight matrices in (1) and (2a) are the same, which is our case, this test statistic is

$$LM^{¤}_{½} = \frac{[ų^0Wy=¾^2 \text{ ¡ } ų^0Wų=¾^2]^2}{nĴ_{½:¯} \text{ ¡ } Tr[W^2 + W^0W]}; \qquad (28)$$

where $ų = y$ ¡ $X¯_{¸}$ are the least squares residuals, $¾^2 = ų^0ų=n$; $nĴ_{½:¯} = ¾^2[(WX¯)^0(I$ ¡ $X(X^0X)^{¡1}X^0)(WX¯_{¸}) + Tr[W^2 + W^0W]]^{¡1}$; and Tr denotes the trace operator. This test statistic has been shown to have good small sample properties (see Anselin et al. 1995).

4.1 Experimental Design

The design of the Monte Carlo simulation follows from the experimental design used by Anselin and Florax (1995), which is standard in this literature. We study the small sample performance of the tests under the null hypothesis $H_o : ½ = 0$: We consider values of ½ and ¸ greater than zero. The chosen values of these parameters are ½ = (0:1; 0:3; 0:5; 0:7; 0:9) and ¸ = (0:1; 0:3; 0:5; 0:7; 0:9):[14] The con¯guration of spatial lag and error dependence is determined by 8 di®erent spatial matrices. Six of these matrices are contiguity matrices that are built based on regular lattices using the rook and queen criterion.[15] These matrices are of size 49 (from a square 7x7 grid), 81 (from a 9x9 grid) and 121 (from a 11x11 grid). In addition, we consider two \real" spatial matrices. The ¯rst corresponds to a distance matrix built accordingly to the distance between the largest cities of each state of the United States, and the second matrix is built based on the distance between the cities and townships within the Statistical Metropolitan Area of Boston (SMSA), Massachussets. These two matrices are of size 51 and 110 respectively.

We explore the ¯nite-sample properties of the test under two distributions of the error Ã: We consider errors Ã$_i$ that are i:i:d: N (0; 1); and errors Ã$_i$ that are i:i:d: LogN (0; 1):[16] The matrix X has a column of constants and 2 columns corresponding to two variates drawn independently from a uniform (0; 10) distribution. This matrix remains ¯xed in the experiment. The dependent variable is then generated according to (1').

We consider a total of 400 cases that are determined by the combination of the two chosen distributions for Ã, the ¯ve values of ½ and ¸; and the eigth weight matrices. For each case, we compute the GMM and the LM$^{¤}_{½}$ test statistics.[17] Results are obtained based on 1000 replications. The properties of the tests are evaluated at ® = 0:05 and ® = 0:10 critical values of their asymptotic chi-squared distribution with one degree of freedom. The empirical size, de¯ned by the proportion of incorrect rejections, and the power of the tests, de¯ned by the proportion of correct rejections, are reported. In addition, we report probability values from the Kolmogorov-Smirnov (KS) test of the hypothesis of departures from the asymptotic chi-squared null distribution of the test statistics.

4.2 Monte Carlo Results

Tables 1-4 summarize the results from the Monte Carlo exercises. Tables 1a through 1c present the empirical size of the tests under normal and lognormal errors for each of the weight matrices considered in the experiments. Tables 2a through 3c report the empirical rejections under the alternative hypothesis for ½ = (0:1; 0:3; 0:5; 0:7; 0:9) for the two error distributions, respectively, and Tables 4a-c contain the P-values for the Kolmogorov-Smirnov test statistic.[18]

Using the normal approximation to the binomial distribution, 95% con¯dence intervals for the estimated empirical sizes are [0:0365; 0:0635] and [0:0814; 0:1186] respectively. Results from the experiments show that there are considerable size distortions for both the GMM and the Lagrange Multiplier tests under both distributions. Although the estimated size distortions decrease with the sample size, they increase with the value of the spatial error parameter ¸: For the regular and irregular spatial matrices and under both error distributions, rejection frequencies of the tests are inside the estimated con¯dence intervals only for sample sizes larger than 81 and for values of ¸ <= 0:3: For the \real" spatial matrices the robust test shows better size performance than the spatial GMM tests. Rejection frequencies for the GMM tests are outside the estimated con¯dence intervals for both spatial matrices, while they are inside the estimated con¯dence interval for the robust test for the contiguity matrix of SMSA cities in Massachusetts and for values of ¸ <= 0:5: An encouraging result is that, for given values of ¸; the empirical rejections of the tests approach the estimated con¯dence intervals as we increase the sample size. This is particularly the case for the queen and the irregular weight matrices.

Furthermore, the empirical size of the GMM tests are similar under both error distributions for the rook and irregular weight matrices (mainly when N=110) and tend to be closer to the con¯dence interval for the queen matrices. In contrast, the empirical size of the robust Lagrange Multiplier test tend to be bigger when the error distribution is Lognormal (except for the queen matrices). However, the di®erences decrease as the sample size increases.

Our experiments suggest that the GMM tests and the robust Lagrange Multiplier tests have good size performance for relatively small sample sizes when using regular weight matrices, but for low values of the spatial error parameter (¸).[19] On the other hand, the GMM tests must be used with caution when using irregularly shaped weight matrices, particularly if the sample size is small. Under the speci¯c conditions of our experiment these tests presented considerable size distortions for a sample size of 110 and for low values of ¸:

These results are consistent with those obtained from the Kolmogorov-Smirnov goodness -of-¯t test. P-values of the Kolmogorov-Smirnov statistic indicate rejection of the null hypothesis that the empirical distribution of the GMM and robust Lagrange Multiplier tests is equal to their asymptotic chi-squared distribution for small sample sizes (N = 49 and N = 51) under the regular and irregular weight matrices. On the other hand, the test indicates non-rejection for sample sizes N = 81; N = 110 and N = 121; but when ¸ takes small values.

Overall, our experiments indicate that the small sample distribution of the GMM tests and the robust Lagrange Multiplier tests is well approximated by their asymptotic chi-squared distribution for relatively small sample sizes, but only for small values of the spatial error parameter (see Table 6a-c). When the \degree" of spatial error correlation is high (ie., the parameter takes values bigger than 0.5) the tests present considerable size distortions.

Analyzing the results from the empirical power computations, we ¯nd that, for the regular weight matrices, all test statistics reach empirical rejection frequencies close to one for relatively low values of the spatial lag parameter ½ under both distributions and for all sample sizes. As expected, the rejection frequency approaches one as we increase the sample size. However, we ¯nd important variations across error distributions and across these weight matrices. For example, under normal errors, for the queen matrices and when ¸ is still small, the empirical power of the tests reach one for values of ½ >= 0:3 and sample sizes N = 81 and N = 121: Except when ¸ = 0:9; power is always one for ½ >= 0:5: For the rook matrices, independently of the value of ¸; power reaches one only for values of ½ >= 0:7: This implies that for given sample size and given values of ¸; the tests have better power performance for the queen matrices (relative to the the rook matrices).

On the other hand, for the irregular matrices, empirical rejections are higher for the smaller weight matrix (N = 51) for given values of ½ and ¸. This is in contrast to the results obtained for the regular lattices for which we ¯nd better power performance as sample size increases. In addition, only for values of ½ = 0:9 both the GMM and the robust tests reach empirical rejections close to 0.95. For all other considered values of ½ empirical rejections were below this value and decreased as ¸ increased.

The results also suggest that the tests have better empirical power when the errors are Lognormal (relative to normal errors) for both regular and irregular weight matrices. The experiments show that the GMM and the robust Lagrange Multiplier tests have higher empirical rejection frequencies for all considered values of ½: In addition, power decreases as the value of the spatial error parameter increases but only for very low values of ½: For instance, for the rook and queen matrices empirical rejections reached one for values of ½ >= 0:3 (except for the rook of size N = 49). Similar results are obtained for the irregular matrices.

Unless some size distortion corrections are taken into account, it is not very meaningful (or perhaps it is better to say \fair") to compare the power properties of test statistics that have di®erent size distortions. For example, the experiments show that the robust LM test has a higher rejection frequency than the GMM tests under normal errors for the regular matrices and for values of ½ <= 0:5 (for ½ <= 0:7 and ½ <= 0:9 empirical rejections are similar for given values of ¸). On the other hand, for the irregular matrices we observe the opposite. Empirical rejections tend to be higher for the GMM tests. Because we do not control for size we cannot conclude that GMM tests have better power performance than the robust LM test for irregular matrices or viceversa for regular matrices.

5. Conclusion

In this paper, we have formulated versions of the Wald, the Likelihood ratio and the Lagrange multiplier tests based on GMM estimation to test the null hypothesis of non-spatial lag dependence in a spatial lag model with a spatial autoregressive disturbance term. Based on the work of Kelejian and Robinson (1993) and Kelejian and Prucha (1996), we obtained restricted and unrestricted GMM estimators, which were used in the computation of the test statistics. We also have presented some evidence of the small sample properties of these tests from a Monte Carlo experiment.

These GMM estimators as well as the GMM test statistics are fairly easy to compute. The tests statistics are asymptotically equivalent. In addition, they are numerically equivalent. The numerical equivalence occurs because the derivatives of the moment conditions with respect to the spatial lag parameter do not depend on the parameters and because the imposed constraint is linear. However, this requires that we use the same estimator of the asymptotic covariance matrix of the moment conditions in the computation of the tests.

The results from the Monte Carlo simulations indicate that the GMM and the LM tests, despite their good power properties, present considerable high empirical size distortions, particularly for high values of the spatial error parameter. In other words, these tests tend to reject the null hypothesis of non-lag spatial dependence more frequently than expected when the errors exhibit a high \degree" of spatial correlation. Size distortions decreased as we increased the sample size. However, they increased as the value of ¸ increased.

Overall, the experiments indicate that the small sample distribution of the GMM tests and the robust Lagrange Multiplier tests is well approximated by their asymptotic chi-squared distribution for relatively small sample sizes (their empirical size is close to the size chosen for the test), but for small values of the spatial error parameter. In addition, the empirical power approaches one for low values of the spatial lag parameter under reasonable small sample sizes. However, we ¯nd variations across di®erent spatial matrices for given sample size and values of the spatial error parameter.

The GMM tests for spatial lag dependence in the presence of spatial error dependence constitute a good alternative to available test procedures in spatial econometrics. However, it is important to note that further work in the study of their ¯nite sample properties is needed, particularly for irregular (\real world") spatial structures while paying close attention to the trade-o® between sample size and degree of spatial error correlation.

Endnotes

[1]For other spatial models see Anselin (1988a).

[2]The computational complexities in the maximum likelihood approach arise from the spatial weight matrix and from the Jacobian of the transformation of the log likelihood, which involves the computation of the determinant of a non-triangular matrix. The use of the eigenvalues of the weight matrix (as suggested by Ord (1975)) considerably simplifies the procedure. In addition, a non-linear optimization of a concentrated log likelihood is required to obtain the estimates of the spatial parameters (see also Anselin (1988a)).

[3]A good survey of this literature is Koenker (1982).

[4]See Anselin (1990) for an application of this re-sampling method to the spatial lag model.

[5]We tried to compare the LM_G with its maximum likelihood counterpart LM_{ρ}^{A} (Anselin (1988a)), however, we had problems with the minimization algorithm used to estimate the restricted maximum likelihood estimators. We tried the nlminb Splus function, which allows for the imposition of parameter constraints. However, we were not able to obtain restricted maximum likelihood estimates of the parameters for 1000 replications. It is important to note that there are no studies in the spatial econometrics literature that have analyzed the small sample performance of this test statistic. This test has always been excluded from other Monte Carlo experiments, because it's computation has a "high computational burden" (See Anselin et. al. (1995)).

[6]This is also the case in Kelejian and Robinson's (1993) model, where, $E[My\,u' \mid X] = M(I - \rho M)^{-1}\Sigma_u \neq 0$.

[7]The restriction on the rank of H rules out the case in which $k = 1$ and X is a constant regressor. In addition, note that when X contains a constant regressor and W is row normalized X^* does not include the constant column of X.

[8]We have omitted the presentation of most of the regularity conditions assumed for the consistency of these estimators. For this, and for the proofs of consistency, see Kelejian and Robinson's (1993) paper.

[9]Using an over parameterization of (13), they also suggested an ordinary least squares estimator $\hat{\theta}_{ols} = (\hat{\lambda}, \hat{\lambda}^2, \hat{\sigma}^2_{\chi})$ obtained from regressing j_n on J_n. In our Monte Carlo experiment we use these OLS estimators as the initial values for the non-linear least squares

estimators.

[10]Note that we also use a consistent estimate of -:

[11]A di®erent choice of the weighting matrix ¡ in the unrestricted and restricted criterion functions would imply that a test statistic based on the di®erence of the criterion functions at the restricted and unrestricted estimates will not be asymptotically chi-squared.

[12]The asymptotic properties of these test statistics are given in Newey and West (1987) and the detailed proofs in Newey and West (1985). See also Newey and Macfadden (1994).

[13]A normalization of the criterion function by 1=2 implies equality between the Hessian matrix and the asymptotic variance of the gradient of the criterion function. When GMM estimators are optimal (ie., have minimum asymptotic variance matrix within their class) this information equality property holds and the test statistics will have an asymptotic chi-square distribution, with degrees of freedom equal to the number of components of a(b) (see Newey and Macfadden (1994)).

[14]As in Anselin and Florax (1995), we did not consider negative values of ½: According to Anselin and Rey (1991), complications arise when negative parameter values are considered.

[15]See Anselin (1988a) for a description of these and other spatial weight matrices based on regular and irregular spatial con¯gurations.

[16]As in Anselin (1995), we use Ã = exp(0:69e_i ¡ 1:272) where e_i is a standard normal random variable.

[17]Because the GMM tests are numerically equivalent, we only report results for the LM_G:

[18]Results of the empirical rejections under the alternative hypothesis at ® = 0:10 are not reported, however, they are available upon request.

[19]This result is not surprising for the Lagrange Multiplier test, which is robust only to local misspeci¯cation of the spatial error parameter.

[20]Our results of signi¯cant empirical size distortions of $LM^{¤}_{½}$; which are in contrast to the results obtained in Anselin et al. (1995), may be due to the high value of ¸ chosen in the experiments. In addition, we note that in our Monte Carlo experiments, we are using rook

weight matrices, which have been found to produce relatively poorer test performances to those obtained from other weight matrices (for example weight matrices built based on the queen criterion) for identical sample sizes. In future research, we will explore the finite-sample properties of the tests using other weight matrices, and assuming different values for λ.