Luc Anselin

Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity

Several diagnostics for the assessment of model misspecification due to spatial dependence and spatial heterogeneity are developed as an application of the Lagrange Multiplier principle. The starting point is a general model which incorporates spatially lagged dependent variables, spatial residual autocorrelation and heteroskedasticity. Particular attention is given to tests for spatial residual autocorrelation in the presence of spatially lagged dependent variables and in the presence of heteroskedasticity. The tests are formally derived and illustrated in a number of simple empirical examples.

INTRODUCTION

The nature of the measurement problems associated with data collected for aggregate spatial units (counties, states, census tracts) is such that a variety of sources of misspecification can lead to violations of the standard assumptions underlying regression analysis. Of primary interest to geographers and regional scientists are the problems present in cross-sectional and pooled space-time analysis: spatial dependence and spatial heterogeneity. Spatial dependence may be caused by different kinds of spatial spill-over effects, while heteroskedasticity could easily result from the heterogeneity inherent in the delineation of spatial units and from contextual variation over space.

These problems are familiar and have led to a large body of literature on specification, estimation, and testing in spatial process models. Recent overviews of the relevant issues can be found in, among others, Cliff and Ord (1973, 1981), Bartels and Hordijk (1977), Haining (1978), Brandsma and Ketellapper (1979), Bennett (1979), Hordijk (1979), Griffith (1980, 1983), Anselin (1980, 1984, 1987), Blommestein (1983), Upton and Fingleton (1985).

In this paper, the focus is on two issues which have received much less attention:

Part of the research on which this paper is based is supported through grant SES-8600465 from the National Science Foundation. Research assistance was provided by Serge Rey.

Luc Anselin is associate professor of geography, University of California, Santa Barbara.

Geographical Analysis, Vol. 20, No. 1 (January 1988) © 1988 Ohio State University Press
Submitted 2/87. Revised version accepted 6/87.

1. testing for multiple sources of misspecification in spatial models;
2. testing for spatial dependence when other forms of misspecification are present.

These issues include as special cases the problem of testing for residual spatial autocorrelation in mixed regressive-autoregressive spatial models (i.e., with a spatially lagged dependent variable) and when heteroskedasticity is present (e.g., as the result of spatial contextual variation).

As is well known, the multidirectional nature of spatial dependence often precludes the straightforward use of least-squares regression and limits the applicability of a wide range of standard econometric procedures for many models of interest to geographers. This has necessitated the development of a specialized methodology in spatial econometrics and spatial statistics, largely based on the maximum likelihood principle.[1] Of the three familiar asymptotic testing principles based on maximum likelihood estimation, i.e., the Wald test (W), Likelihood Ratio test (LR), and the Lagrange Multiplier (LM) test, the first two have received most of the attention in spatial econometrics. Even though LM tests are sometimes mentioned, they are seldom carried out and most of the inference in spatial models is still based on the Wald (asymptotic t-test) or Likelihood Ratio tests, e.g., in Cliff and Ord (1973, 1981), Brandsma and Ketellapper (1979), Anselin (1980), and Upton and Fingleton (1985). Exceptions are the LM test for residual spatial autocorrelation in Burridge (1980), and for common factors in Burridge (1981) (see also the discussion in Anselin 1987).

In this paper, an attempt is made to fill this gap in the spatial econometric literature, by applying the Lagrange Multiplier principle to a variety of misspecification problems in spatial analysis. The point of departure is a general specification, which incorporates a spatially lagged dependent variable, residual spatial autocorrelation and heteroskedasticity. The LM approach results in a number of very practical tests, since they are based on the estimation of the model under the null hypothesis, i.e., in its most simple form. In many situations, this will avoid nonlinear procedures and allow the use of ordinary-least-squares results.

The outline of the remainder of the paper is as follows. First, the general model is presented in a formal way and the application of the LM approach discussed in more detail. Next, the actual test statistics are derived for the general case. In the third section, two special cases of particular interest to spatial analysts are considered more closely. The first deals with testing for spatial residual autocorrelation in the presence of spatially lagged dependent variables, the second with testing for spatial residual autocorrelation in the presence of heteroskedasticity. Finally, the tests are illustrated with a simple empirical example and some concluding remarks are formulated.

THE GENERAL APPROACH

Consider the following general specification:

$$y = \rho W_1 y + X\beta + \epsilon$$

$$\epsilon = \lambda W_2 \epsilon + \mu \tag{1}$$

[1] Notable exceptions are the Moran, Geary and Cliff-Ord generalized weight tests for spatial autocorrelation (Cliff and Ord 1973), and the heuristic approaches based on randomization, such as the generalized spatial autocorrelation statistics of Hubert and Golledge (see Hubert, Golledge, and Costanzo 1981; Hubert et al. 1985). A more extensive discussion is given in Anselin (1987).

with $\mu \sim N(0, \Omega)$, and the diagonal elements of the error covariance matrix Ω as

$$\Omega_{ii} = h_i(z\alpha) \qquad h_i > 0 \ .$$

In this specification, β is a $K \times 1$ vector of parameters associated with exogenous (i.e., not lagged dependent) variables X ($N \times K$ matrix),[2] ρ is the coefficient of the spatially lagged dependent variable, and λ is the coefficient in a spatial autoregressive structure for the disturbance ϵ. The disturbance μ is taken to be normally distributed with a general diagonal covariance matrix Ω. The diagonal elements allow for heteroskedasticity as a function of $P+1$ exogenous variables z, which include a constant term. The P parameters α are associated with the nonconstant terms, such that, for α=0, it follows that h $= \sigma^2$ (the classic homoskedastic situation). The two $N \times N$ matrices W_1 and W_2 are standardized spatial weight matrices, associated with a spatial autoregressive process in the dependent variable and the disturbance term respectively. This allows for the two processes to be driven by a different spatial structure, e.g., as suggested in Hordijk (1979). In all, the model has $3+K+P$ unknown parameters, in vector form:

$$\theta \ = [\rho, \beta', \lambda, \sigma^2, \alpha',]' \ . \tag{2}$$

Several familiar model structures result when subvectors of the parameter vector (2) are set to zero. Specifically, the following situations correspond to four traditional spatial autoregressive models commonly discussed in the literature (see e.g., Hordijk 1979; Anselin 1980, 1987; Bivand 1984):

— for ρ=0, λ=0, α=0 (P+2 constraints):

$$y = X\beta + \epsilon \ , \tag{3}$$

that is, the classical linear regression model.

— for λ=0, α=0 (P+1 constraints):

$$y = \rho W_1 y + X\beta + \epsilon \ , \tag{4}$$

that is, the mixed regressive-spatial autoregressive model (which includes the common factor specifications, i.e., with WX, as special cases).

— for ρ=0, α=0 (P+1 constraints):

$$y = X\beta + (I - \lambda W_2)^{-1}\mu \ , \tag{5}$$

that is, the linear regression model with a spatial autoregressive disturbance.

— for α=0 (P constraints):

$$y = \rho W_1 y + X\beta + (I - \lambda W_2)^{-1}\mu \ , \tag{6}$$

[2] Note that the exogenous variables may include spatial lags, such as WX in the common factors approach (see Burridge 1981; Blommestein 1983; Bivand 1984; Anselin 1987). These lagged exogenous variables can be treated in the same way as the other exogenous variables with respect to estimation and testing, although they may form a source of multicolinearity.

that is, the mixed regressive-spatial autoregressive model with a spatial autoregressive disturbance.

Four more specifications can be obtained by allowing heteroskedasticity of a specific form (i.e., a specific $h(z\alpha)$) in models (3)-(6). The different models are constructed by imposing constraints on the more general form (1). In a model specification context, the opposite viewpoint is more relevant. Indeed, the unconstrained models can be viewed as alternative hypotheses to the constrained model, representing a particular form of misspecification. For example, in (3), the null hypothesis (H_0) can be taken as the classical linear regression model, and the types of misspecifications of interest (i.e., the alternative hypothesis, H_1) include the omission of a spatially lagged dependent variable ($\rho \neq 0$), residual spatial autocorrelation ($\lambda \neq 0$), and heteroskedasticity ($\alpha \neq 0$).[3]

In general terms, tests against misspecification of a particular form are carried out by taking the restricted model as H_0 and the more general model as H_1, and by considering the situation as an omitted variable problem.

Formally, for specification (1), the parameter vector (2) can be partitioned as

$$\theta = [\theta_1', \theta_2']' \tag{7}$$

where θ_1 pertains to the parameters included in the null hypothesis,[4] and θ_2 to the remainder. Hypotheses of interest can be expressed as

$$H_0 : \theta_1 = 0$$

$$H_1 : \theta_1 = 0 + \delta ,$$

with $\delta \rightarrow 0$ with $N^{-1/2}$, i.e., local alternatives.

The Lagrange Multiplier or score test statistic is

$$\mathrm{LM} = d'I^{11}d \tag{8}$$

and is distributed as χ^2 with q degrees of freedom (asymptotically), where q corresponds to the dimension of θ_1 in (7), and

$$d = (\partial L/\partial\theta) \tag{9}$$

i.e., the score, evaluated at the null hypothesis, $\theta_1=0$, with L as the log-likelihood for the encompassing model. Furthermore, I^{11} is the partitioned inverse of the information matrix for the encompassing model, evaluated at the null, and partitioned conforming to the partitioning in (7).

The properties of the Lagrange Multiplier test have been discussed extensively in the econometric literature, as they have recently been applied to a wide range of situations, overviewed in, among others, Breusch and Pagan (1980), Engle

[3] Following Breusch and Pagan (1979), the heteroskedasticity is taken to be of prespecified form $h(z\alpha)$. Although this may seem to limit the generality of the approach, it encompasses a wide range of situations of interest in applied work, such as testing for random coefficient variation. It can also be shown to be very similar to the specification-robust test of White (1980), where the z would correspond to all cross-products of the x.

[4] For simplicity of notation and given the particular types of tests which are of interest in this paper, H_0 will be taken as a zero constraint on the parameters. In general, it can represent any linear and nonlinear parameter combination.

(1982, 1984), and Davidson and MacKinnon (1983, 1984). In most situations, the tests reduce to simple expressions in regression residuals, and usually turn out to be functions of the R^2 in a simple auxiliary regression. However, as will be shown in the next section, many of these simplifying results do not hold in the spatial case, due to the multidirectional nature of dependence and its effect on the structure of the Jacobian terms in the likelihood function. The spatial case therefore merits attention in its own right and cannot be found as a direct extension of tests developed for similar situations in time-series data.

A LAGRANGE MULTIPLIER TEST FOR SPATIAL DEPENDENCE AND SPATIAL HETEROGENEITY

In this section, a Lagrange Multiplier test will be developed for the situation where several sources of misspecification are considered. In other words, because spatial dependence (in the form of an omitted spatially lagged variable or spatial autocorrelation in the disturbance term) as well as spatial heterogeneity (in the form of heteroskedasticity) may be present, a test is used which has power against both. If this broad scope for sources of misspecification is unwarranted, the test may have low power (overtesting). On the other hand, if relevant sources of misspecification are ignored when testing for one particular type, the one-directional test will lack robustness and may yield misleading information. This situation may occur in spatial analysis when testing for spatial residual autocorrelation in the presence of heteroskedasticity, or when a spatially lagged dependent variable is omitted.

Formally, model (1) is taken as the alternative, incorporating the three sources of misspecification as the null hypothesis, model (3):

$$H_0 \, : \, \theta_1 = [\rho, \lambda, \alpha']' = 0 \tag{10}$$

where θ_1 is a $(2+P) \times 1$ vector.[5]

In order to operationalize the test statistic (8), the score (9) and relevant partitioning of the information matrix have to be derived for the general model (1), and evaluated at H_0, i.e., with (10) holding.

In the following derivations, the nonlinear form of the model is used, with the simplifying notation of $A = I - \rho W_1$; $B = I - \lambda W_2$. Since $E\,[\mu\mu'] = \Omega$, there exists a vector of homoskedastic random disturbances ν, as $\nu = \Omega^{-1/2}\mu$ and the disturbance in (1) becomes $\epsilon = B^{-1}.\Omega^{1/2}\nu$. After some elementary manipulations, the nonlinear form of (1) results as

$$\Omega^{-1/2}.B.(Ay - X\beta) = \nu \tag{11}$$

where ν is normally distributed with covariance matrix I. In a fairly straightforward way, the corresponding Jacobian can be found as $|\Omega^{-1/2}.B.A| = |\Omega^{-1/2}|.\,|B|.\,|A|$, and the associated log-likelihood function as

$$L = -\,(N/2).\ln(\pi) - (1/2).\ln|\Omega| + \ln|B| + \ln|A| - (1/2)\nu'\nu \tag{12}$$

with

[5] In the notation which follows, the vector α will be considered to contain σ^2 as element α_0.

$$\nu'\nu = (Ay - X\beta)'B'\Omega^{-1}B.(Ay - X\beta) \tag{13}$$

as a sum of squares of appropriately transformed error terms.

Before the actual test can be derived, two intermediate results are necessary: the score vector and the information matrix. The derivations are based on the general approach outlined in detail in, e.g., Ord (1975), Anselin (1980), and Upton and Fingleton (1985), and are fairly tedious but straightforward applications of matrix calculus. In the interest of clarity of exposition, the details are omitted here. The resulting expressions for the elements of the score vector are[6]

$$\partial L/\partial\beta = \nu'(\Omega^{-1/2}.B.X) \ , \tag{14}$$

$$\partial L/\partial\rho = -\ tr\, A^{-1}W_1 + \nu'\Omega^{-1/2}.B.W_1.y \ , \tag{15}$$

$$\partial L/\partial\lambda = -\ tr\, B^{-1}W_2 + \nu'\Omega^{-1/2}.W_2.(Ay - X\beta) \ , \tag{16}$$

and

$$\partial L/\partial\alpha_p = -\ (1/2)\ tr\, \Omega^{-1}.H_p + (1/2)\ \nu'\Omega^{-3/2}.H_p.B.(Ay - X\beta) \tag{17}$$

for $p=1,\ldots,P$.

The elements of the information matrix I, i.e., $-\ E\ [\partial^2 L/\partial\theta\partial\theta']$ are, for the relevant combinations of parameters,

$$I\beta\beta' = X'B'\Omega^{-1}BX \tag{18}$$

$$I\beta\rho = (BX)'\Omega^{-1}.B.W_1.A^{-1}.X\beta \tag{19}$$

$$I\beta\lambda = 0 \tag{20}$$

$$I\beta\alpha' = 0 \tag{21}$$

$$\begin{aligned} I\rho\rho = {} & tr\,(W_1.A^{-1})^2 + tr\,\Omega.(B.W_1.A^{-1}.B^{-1})'\ \Omega^{-1}(B.W_1.A^{-1}.B^{-1}) \\ & + (B.W_1.A^{-1}.X\beta)'\ \Omega^{-1}(B.W_1.A^{-1}.X\beta) \end{aligned} \tag{22}$$

$$I\rho\lambda = tr\,(W_2.B^{-1})'\ \Omega^{-1}.B.W_1.A^{-1}.B^{-1}.\ \Omega + tr\, W_2.W_1.A^{-1}.B^{-1} \tag{23}$$

$$I\rho\alpha_p = tr\,\Omega^{-1}.H_p.B.W_1.A^{-1}.B^{-1} \tag{24}$$

$$I\lambda\lambda = tr\,(W_2.B^{-1})^2 + tr\,\Omega.(W_2.B^{-1})'\ \Omega^{-1}.W_2.B^{-1} \tag{25}$$

$$I\lambda\alpha_p = tr\,\Omega^{-1}.H_p.W_2.B^{-1} \tag{26}$$

$$I\alpha_p\alpha_q = (1/2)\ tr\,\Omega^{-2}.H_p.H_q \ . \tag{27}$$

Expressions (14)-(17) and (18)-(27) need to be evaluated under constrained estimation, i.e., with the parameter values included in the null hypothesis set to zero, and with the other parameters (β and σ^2) set to their ordinary-least-squares estimates (with $\sigma^2 = e'e/N$).

[6] In the notation below, *tr* stands for trace of a matrix, and α_p stands for the pth element of the vector α, with $p=0,1,\ldots,P$. H_p stands for the diagonal matrix with elements $\partial h/\partial\alpha_p$, where h is $h(z\alpha)$, or, explicitly, for $(\partial h/\partial s).z_p$, where $s=z\alpha$ and z_p is the pth element of the z vector.

For the score vector d, only (15) - (17) need to be considered, since (14) equals zero as a result of the conditions for maximum likelihood estimation. Also, under H_0, $A=B=I$ and $\Omega=\sigma^2.I$. With the parameter vector partitioned as $\theta = [\rho\ \lambda\ \alpha' \mid \sigma^2\ \beta']'$, and, after substituting the appropriate values in expressions (15) - (17), it follows that (using straightforward matrix algebra), with e as residual, and β and σ^2 standing for the parameter estimates in the constrained model:

$$d = \begin{bmatrix} e'W_1.y/\sigma^2 \\ e'W_2.e/\sigma^2 \\ (1/2).\sigma^{-2}.[-tr\,H_1 + e'H_1.e/\sigma^2] \\ . \\ . \\ (1/2).\sigma^{-2}.[-tr.H_p + e'H_p e/\sigma^2] \end{bmatrix}$$

The first two elements correspond to N times the regression coefficient of $W_1 y$ on e and $W_2 e$ on e respectively. The latter corresponds to the unstandardized Moran statistic for e (see Cliff and Ord 1981), and is called R_{01} in Anselin (1982) (see also Bivand 1984). The last P elements are the same as in Breusch and Pagan (1979, p. 1289).

The information matrix is partitioned as

$$I = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

or, more specifically, as

$$I = \left[\begin{array}{ccc|cc} I\rho\rho & I\rho\lambda & I\rho\alpha' & I\rho\sigma^2 & I\rho\beta' \\ & I\lambda\lambda & I\lambda\alpha' & I\lambda\sigma^2 & I\lambda\beta' \\ & & I\alpha\alpha' & I\alpha\sigma^2 & I\alpha\beta' \\ \hline & & & I\sigma^2\sigma^2 & I\sigma^2\beta' \\ & & & & I\beta\beta' \end{array}\right] \tag{28}$$

where the indexed I refer to the corresponding expression in (18) - (27), and have to be evaluated under the null. After some matrix algebra, the following expressions result for I_{11}:

$$I\rho\rho = tr\{(W_1)^2 + W_1'W_1\} + \sigma^{-2}.(W_1.X\beta)'(W_1.X\beta) \ ;$$

$$I\rho\lambda = tr\{W_2.W_1 + W_2'W_1\} \ ;$$

$$I\rho\alpha_p = 0 \ ;$$

$$I\lambda\lambda \;=\; tr\,\{(W_2)^2 + W_2'W_2\}\ ;$$

and for each element p,q of $\alpha\alpha'$,

$$I\alpha\alpha' \;=\; (1/2).\sigma^{-4}.tr\,H_p.H_q\ .$$

Using the notation $T_{ij} = tr\,\{W_i.W_j + W_i'W_j\}$, yields

$$I_{11} = \begin{bmatrix} \sigma^{-2}.(W_1X\beta)'(W_1X\beta) + T_{11} & T_{12} & 0 \\ T_{21} & T_{22} & 0 \\ 0 & 0 & I\alpha\alpha' \end{bmatrix}$$

which is block-diagonal in $\{\rho,\lambda\}$ and α.

Also, I_{12} becomes:

$$I_{12} = \begin{bmatrix} 0 & \sigma^{-2}.(X'W_1.X\beta)' \\ 0 & 0 \\ [(1/2).\sigma^{-4}.\ tr\,H_p] & 0 \end{bmatrix}$$

and I_{22} is the familiar maximum likelihood information matrix in the linear regression model:

$$I_{22} = \begin{bmatrix} (N/2).\sigma^{-4} & 0 \\ 0 & \sigma^{-2}.(X'X). \end{bmatrix}$$

The main difference between the spatial model and situations in time series analysis lies in the effect of the spatially lagged dependent variable, i.e., the covariance between ρ and λ (and ρ and β), which results in an information matrix (28) which is not block-diagonal. As a consequence the expression for the inverse I^{11} is less straightforward, though not analytically prohibitive to obtain (due to the sparseness of I_{12} and I_{22}). A series of partitioned inversions result in $I^{11} = [I_{11} - I_{12}.I^{22}.I_{21}]^{-1}$ as the inverse of

$$\begin{bmatrix} \sigma^{-2}.(W_1X\beta)'M.(W_1X\beta) + T_{11} & T_{12} & 0 \\ T_{12} & T_{22} & 0 \\ 0 & 0 & C \end{bmatrix}$$

with $M = I - X(X'X)^{-1}X'$ and a typical p,q element of C as $(1/2).\sigma^{-4}.\ tr\ \{H_p.H_q - (1/N)(H_p)^2\}$.

This matrix is block-diagonal between the spatially dependent and the heteroskedastic components. A partitioned multiplication of $d'I^{11}.d$ yields the Breusch and Pagan (1979) result for the heteroskedastic part, and does not need to be considered any further. In their notation,

$$(1/2).f'Z(Z'Z)^{-1}.Z'f \tag{29}$$

with $f_i = (\sigma^{-2}.e_i{}^2 - 1)$ and Z as the NxP matrix containing the z vectors for each observation.

The analysis of the spatial dependent component is less straightforward. The following simplifying notation is introduced:

$$R_y = e'W_1.y/\sigma^2$$

$$R_e = e'W_2.e/\sigma^2$$

and

$$D = \sigma^{-2}.(W_1X\beta)'M.(W_1X\beta),$$

i.e., σ^{-2} times the residual sum of squares in a regression of $W_1.X\beta$ on X (the spatially lagged predicted values on the original regressors). In addition, setting $E = (D + T_{11}).T_{22} - (T_{12})^2$ straightforward algebra yields, for $d'I^{11}d$ corresponding to the spatial elements,

$$E^{-1}.\{(R_y)^2.T_{22} - 2.R_y.R_e.T_{12} + (R_e)^2.(D + T_{11})\} . \tag{30}$$

This rather awkward expression simplifies greatly for the case where the spatial weight matrices W_1 and W_2 are the same, and consequently, $T_{11} = T_{12} = T_{22} = T = tr\{(W' + W).W\}$. The simplified expression becomes

$$D^{-1}.(R_y - R_e)^2 + (1/T)(R_e)^2 . \tag{31}$$

Note that the second element in (31) is the LM test on spatial residual autocorrelation [distributed asymptotically as $\chi^2(1)$], reported in Burridge (1980).

The full test statistic is the sum of (30) and (29), or (31) and (29). Since the two last elements in this sum correspond to one-directional tests against a specific form of misspecification, one would suspect that the first term corresponds to a one-directional test for the omitted variable Wy. However, this is not the case. It can be shown (using the same general reasoning as above) that the correct statistic, distributed as χ^2 with one degree of freedom, is

$$(R_y)^2.(D + T)^{-1} . \tag{32}$$

The failure of the additivity of the one-directional tests to form the overall test is in contrast to the results in Jarque and Bera (1980) and Bera and Jarque (1982). It is the presence of the spatially lagged dependent variable that results in a complex interaction between ρ and λ. This follows from the structural relationships between a spatial autoregressive process in the dependent variable and a spatial process in the disturbance, e.g., used in the common factor approach.

It should be clear that the expressions (29)-(32) can form the basis for a wide range of tests, one-directional as well as multidirectional. Neither of these necessitates nonlinear estimation, and they can be fairly easily implemented in traditional regression packages that have some matrix manipulation capacities.[7]

[7] A microcomputer implementation is provided in Anselin (1986). Interested readers should contact the author for further information.

In addition, the general framework, using expressions (14)-(27), can be applied to several special cases, either by setting certain combinations of parameters to zero a priori, or by incorporating them in the estimation process. In the latter case, the estimation becomes more complex. Two special cases of particular interest to spatial analysts are examined in the next section.

SPECIAL CASES

In this section, two situations are examined more closely, in which a one-directional test against spatial residual autocorrelation (H_0: $\lambda=0$) is carried out in the presence of other sources of misspecification. The treatment differs from the one-directional tests discussed in the previous section. There, the other parameters are a priori set to zero (i.e., $\rho=0$, $\alpha=0$). Here, some of the parameters are included in the model estimation, i.e., they are assumed to be non-zero. The two cases consist of testing for spatial residual autocorrelation in the presence of a spatially lagged dependent variable (i.e., $\rho \neq 0$, $\alpha = 0$), and in the presence of heteroskedasticity of a prespecified form (i.e., $\alpha \neq 0$, $\rho = 0$).

Testing for Spatial Residual Autocorrelation in the Presence of a Spatially Lagged Dependent Variable

In this first case, the model under consideration is

$$B.(Ay - X\beta) = \mu$$

with $E[\mu\mu'] = \Omega = \sigma^2.I$. The partitioned parameter vector is $\theta = [\lambda \mid \rho\beta'\sigma^2]'$ so that the model under the null hypothesis corresponds to

$$Ay - X\beta = \mu \ . \tag{33}$$

Although the ML estimation of model (33) necessitates nonlinear optimization, it can be shown that the concentrated log-likelihood involves a search over only one parameter (ρ) and thus can be carried out in a fairly straightforward way (for details, see Ord 1975; Anselin 1980, 1986; Bivand 1984; Upton and Fingleton 1985).

Following the same approach as before, the relevant score vector becomes

$$\partial L/\partial\lambda = -\ tr\, B^{-1}.W_2 + \sigma^{-2}.(Ay - X\beta)'B'W_2.(Ay - X\beta)$$

which, under the null hypothesis, with $\lambda = 0$ and $B = I$ (and also, $tr\, W_2 = 0$, by construction), becomes

$$\partial L/\partial\lambda = \sigma^{-2}.(Ay - X\beta)'W_2.(Ay - X\beta) \ . \tag{34}$$

Since $Ay - X\beta$ is the ML residual for (33), expression (34) is similar to R_e in the previous section, i.e., N times a Moran I coefficient, or the coefficient in a regression of $W_2.e$ on e (with e now as the ML residual).

The information matrix for this case follows directly from imposing the restriction $\alpha = 0$ on expressions (19)-(28). To obtain I^{11}, expressions for I_{11}, I_{12}, and I^{22} are needed under the null. In this respect, note that I^{22} is the estimated asymptotic covariance matrix for (33). Also, I_{11} and I_{12} turn out to take very simple forms. The first becomes

$$I_{11} = I\lambda\lambda = tr\{W_2.W_2 + W_2'W_2\} = T_{22}$$

and I_{12} turns out to contain only one nonzero element:

$$I\lambda\rho = tr\{W_2.W_1.A^{-1} + W_2'W_1.A^{-1}\} = T_{21A} \ .$$

Consequently,

$$\begin{aligned} I^{11} &= [I_{11} - I_{12}.I^{22}.I_{21}]^{-1} \\ &= [T_{22} - (T_{21A})^2.\mathrm{var}(\rho)]^{-1} \end{aligned}$$

and the LM statistic for H_0: $\lambda = 0$, is found as

$$(e'W_2.e/\sigma^2)^2.\{T_{22} - (T_{21A})^2.\mathrm{var}(\rho)\}^{-1} \sim \chi^2(1) \tag{35}$$

or, equivalently, as

$$(e'W_2.e/\sigma^2).\{T_{22} - (T_{21A})^2.\mathrm{var}(\rho)\}^{-1/2} \sim N(0,1) \ ,$$

The first term in this expression is N times a Moran I coefficient for the appropriate residuals. The test can be computed fairly easily from the output of the ML estimation of (33), with the additional evaluation of one cross-product and the relevant traces.

Testing for Spatial Residual Autocorrelation in the Presence of Heteroskedasticity

This case concerns the linear regression model with a prespecified form of heteroskedasticity. This can correspond to a particular model assumption, or can occur in more general situations, e.g., when random coefficients or variable coefficients are included.

The model under consideration is $\Omega^{-1/2}.B.(y - X\beta) = \nu$ with $E[\nu\nu'] = I$.

The partitioned parameter vector is $\theta = [\lambda \mid \beta' \alpha']'$ so that the model under the null hypothesis corresponds to

$$\Omega^{-1/2}.(y - X\beta) = \nu \ . \tag{36}$$

Maximum likelihood estimates for the parameters in model (36) can be obtained using one of several iterative techniques (for an overview, see, e.g., Swamy 1971; Raj and Ullah 1981). Alternatively, an explicit nonlinear optimization can be applied (as in Magnus 1978).

As before, the relevant score vector is found as a special case of (17):

$$\partial L/\partial\lambda = -\ tr\,B^{-1}.W_2 + \nu'\Omega^{-1/2}.W_2.(y - X\beta) \ ,$$

which, under the null hypothesis, with $\lambda = 0$ and $B = I$, becomes

$$\partial L/\partial\lambda = (y - X\beta)'\Omega^{-1}.W_2.(y - X\beta) \ ,$$

or, with e as the ML residuals in (36),

$$\partial L/\partial\lambda = e'\Omega^{-1}.W_2.e \ ,$$

the cross-product between e and $W_2.e$ in the metric Ω^{-1} (i.e., weighted by the inverse diagonal elements of Ω).

Also, the appropriate information matrix is found from (19)-(28). Since, under the null, $I\lambda\alpha_p = tr\ \Omega^{-1}.H_p.W_2 = 0$, the resulting information matrix will be block diagonal in λ and $[\beta'\ \alpha']$. It therefore follows that I^{11} can be found as the inverse of $I\lambda\lambda$. With

$$I\lambda\lambda = tr\,\{W_2.W_2 + \Omega.W_2'.\Omega^{-1}.W_2\} = T$$

the Lagrange Multiplier statistic for H_0: $\lambda = 0$, is found as

$$[e'\Omega^{-1}.W_2.e]^2 / T \sim \chi^2(1)$$

or, equivalently, as

$$[e'\Omega^{-1}.W_2.e].\,T^{-1/2} \sim N(0,1) \ . \tag{38}$$

Again, the first term in the expression is similar to a Moran I statistic. The statistic can be evaluated from the output of the ML estimation of (36), with an additional computation of one cross-product and a trace.

EMPIRICAL ILLUSTRATION

The tests developed in the previous sections are illustrated in a simple empirical example. A familiar data set and simple spatial model specification are chosen in order to focus on the comparison of the Lagrange Multiplier tests to more standard approaches. The data set on Irish agricultural consumption, originally studied by O'Sullivan, and described in detail in Cliff and Ord (1981, pp. 208–210, and p. 230) is selected because of its widespread use as a standard of reference (see, e.g., the empirical examples in Burridge 1981, Blommestein 1983, and Bivand 1984). The specific focus in this section will be on tests for the presence of spatial dependence, either in the form of spatial residual correlation, or in the form of an omitted spatially lagged dependent variable, or both. In particular, the LM tests will be compared to the inferences drawn by Burridge (1981), using a generalized spatial weight matrix, and those reported in Bivand (1984), where a standardized binary contiguity matrix is used.

In all, five model specifications are considered. Model 1 consists of a simple regression, which relates the percentage of self-consumption of agricultural output to an arterial road network accessibility index (see the description in Cliff and Ord 1981 for details):

$$y = \alpha + \beta_1 x \ .$$

Models 2 and 3 introduce a spatially lagged dependent variable, respectively for a binary weight matrix and a generalized weight matrix (each standardized such that the row elements sum to one):

$$y = \rho.W.y + \alpha + \beta_1 x \ .$$

Models 4 and 5 include a common factor hypothesis, i.e., the spatial autoregressive form of the model which results when spatial residual correlation is present in Model 1, again for a binary as well as a generalized weight matrix:

$$y = \rho.W.y + \alpha + \beta_1 x + \beta_2 Wx ,$$

where the constraint $\rho.\beta_1 = - \beta_2$ should be satisfied.

The maximum likelihood estimation results (using the unconstrained form for Models 4 and 5), and some measures of goodness of fit, such as the estimated residual variance, the maximized log-likelihood, Akaike Information Criterion (AIC), and the Schwartz Criterion (SC), are reported in Table 1 (for a detailed discussion of these measures, see Anselin 1987). The reader familiar with the literature will note slight differences with the results presented by Burridge and Bivand, particularly for the estimated standard errors and measures of fit. For example, the likelihood and resulting AIC given by Bivand pertain to the concentrated log-likelihood function, and thus differ from the values in Table 1 by a constant term. More important differences result from the lack of a common practice to estimate the asymptotic variance matrix for the parameters. The variance matrix is often estimated by its sample equivalent (and not based on the explicit derivation of expected values of the information matrix, which is used here), and simplifying assumptions may or may not be introduced for the traces of the matrices involved. In this respect, setting $tr\ (W.A^{-1})^2$ equal to $tr\ (W.A^{-1})'(W.A^{-1})$ (as suggested in Bivand 1984, p. 32) will result in inaccuracies. For example, for Model 2, these values are respectively 28.70 and 32.92.

TABLE 1
Estimation Results

	Model 1	Model 2 (binary)	Model 3 (general)	Model 4 (binary)	Model 5 (general)
ρ		0.731	0.646	0.569	0.622
		(0.115)	(0.126)	(0.174)	(0.156)
α	− 8.448	− 6.249	− 6.713	− 12.438	− 7.617
	(3.193)	(2.007)	(2.081)	(4.617)	(4.100)
β_1x100	0.527	0.239	0.276	0.207	0.270
	(0.071)	(0.054)	(0.058)	(0.070)	(0.072)
β_2x100				0.232	0.035
				(0.166)	(0.154)
R^2	0.697	0.874	0.863	0.871	0.862
σ^2	12.537	5.255	5.674	5.328	5.738
L	− 60.755	− 51.653	− 52.320	− 50.803	− 52.301
AIC	125.509	109.306	110.640	109.606	112.603
SC	128.026	113.080	114.414	114.639	117.635

Estimated standard error in parenthesis. R^2 is the unadjusted R^2 for Model 1 (OLS), and the squared correlation between predicted values and actual values for the other models (ML).

In terms of model discrimination, the results in Table 1 show Model 2, i.e., the simple spatial autoregressive form with the binary weight matrix, with the best corrected fit (lowest AIC or SC). Based on these criteria, there seems to be no basis to expect spatial residual autocorrelation, which would show Models 4 or 5 as best. Rather, the effect of an omitted spatial lag is more likely.

In terms of diagnostics, the different tests for multiple sources of misspecification for Model 1 are considered first. These include a test for an omitted spatially lagged dependent variable S, in combination with spatial residual autocorrelation R and heteroskedasticity H (SRH), as well as the tests against two sources of misspecification, SR and RH. The results in Table 2 show clear evidence of misspecification, particularly in the form of spatial dependence (SR). In order to assess which form of spatial dependence is more appropriate, the one-directional tests can be used. It should be noted that in the process of carrying out multiple comparisons, as in this simple illustration, the critical levels used as the basis for rejection of the respective null hypotheses should be adjusted. For example, the use of Bonferroni bounds would roughly consist of dividing the overall level which is to be attained by the number of comparisons (see Savin 1980 for a discussion). Since the emphasis here is on simple illustration, this is not pursued in further detail.[8]

TABLE 2

Lagrange Multiplier Tests for Multiple Misspecifications

	Binary Weights	General Weights
LM_{SRH}	18.237	17.270
	(0.003)	(0.004)
LM_{SR}	16.058	15.091
	(0.000)	(0.001)
LM_{RH}	7.420	9.907
	(0.115)	(0.042)

Probability levels in parentheses; S: spatially lagged dependent variable; R: residual spatial autocorrelation; H: heteroskedasticity. All statistics are χ^2, SRH with five degrees of freedom, SR and RH with two.

One-directional test results for spatial residual autocorrelation and for an omitted spatially lagged dependent variable are reported in Tables 3 and 4. For spatial residual autocorrelation, five tests are compared: the Moran I statistic under the assumption of normality (Cliff and Ord 1981, pp. 202–3), and under the assumption of randomization (Cliff and Ord 1981, p. 21), the Wald statistic (i.e., the asymptotic t-test on ρ in Models 4 and 5), the LR test (i.e., the LR test on ρ in Models 4 and 5), and the LM test. To facilitate comparison, all statistics are expressed as χ^2 variates. Here again, the reader familiar with the literature will note slight differences with the results for the Moran statistics reported in Cliff and Ord (1981, p. 210). The reason for these differences lies in the propagation of rounding errors in the calculation of the moments for the residuals. For example, a replication of the analysis using the residuals given in Cliff and Ord (1981, p. 208) yields a Moran I statistic of 0.436 and a normal z-value of 3.558, whereas a computation carried out at a higher level of precision (i.e., with the residuals kept in memory) yields a Moran I of 0.429 and a z-value of 3.514. Although these differences are not large enough to change a qualitative interpretation of the results, in particular circumstances they may give conflicting indications.[9]

All test results show evidence of spatial residual correlation, and at a higher level of significance for the generalized weight matrix. The W, LR, and LM tests show the usual inequality: LM $\leqslant$ LR $\leqslant$ W. The magnitudes for the Moran tests are between the values obtained for W and LR, with higher values under the normality assumption.

In comparison, the W, LR (using the estimates for Models 2 and 3), and LM

[8] A one-directional test against random coefficient variation yielded a LM value of 2.179, which is clearly nonsignificant (p=0.54). Therefore, heteroskedasticity is not further considered in this example.

[9] It should be noted that all results reported here were obtained using the software in Anselin (1986) on the IBM PC-AT with a numerical co-processor, with all internal calculations carried out with an 80-bit precision (or about 19-digit precision).

TABLE 3
Tests for Spatial Residual Autocorrelation

	Binary Weights	General Weights
Moran (N)	9.880 (0.002)	12.351 (0.000)
Moran (R)	7.833 (0.005)	10.438 (0.001)
Wald	10.706 (0.001)	15.887 (0.000)
LR	5.917 (0.015)	8.246 (0.004)
LM	5.241 (0.022)	7.728 (0.005)

Probability levels in parentheses. All statistics are distributed as $\chi^2(1)$.

TABLE 4
Tests for an Omitted Spatially Lagged Dependent Variable

	Binary Weights	General Weights
Wald	40.722 (0.000)	26.171 (0.000)
LR	18.204 (0.000)	16.869 (0.000)
LM	14.559 (0.000)	14.700 (0.000)

Probability levels in parentheses. All statistics are distributed as $\chi^2(1)$.

tests for an omitted spatially lagged dependent variable, shown in Table 4, give much larger values, and a clear indication of misspecification.

Finally, when a spatially lagged dependent variable is included, two tests on spatial residual autocorrelation are compared in Table 5, the traditional Moran statistic under the assumption of randomization, and the LM test developed in this paper. The values obtained are fairly close for the binary weights, but clearly different for the general weights, with the Moran statistic less likely to reject the null hypothesis. In fact, there is clearly no indication of persistent residual autocorrelation in the case of binary weights (giving a formal basis for the results of Bivand), and very little indication for autocorrelation with the generalized weights (in contrast to the suggestion made in Burridge, based on an approximate z-value for the Moran test). In this context, the higher value for the LM test in Model 5 may be an indication of spurious correlation due to the use of a misspecified weight matrix (since model discrimination criteria show a binary weight matrix as more appropriate).

TABLE 5
Test on Spatial Residual Autocorrelation in the Presence of a Spatially Lagged Dependent Variable

	Binary Weights	General Weights
Models 2,3		
Moran (R)	0.027 (0.869)	0.334(−) (0.563)
LM	0.048 (0.827)	1.564 (0.211)
Models 4,5		
Moran (R)	0.440 (0.507)	0.285(−) (0.593)
LM	0.579 (0.447)	3.592 (0.058)

Probability levels in parentheses. All statistics are distributed as $\chi^2(1)$.

CONCLUDING REMARKS

The various LM statistics outlined in this paper provide a means to test for the presence of a variety of combinations of possible sources of misspecification. Several aspects are very attractive: the tests are computed on the basis of the simpler specification; the tests provide indications of the presence of multiple kinds of misspecification; and the tests have clear asymptotic properties. Therefore, they provide a useful alternative to ad hoc procedures, such as the uncritical use of the Moran *I* statistic under a randomization assumption.

The formulations for testing for spatial residual autocorrelation in the presence of either spatially lagged dependent variables or heteroskedasticity turn out to be closely related to the Moran *I* statistic. In each case, a basic element of the test consists of a (weighted) regression of the lagged residuals on the originals. The main difference with the Moran *I* lies in the scaling factor to be applied (a complex expression in traces of the weight matrices) and in the asymptotic distribution.

As is the case with all tests based on asymptotic properties, the performance in small samples is not necessarily satisfactory, in particular when the assumptions underlying the likelihood function may not be satisfied. This remains to be investigated. In this respect, it should be noted that a direct extension of properties obtained in standard econometric practice may not be appropriate in situations of interest to geographers and regional scientists. As the derivations in this paper show, the additional complexity introduced by the multidirectional spatial dependence merits special attention.

LITERATURE CITED

Anselin, L. (1980). *Estimation Methods for Spatial Autoregressive Structures.* Regional Science Dissertation and Monograph Series #8. Ithaca, N.Y.: Cornell University.

——— (1982). "A Note on Small Sample Properties of Estimators in a First-order Spatial Autoregressive model." *Environment and Planning* A 14, 1023–30.

——— (1984). "Specification Tests on the Structure of Interaction in Spatial Econometric Models." *Papers, Regional Science Association* 54, 165–82.

——— (1986). "Estimation and Model Validation of Spatial Econometric Models Using the GAUSS Microcomputer Statistical Software." Working paper, Department of Geography, University of California, Santa Barbara.

——— (1987). "Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches." *International Regional Science Review* 11.

Bartels, C. P. A., and L. Hordijk (1977). "On the Power of the Generalized Moran Contiguity Coefficient in Testing for Spatial Autocorrelation among Regression Disturbances." *Regional Science and Urban Economics* 7, 83–101.

Bennett, R. (1979). *Spatial Time Series.* London: Pion.

Bera, A. K., and C. M. Jarque (1982). "Model Specification Tests, a Simultaneous Approach." *Journal of Econometrics* 20, 59–82.

Bivand, R. S. (1984). "Regression Modeling with Spatial Dependence: An Application of Some Class Selection and Estimation Methods." *Geographical Analysis* 16, 25–37.

Blommestein, H. (1983). "Specification and Estimation of Spatial Econometric Models. A Discussion of Alternative Strategies for Spatial Economic Modelling." *Regional Science and Urban Economics* 13, 251–70.

Brandsma, A. S., and R. H. Ketellapper (1979). "Further Evidence on Alternative Procedures for Testing of Spatial Autocorrelation among Regression Disturbances." In *Exploratory and Explanatory Statistical Analysis of Spatial Data,* edited by C.P.A. Bartels and R.H. Ketellapper, pp. 113–36. Boston: Martinus Nijhoff.

Breusch, T. S., and A. R. Pagan (1979). "A Simple Test for Heteroskedasticity and Random Coefficient Variation." *Econometrica* 47, 1287–94.

——— (1980). "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." *Review of Economic Studies* 67, 239–53.

Burridge, P. (1980). "On the Cliff-Ord Test for Spatial Correlation." *Journal of the Royal Statistical Society B* 42, 107–8.

——— (1981). "Testing for a Common Factor in a Spatial Autoregression Model." *Environment and Planning* A 13, 795-800.

Cliff, A. D., and J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion.

——— (1981). *Spatial Processes: Models and Applications*. London: Pion.

Davidson, R., and J. G. MacKinnon (1983). "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test." *Economics Letters* 12, 269-75.

——— (1984). "Model Specification Tests Based on Artificial Linear Regressions." *International Economic Review* 25, 485-502.

Engle, R. F. (1982). "A General Approach to Lagrange Multiplier Model Diagnostics." *Journal of Econometrics* 20, 83-104.

——— (1984). "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." In *Handbook of Econometrics*, Vol. II, edited by Z. Griliches and M. D. Intriligator, pp. 775-826. Amsterdam: North Holland.

Griffith, D. A. (1980). "Towards a Theory of Spatial Statistics." *Geographical Analysis* 12, 325-39.

——— (1983). "The Boundary Value Problem in Spatial Statistical Analysis." *Journal of Regional Science* 23, 377-87.

Haining, R. (1978). *Specification and Estimation Problems in Models of Spatial Dependence*. Studies in Geography no. 24. Evanston, Ill.: Northwestern University.

Hordijk, L. (1979). "Problems in Estimating Econometric Relations in Space." *Papers, Regional Science Association* 42, 99-115.

Hubert, L. J., R. G. Golledge, and C. Costanzo (1981). "Generalized Procedures for Evaluating Spatial Autocorrelation." *Geographical Analysis* 13, 224-33.

Hubert, L. J., R. G. Golledge, C. M. Costanzo, and N. Gale (1985). "Measuring Association between Spatially Defined Variables: An Alternative Procedure." *Geographical Analysis* 17, 36-46.

Jarque, C. M., and A. K. Bera (1980). "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals." *Economics Letters* 6, 255-9.

Magnus, J. R. (1978). "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix." *Journal of Econometrics* 7, 281-312.

Ord, J. K. (1975). "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70, 120-6.

Raj, B. and A. Ullah (1981). *Econometrics, a Varying Coefficients Approach*. New York: St. Martins Press.

Savin, N. E. (1980). "The Bonferroni and Scheffe Multiple Comparison Procedures." *Review of Economic Studies* 47, 255-74.

Swamy, P. A. V. B. (1971). *Statistical Inference in Random Coefficient Regression Models*. New York: Springer Verlag.

Upton, G. J. G., and B. Fingleton (1985). *Spatial Data Analysis by Example*. New York: Wiley.

White, H. (1980). "A Heteroskedastic-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48, 817-38.