1997

Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors

Luc Anselin

Harry H. Kelejian

Digital Commons Citation

TESTING FOR SPATIAL ERROR AUTOCORRELATION IN THE PRESENCE OF ENDOGENOUS REGRESSORS

LUC ANSELIN

Regional Research Institute and Department of Economics, West Virginia University, Morgantown, WV 26506-6825 USA (lanselin@wvu.edu)

HARRY H. KELEJIAN

Department of Economics, University of Maryland, College Park, MD 20742 USA (kelejian@econ.umd.edu)

This paper examines the properties of Moran's *I* test for spatial error autocorrelation when endogenous variables are included in the regression specification and estimation is carried out by means of instrumental variables procedures (such as two-stage least squares). The asymptotic distribution of the statistic is formally derived in a general model that encompasses endogeneity due to system feedbacks as well as spatial interaction (in the form of spatially lagged dependent variables). The small-sample performance of the test is assessed in a series of Monte Carlo simulation experiments, and the test is compared to a number of ad hoc approaches that have been suggested in the literature. While some of these ad hoc procedures perform surprisingly well, the new test is the only acceptable one in the presence of spatially lagged dependent variables. The test is straightforward to compute and should become part of routine specification testing of models with endogeneity that are estimated for cross-sectional data.

INTRODUCTION

In empirical regional and urban economic research, many models are estimated with cross-sectional observations on aggregate spatial units such as counties or census tracts. Such data sets are likely to exhibit a lack of independence in the form of spatial autocorrelation. This is increasingly acknowledged in applied econometric work and approached by implementing diagnostic tests for the presence of spatial autocorrelation (for reviews, see Anselin 1988b, 1988c; Anselin and Bera 1997). In the typical case, the regression model is linear, the regressors are exogenous, and the diagnostic is formulated in terms of the ordinary least squares residuals. The most popular test applied in this context is Moran's *I* test for spatial autocorrelation (Moran 1950), which was generalized to regression residuals by Cliff and Ord (1972) (see also Cliff and Ord 1973, 1981). Moran's *I* test is similar in structure to the familiar Durbin-Watson test for serial correlation in the time domain and shares many of its properties. In particular, it is locally best invariant (King 1981) and its exact distribution depends on the explanatory variables in the model and the dependence structure embodied in the alternative hypothesis (King 1987; Sen 1990; Tiefelsdorf and Boots 1995). Assuming

normality for the errors and fixed regressors, Moran's I is an exact test. Its distribution can be computed in small samples by means of numerical integration following the procedure suggested in Tiefelsdorf and Boots (1995). However, this is typically not done in practice. Instead, an asymptotic approximation is used that is based on the expected value and variance of the statistic under the null hypothesis of independence and normality (Cliff and Ord 1973; Sen 1976; see also Terui and Kikuchi 1994). In most cases, this approximation is quite satisfactory and the test has demonstrated superior power in several Monte Carlo simulation studies (e.g., Hordijk 1974; Bartels and Hordijk 1977; Anselin and Rey 1991; Anselin and Florax 1995).

Moran's I test is a powerful misspecification test. However, it does not provide an indication of the nature of the spatial process that causes the autocorrelation, specifically whether it is due to an error process or an omitted spatially lagged dependent variable. Other tests, based on the Lagrange multiplier principle (e.g., Burridge 1980; Anselin 1988a, 1988c; Anselin et al. 1996) are more direct in this respect, but they are asymptotic in nature and depend on the assumption of normal error terms. Also, when the misspecification is error autocorrelation, Moran's I consistently achieves higher power in small samples than the corresponding Lagrange multiplier (LM) test. On the other hand, LM tests are easier to implement in practice since they do not involve moment calculations as is the case for Moran's I (see Anselin and Hudak 1992).

Moran's I and the LM tests are typically considered in a context where all explanatory variables are exogenous. In the practice of empirical regional research however, the potential endogeneity of variables in a model is increasingly acknowledged (e.g., Kelejian and Robinson 1993; Holtz-Eakin 1994). Little is known about the properties of tests for spatial error autocorrelation when such endogeneity is present.

The objectives of this paper are twofold. First, from a theoretical perspective, the asymptotic distribution of Moran's I statistic is formally derived based on residuals that are obtained from an instrumental variables (IV) procedure such as two-stage least squares (2SLS) in a general model that encompasses endogeneity due to system feedbacks as well as spatial interaction (spatially lagged dependent variables). In the process, the set of assumptions needed to rigorously obtain the asymptotic properties is stressed. Second, focusing on situations encountered in practice, this paper assesses the small-sample performance of the suggested test, as well as a variation that is ad hoc in nature but has been considered in practice (e.g., Holtz-Eakin 1994). This variation boils down to the application of the moments derived in Cliff and Ord (1972) to IV residuals. This paper also empirically investigates the effect of endogeneity on the standard OLS-based Moran's I and LM statistics, i.e., when the endogeneity is ignored in the estimation. This is the first study to address this aspect.

The remainder of the paper consists of six sections. First, the formal framework for the derivations in terms of the general model specification is outlined.

Next, the necessary assumptions are spelled out. This is followed by a discussion of instrumental variable estimation and the asymptotic distribution of the statistic for the general model as well as for a number of special cases. Next, the various diagnostics are evaluated empirically in a series of Monte Carlo simulation experiments. The paper closes with some concluding comments and practical recommendations. Formal proofs are relegated to the appendices.

THE GENERAL MODEL

Consider a regression model that contains exogenous explanatory variables, as well as endogenous variables and spatially lagged dependent variables:

$$y_1 = X_1\beta + \rho W_N^* y_{1*} + Y_2\alpha + \varepsilon, \tag{1}$$

or, more concisely,

$$y_1 = Z_1\delta + \varepsilon \tag{2}$$

where $Z_1 = (X_1, W_N^* y_{1*}, Y_2)$ and $\delta' = (\beta', \rho, \alpha')$. The notation is as follows: y_1 is the N by 1 vector of observations on the dependent variable; X_1 is an N by k_1 matrix of observations on k_1 exogenous variables with β as the corresponding k_1 by 1 vector of parameters; W_N^* is an N by $(N+J)$ spatial weights matrix whose elements are known constants; y_{1*} is the $(N+J)$ by 1 vector $y_{1*} = (y_1', y_{11}')$, with y_{11} as a J by 1 vector of observations on J values of the dependent variable that correspond to border locations; ρ is a scalar spatial autoregressive parameter; Y_2 is an N by g matrix of observations on g endogenous regressors, with α as the corresponding g by 1 parameter vector; and ε is the N by 1 vector of error terms. Throughout the paper, the subscript of the spatial weights matrix will be used to indicate its dependence on the sample size (N).

The weights matrix W_N^* can be partitioned to make it conformable to the partition of y_{1*}, such that $W_N^* = (W_N, \tilde{W}_N)$, where W_N is N by N, and $\tilde{W}_N$ is N by J. Given this notation, equation (1) can also be written as

$$y_1 = X_1\beta + \rho(W_N y_1 + \tilde{W}_N y_{11}) + Y_2\alpha + \varepsilon. \tag{3}$$

In practice, the border effects are typically ignored. This constitutes a special case of the general model, with $J = 0$ and no term of the form $\tilde{W}_N y_{11}$ in equation (3). Border observations are explicitly considered in order to ensure a correct model specification for each sample size N. Clearly, their omission would lead to measurement error problems (see, e.g., Griffith 1983). This is important in the derivation of the asymptotic properties and is further discussed below.

The elements of y_1 and y_{11} are not necessarily assumed to form a complete system. In other words, it is possible for the elements of y_{11} to interact among themselves and to interact with the elements in y_1, as well as with other neighboring elements (not part of y_{11}). In what follows, the feedback between y_{11} and y_1 is explicitly accounted for.

In addition to endogeneity in the form of a spatially lagged dependent variable, the general model also involves system feedbacks contained in the matrix of endogenous variables Y_2. This implies the existence of other unspecified equations in which the elements of Y_2 are partially determined by the elements of y_1 along with other variables. For this reason, the elements of Y_2 cannot be assumed to be unrelated to the elements of ε.

Finally, let X_2 be an N by k_2 matrix of available observations on k_2 additional exogenous variables. These are a subset of the exogenous variables appearing in the equations that determine the elements of Y_2, but they do not appear as exogenous variables in (1). The exact specification of the equations in which these excluded exogenous variables appear does not need to be known.

An illustration of the type of specification considered here can be found in the study of regional productivity measures of Kelejian and Robinson (1997). For example, y_{1i} (as the i-th element in y_1) could represent a productivity measure for the i-th region. Then, y_{1i} may be explained in terms of factors specific to the i-th region, such as weather (exogenous variables); in terms of a similar productivity measure for neighboring areas, due to production externalities or technology borrowing (a spatially lagged dependent variable); and in terms of factor inputs in the i-th region that, in turn, are partially dependent upon the productivity in region i (system feedback variables).

STATISTICAL ASSUMPTIONS

A detailed outline of the set of assumptions on which the derivation of the asymptotic properties will be based is now presented. First, some additional notation is needed. In general, denote the i,j-th element, the i-th row, and the j-th column of a matrix A as, respectively, a_{ij}, $a_{i.}$, and $a_{.j}$. Each assumption is now considered in turn.[2]

> ASSUMPTION 1: w_{Nij} *does not depend upon* N, *so that* $w_{Nij} \equiv w_{ij} \, \forall \, N$. *Furthermore,* $|w_{ij}| \leq c_w < \infty$, *for all* i, j = 1,..., N, *and* N > 1, *with* c_w *as a finite upper bound on the magnitude of the spatial weights.*

This assumption requires that the elements of W_N are not functions of the sample size. Essentially, this implies that the large-sample analysis is conditional upon a given sequence of weights matrices. It also ensures that a sample of size $N + 1$ corresponds to all the units belonging to the sample of size N, plus one additional one. This is required for the asymptotics (as $N \to \infty$) to work properly. This assumption would be violated if, for example, for each N the sample were to correspond to a random selection of N units from the population of all possible units. If this were the case, the sample of size $N + 1$ need not contain all, or even any, of the units contained in the sample of size N.

---

[2]Note that the assumptions below that directly involve a weights matrix pertain to W_N. The weights matrix W_N^* is involved indirectly in assumptions 8 and 9 through Z_1 from equation (2).

More importantly, assumption 1 rules out situations in which the elements of the weights matrix are re-specified as additional observations are added to the sample. For example, this is the case when the number of neighbors for each observation is allowed to change as the sample increases in size. This would happen in the familiar row-standardized case, where each nonzero weight in the i-th row of W_N equals $1/d_i$, with d_i as the number of neighbors for observation i. This problem is avoided by specifying the more general W_N^* as the weights matrix in (1), such that d_i is the number of nonzero elements in the i-th row of W_N^*, which will exceed the number of nonzero elements in W_N when border locations are neighbors of observation i. However, the $1/d_i$ derived from the row elements in W_N^* will not change as new observations are added to the sample (i.e., all the neighbors of observation i are already taken into account), so that assumption 1 is not violated. Consequently, the weights structure of the model is correctly specified for each sample size.

Assumption 1 may be unduly restrictive, but the technical apparatus required to treat more complex ways in which the elements of the weights matrix are affected by increasing sample size are beyond the scope of the current paper. For an in-depth treatment of the difficulties encountered in formalizing asymptotics for spatial samples, see, e.g., Cressie (1993: 100–101) and Kelejian and Prucha (1995).

ASSUMPTION 2: *Let* r_N *be the number of rows in* W_N *that consist entirely of zero elements. Then,* $0 \leq r_N \leq \lambda_1$ *for all N, where* λ_1 *is a finite constant.*

This imposes a bound on the number of observations that are unconnected, or "islands," in the sense that they have no neighbors specified in W_N. Hence, assumption 2 ensures that, at most, λ_1 elements of y_1 are unrelated to other elements of y_1 (through W_N), even as $N \to \infty$. If this were not the case, a situation could arise where the sample would grow in size by adding unconnected observations. Since Moran's I statistic is based on sums of cross products of residuals for neighboring locations (as determined by the nonzero elements in W_N), the addition of unconnected observations to the sample would only add zero to this sum. This violates a fundamental assumption of the central limit theorem needed to determine the large-sample distribution of the test, since central limit theorems are based on sums of an infinite number of terms. Assumption 2 will be satisfied in most situations where there is no abrupt change in the structure of spatial dependence that is reflected in the weights matrix. For example, this will be the case for regular lattice structures.

ASSUMPTION 3: $w_{ij} \neq 0$ *if and only if* $w_{ji} \neq 0$; *however,* w_{ij} *need not equal* w_{ji}.

This requires a spatial weights matrix whose binary (boolean) form is symmetric. In other words, if unit i is viewed as neighboring unit j, then the reverse should be the case as well, although the strength of potential interaction (the value of the spatial weights) need not be identical. In practice, this assumption

will be satisfied by most spatial econometric model specifications, although it precludes one-directional forms of spatial autocorrelation such as may be present in network flows. Since the latter can be treated more directly as an extension of the time series case, this is not very restrictive.

ASSUMPTION 4: *The sequence of spatial weights matrices* W_N *is such that its elements* w_{Nij} *satisfy the following constraints:*

(a) $w_{Ni,i+j} = 0$ *and* $w_{Ni.} w'_{N(i+j)} = 0$, *for all* $\lambda_2 < j \le N - i$, $i = 1, \ldots$
   $N - \lambda_2 - 1$, *and* $N > 1$, *where* λ_2 *is a finite constant,*

(b) $w_{Nii} = 0$, $\forall i$,

(c) $\displaystyle \lim_{N \to \infty} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Nij}/N = s_1$, *with* s_1 *as a finite constant, and*

(d) $\displaystyle \lim_{N \to \infty} \text{tr}[(W_N + W_N')(W_N + W_N')]/N = s_2$, *with* s_2 *as a finite constant.*

The main point of assumption 4 is that the spatial weights matrices should be such that, regardless of sample size, each element of y_1 is related to at most λ_2 other elements of y_1. In other words, and in conjunction with assumption 3, this implies that there are at most λ_2 nonzero elements in each row and column of W_N. This can be interpreted as a bound on the extent of spatial interaction (dependence) allowed in the sample (see, e.g., the discussion of mixing conditions in Anselin 1988c, Ch. 5). Furthermore, the spatial sample is assumed to grow in such a manner that units that are "sufficiently" apart from i are not neighbors of i, nor do they have neighbors in common (i.e., they are not second-order neighbors). More precisely, this implies a form of spatial ordering of the observations, such that observations that are separated by λ_2 units (a finite constant) are neither first- nor second-order contiguous. This does not pertain to the way data would be collected in practice, but is a formal mathematical (conceptual) requirement imposed to ensure that the spatial structure (the structure of W_N) does not change as the sample increases to infinity. It is satisfied by all spatial layouts in which a form of distance decay (in a proper metric, see, e.g., Weibull 1976) operates to limit the extent of interaction. As mentioned earlier, in the approach used here, all neighbors of i are contained in an augmented weights matrix W_N^*.

Assumptions 1 and 4 may be seen as unduly restrictive because they impose a given ordering of the data. However, it is important to make this ordering explicit since, without it, a much more complex technical apparatus is needed to handle the way in which spatial weights matrices evolve with an increasing sample size. The assumption is not often stated in the literature but is typically implicit. For example, in the first formal derivation of the asymptotic normality of Moran's I by Sen (1976), the main theorem is based upon the existence of an independent and identically distributed sequence Z_1, Z_2,... (with zero mean and

unit variance). However, without the assumption of a spatial ordering, the sequence should be expressed as the triangular array Z_{1N}, Z_{2N},..., indicating the dependence on the sample size. Without further restrictions, this would allow Z_{1N} to be different from Z_{1S} for $N \neq S$, invalidating several aspects of the relevant proofs.

The second part of assumption 4 is the standard convention that observations are not neighbors of themselves, that is, the diagonal elements of the spatial weights matrix are zero.

Assumption 4(c) states that the sum of all elements in the weights matrix (the familiar normalizing constant S_0 for Moran's I, e.g., Cliff and Ord 1981: 19) converges to a fixed constant s_1. This implies that the sum is, at most, of order N. Since each row in W_N has, at most, λ_2 nonzero elements that in turn are each less than c_w,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} w_{Nij} < N(\lambda_2 c_w) .$$

Assumption 4(d) is similar in nature, but pertains to another familiar normalizing constant,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (w_{Nij} + w_{Nji})^2 = 2 \sum_{i=1}^{N} \sum_{j=1}^{N} (w_{Nij} w_{Nji} + w_{Nij}^2)$$

$$= tr[(W_N + W_N')(W_N + W_N')] \tag{4}$$

(or $2S_1$ in the notation of Cliff and Ord 1981: 19). For the same reasons as before, the double sums in (4) are also, at most, of order N, and thus the sample size N is the proper deflator to obtain a constant (and finite) limit s_2. For example, using $\sum_{j=1}^{N} |w_{Nij}| < \lambda_2 c_w$, it can be seen that

$$\left| \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Nij} w_{Nji} \right| \leq \sum_{i=1}^{N} \sum_{j=1}^{N} |w_{Nij}| |w_{Nji}| \leq c_w \sum_{i=1}^{N} \left(\sum_{j=1}^{N} |w_{Nij}| \right) \leq N c_w^2 \lambda_2 .$$

ASSUMPTION 5: *The matrix* $(I - \rho W_N)$ *is nonsingular at the true value of* ρ, *for all* N.

This implies that the model is complete with respect to y_1, in the sense that equation (3) can be solved for y_1, as

$$y_1 = (I - \rho W_N)^{-1} [X_1 \beta + \rho(\tilde{W}_N y_{11}) + Y_2 \alpha + \varepsilon] .$$

ASSUMPTION 6: *The error terms* ε_i *are independent and identically distributed (i.i.d.), with zero mean,* $E(\varepsilon_i) = 0$, *fixed variance,* $E(\varepsilon_i^2) = \sigma_\varepsilon^2$, *and finite third absolute moment,* $E|\varepsilon_i|^3 = \mu_3$.

This may be viewed as the null hypothesis, say H_0, which implies, among other things, that the elements of the disturbance vector ε are not spatially auto-correlated.

> ASSUMPTION 7: *The matrices of observations on the exogenous variables,* X_1 *and* X_2, *are nonstochastic, and rank* $(X_1, X_2) = k_1 + k_2$. *In addition, consider an* N *by* k_3 *matrix,* X_3, *whose columns are a subset of the spatially lagged exogenous variables* $(W_N X_1, W_N X_2)$, *with* $k_3 \le k_1 + k_2$, *and let* $X = (X_1, X_2, X_3)$. *Then it is assumed that* $|x_{ij}| \le c_x < \infty$, *with* c_x *as a finite upper bound on the elements of* X, *and* $\lim_{N \to \infty} N^{-1}(X'X) = Q_X$, *where* Q_X *is nonsingular.*

This assumption implies that the analysis is conditional on the observed values of the exogenous variables in X_1 and X_2. Furthermore, both matrices are assumed to be of full rank, and perfect multicollinearity is excluded. Also, the elements of X_1 and X_2 are assumed to be bounded, which precludes certain types of variables, such as powers or exponentials of coordinates of locations in a trend surface regression. The elements of the matrix X will serve as instruments in the estimation procedure outlined in the next section.

Note that if a regression contains a spatially lagged dependent variable, the spatial lags of the exogenous variables, $W_N X_1$, can be used as instruments. (For a rigorous treatment, see Kelejian and Robinson 1993.) However, in the framework outlined here, these spatially lagged variables need not be used since some of the excluded exogenous variables, X_2, may constitute a proper instrument set. In this instance, k_3 would be zero, and the instrument matrix X could simply be (X_1, X_2). However, the large-sample efficiency of the IV estimator will increase with the number of instruments, i.e., with $k_3 > 0$.

The nonsingularity of Q_X is a standard assumption in a large-sample framework and should typically be satisfied in practice since the elements of X are bounded and perfect multicollinearity is excluded.

> ASSUMPTION 8: $\text{plim}_{N \to \infty}(X'X)^{-1}X'Z_1 = \Pi$, *where* Π *is a matrix of finite elements, with rank* $(\Pi) = k_1 + g + 1$.

This is a standard assumption needed to ensure that the parameters of the model are identified. A necessary condition for this is that $k_2 + k_3 \ge g + 1$, i.e., there are at least as many instruments X_2 and X_3 that do not appear in equation (1) as there are regressors that are correlated with the error term, which are the endogenous variables $W_N^* y_{1*}$, and Y_2. Assumption 8 also ensures that the endogenous variables $W_N^* y_{1*}$ and Y_2 are at least correlated with the instruments X.

ASSUMPTION 9:

(a) $\text{plim}_{N \to \infty} N^{-1} Z_1' Z_1 = S_1$,

(b) $\text{plim}_{N \to \infty} N^{-1} \varepsilon' Z_1 = S_2$,

(c) $\displaystyle\operatorname*{plim}_{N\to\infty} N^{-1}\varepsilon'W_NZ_1 = S_3$,

(d) $\displaystyle\operatorname*{plim}_{N\to\infty} N^{-1}Z_1'W_NZ_1 = S_4$,

where S_1, S_2, S_3, *and* S_4 *are matrices of finite elements.*

Assumption 9 ensures that the sample moments of the variables involved in the various operations needed in the estimation and in the proofs converge. Note that some of the elements of the S_i matrices may be zero, for example, the elements of S_2 that pertain to the plim of the cross product of the error terms ε with the nonstochastic (assumption 7) exogenous variables X_1.

INSTRUMENTAL VARIABLES ESTIMATION

As is well known, the presence of the endogenous variables in Y_2 in equation (1) violates the assumption of uncorrelatedness between regressors and error term that is fundamental for the unbiasedness or consistency of the ordinary least squares estimator. Similarly, the elements of a spatially lagged dependent variable $W_N^*y_{1*}$ will be correlated with the error term ε (Anselin 1988c: 58). Hence, an alternative approach must be used to obtain a consistent estimator for the parameter vector δ in equation (2). In addition, the implementation of Moran's I test for spatial error autocorrelation should not be based on residuals obtained from an inconsistent estimator such as ordinary least squares. Instead, such a test should be based upon residuals computed in terms of a consistent estimator of the model parameters.

Linear models with endogenous regressors are typically estimated by means of the two-stage least squares (2SLS) procedure, which is an instrumental variables (IV) method (see, e.g., Judge et al. 1985: 597–9; and, for the spatial case, Anselin 1988c: 81–8; Anselin 1990; Kelejian and Robinson 1993). With $P_X = X(X'X)^{-1}X'$ as the projection matrix associated with the matrix of instruments X (defined in assumption 7), the IV estimator for δ can be expressed as

$$\hat{\delta} = (Z_1'P_XZ_1)^{-1}Z_1'P_Xy_1, \qquad (5)$$

in the notation of equation (2). As shown in a number of places (e.g., Schmidt 1976, Davidson and MacKinnon 1993), the estimator $\hat{\delta}$ is consistent and asymptotically normal under a reasonable set of conditions that are satisfied by assumptions 1–9. More specifically,

$$N^{1/2}(\hat{\delta}-\delta) \xrightarrow{D} N[0, VC(\hat{\delta})], \qquad (6)$$

where $VC(\hat{\delta}) = \sigma_\varepsilon^2 \operatorname{plim} N(Z_1'P_XZ_1)^{-1}$ is the associated limiting covariance matrix.

The IV residuals are obtained as

$$\hat{\varepsilon} = y_1 - Z_1\hat{\delta}. \tag{7}$$

Since $\hat{\delta}$ is a consistent estimator for δ, or $\hat{\delta} \xrightarrow{P} \delta$, it can be shown under the set of assumptions given that $\hat{\sigma}_\varepsilon^2 = (\hat{\varepsilon}'\hat{\varepsilon})/N$ is a consistent estimator for the error variance σ_ε^2, or $\hat{\sigma}_\varepsilon^2 \xrightarrow{P} \sigma_\varepsilon^2$. Hence, approximate small-sample inference for δ may be based on the normal distribution, with $\hat{\delta} \approx N[\delta, \hat{\sigma}_\varepsilon^2(Z_1'P_XZ_1)^{-1}]$.

The consistency of $\hat{\delta}$ and $\hat{\sigma}_\varepsilon^2$ will be exploited in the next section to obtain the large-sample properties of Moran's I statistic based on IV residuals $\hat{\varepsilon}$.

MORAN'S I STATISTIC FOR RESIDUALS FROM IV ESTIMATION

Based on the IV residuals $\hat{\varepsilon}$, Moran's I statistic can be expressed as:

$$I^* = N(\hat{\varepsilon}'W_N\hat{\varepsilon})/S_0(\hat{\varepsilon}'\hat{\varepsilon}), \tag{8}$$

where W_N is the spatial weights matrix, and S_0 is the usual normalizing factor,

$$S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Nij}.$$

Using the previously outlined set of assumptions in conjunction with the consistency of $\hat{\delta}$, appendix A shows formally that $N^{1/2}I^*$ is asymptotically normal with mean zero and finite variance ϕ^2, or

$$N^{1/2}I^* \xrightarrow{D} N[0, \phi^2], \tag{9}$$

where ϕ^2 will be discussed in more detail below. With ϕ^2 replaced by a consistent estimator $\hat{\phi}^2$, an asymptotic test can be constructed such that the null hypothesis of no spatial autocorrelation may be rejected at the α level of significance if

$$\left| N^{1/2}I^*/\hat{\phi} \right| > z_\alpha, \tag{10}$$

where z_α is the value of a standard normal variate corresponding to α. For example, taking a significance level of 5 percent, the null hypothesis would be rejected if the expression in (10) exceeded 1.96.

Note that the exact specification of the error terms under the alternative hypothesis of spatial autocorrelation need not be known to obtain these asymptotic results. However, clearly, the power of the test will depend on how well the weights matrix W_N describes the pattern of spatial autocorrelation under the alternative. (For a more detailed discussion of the effect of misspecified weights on the power of tests, see Florax and Rey 1995.)

Implementation of the Test in the General Case

In the general case where both endogenous variables, Y_2, and a spatially lagged dependent variable, $W_N^* y_{1*}$, are present in the model specification [as in equation (1)], the asymptotic variance of $N^{1/2} I^*$ is shown in appendix A to be

$$\phi^2 = s_2/(2s_1^2) + (4/s_1^2 \sigma_\varepsilon^2) A , \qquad (11)$$

with σ_ε^2 as the error variance, s_1 and s_2 as given in assumption 4, and

$$A = \text{plim} \, [(N^{-1} \varepsilon' W_N Z_1)(N(Z_1' P_X Z_1)^{-1})(N^{-1} Z_1' W_N' \varepsilon)] . \qquad (12)$$

A consistent estimator for ϕ^2, say $\hat{\phi}^2$, is obtained by replacing σ_ε^2 with a consistent estimator, $\hat{\sigma}_\varepsilon^2 = (\hat{\varepsilon}'\hat{\varepsilon})/N$, and s_1, s_2, and A by their finite sample counterparts, respectively

$$\hat{s}_1 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{Nij}/N ,$$

$$\hat{s}_2 = (tr[(W_N + W_N')(W_N + W_N')])/N ,$$

and

$$\hat{A} = (N^{-1} \hat{\varepsilon}' W_N Z_1)(N(Z_1' P_X Z_1)^{-1})(N^{-1} Z_1' W_N' \hat{\varepsilon}) .$$

Note that the first term on the right-hand side of equation (12) (as well as its transpose in the third term) pertains to the probability limit of a series of cross-products of the vector of error terms ε with the spatial lags of the elements of $Z_1 = (X_1, W_N^* y_{1*}, Y_2)$. Specifically, this term can be expressed as

$$\text{plim} \, (N^{-1} \varepsilon' W_N Z_1) = \text{plim} \, (N^{-1} \varepsilon' W_N X_1, N^{-1} \varepsilon' W_N W_N^* y_{1*}, N^{-1} \varepsilon' W_N Y_2) . \qquad (13)$$

While $W_N^* y_{1*}$ and Y_2 in (13) are stochastic, X_1 is not, nor is its spatial lag, $W_N X_1$ (assumption 7). Furthermore, the elements of W_N (of which, at most, λ_2 in any row are nonzero due to assumption 4) are bounded due to assumption 1, and the elements of X_1 are bounded due to assumption 7. Hence it can be shown that $\text{plim} \, (N^{-1} \varepsilon' W_N X_1) = 0$.[3] This simplifies the first term in equation (13).

---

[3]Let $\varphi = N^{-1} X_1' W_N' \varepsilon$. Then a formal proof of this property is based on the fact that $E[\varphi] = 0$, and $\text{Var}[\varphi] = \sigma_\varepsilon^2 N^{-1}(N^{-1} X_1' W_N' W_N X_1)$. Assumptions 1, 4, and 7 imply that $N^{-1} X_1' W_N' W_N X_1$ is bounded and thus $\lim_{N \to \infty} \text{Var}[\varphi] = 0$. Hence, following Chebychev's inequality, $\text{plim} \, (\varphi) = 0$.

Special Case 1: No Endogenous Regressors

When no endogenous regressors are present in the model, then $Z_1 = X_1$, and, using the previous result, plim $(N^{-1}\varepsilon' W_N Z_1) = 0$. Consequently, the term A in equation (11) will equal zero and the variance $\phi^2 = s_2/(2s_1^2)$. In this case, it follows from equation (9) that

$$NI^{*2}/[\, s_2/(2s_1^2)] \xrightarrow{D} \chi^2(1), \tag{14}$$

since the square of a standard normal variate is distributed as chi-squared with one degree of freedom.

Note that equation (8) can also be written as

$$I^* = [(\hat{\varepsilon}' W_N \hat{\varepsilon})/(\hat{\varepsilon}'\hat{\varepsilon})]/(S_0/N) = [(\hat{\varepsilon}' W_N \hat{\varepsilon})/(\hat{\varepsilon}'\hat{\varepsilon})]/\hat{s}_1$$

[see (8) and (12)], and hence,

$$NI^{*2}/[\hat{s}_2/(2\hat{s}_1^2)] = N[(\hat{\varepsilon}' W_N \hat{\varepsilon})/(\hat{\varepsilon}'\hat{\varepsilon})]^2/(\hat{s}_2/2)$$

Also,

$$\hat{s}_2 = tr[(W_N + W_N')(W_N + W_N')]/N = 2tr(W_N W_N + W_N' W_N)/N.$$

Therefore, the test statistic in equation (14) simplifies to

$$[(\hat{\varepsilon}' W_N \hat{\varepsilon})/\hat{\sigma}_\varepsilon^2]^2/\ tr(W_N W_N + W_N' W_N) \ , \tag{15}$$

which is Burridge's (1980) Lagrange multiplier (LM) statistic for spatial error autocorrelation in regression residuals (see also Anselin 1988a).

In sum, the results of this special case imply that in the standard regression context the asymptotic distribution of N times the square of Moran's I is equivalent to that of the LM-error statistic. It is important to note that this result is obtained without having to resort to a normality assumption for the error terms.

Special Case 2: No Spatially Lagged Dependent Variables

A special case of great interest in practice is when the model specification contains only exogenous variables and endogenous variables, Y_2, that reflect system feedbacks, or $Z_1 = (X_1, Y_2)$. This is the standard (i.e., nonspatial) situation in which the 2SLS procedure is used to obtain consistent parameter estimators. Then, under a reasonable set of assumptions and in the absence of spatial autocorrelation (of either spatial error or spatial lag form) in the equations that determine the endogenous variables in Y_2, the term A in equation (11) equals zero. (The precise conditions under which this holds and a proof are given in appendix B.) Consequently, under these conditions, a test for spatial error auto-

correlation in models with (nonspatial) endogenous regressors can be based on the simplified form of the statistic in equations (14) or (15), which is identical in form to the usual LM-error test but based on the residuals obtained from IV estimation [equation (7)].

MONTE CARLO SIMULATION STUDY

An initial assessment of the size and power of the IV-based Moran's I test is obtained from a small number of Monte Carlo simulation experiments. Both the asymptotic form of the test (equations 9–13) as well as an ad hoc finite sample approximation based on the standard Cliff-Ord moments calculated in terms of IV residuals are considered. Note that in contrast to what holds in the classic linear regression model, there is no formal basis for this approximation in the presence of endogenous regressors.

Also considered are the size and power of Moran's I and the LM-error tests computed from the residuals of an OLS estimation. These tests are clearly ad hoc, since they represent a situation in which the endogeneity of a subset of regressors is ignored. Consequently, OLS will generally be both biased and inconsistent. However, this approach may seem attractive in practice, since OLS is simple to compute relative to more complex consistent procedures and often is superior in terms of mean squared error. It is therefore important to assess the extent to which the properties of these diagnostics are affected. This aspect of the simulations extends earlier work on the effect of misspecifications on the size and power of spatial autocorrelation tests, e.g., when heteroskedasticity or non-normality are present or when the spatial process is misspecified (see Anselin and Rey 1991; Anselin and Florax 1995; Florax and Rey 1995; Kelejian and Robinson 1995).

Experimental Design

The experimental design for the Monte Carlo simulations is based on a format extensively used in earlier studies in the standard linear regression model (e.g., Anselin and Rey 1991; Anselin and Florax 1995). A new feature is the consideration of endogeneity.

Two different contexts are considered. In the first, the empirical Type I errors of the tests are assessed when endogeneity is present in the form of systems feedbacks, but without a spatial lag in the model (see special case 2 in the previous section). A small system of two equations, with two endogenous variables and four exogenous variables, is specified:

$$y_{1i} = \alpha_1 + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 y_{2i} + \varepsilon_{1i} \tag{16}$$

$$y_{2i} = \alpha_2 + \beta_3 x_{3i} + \beta_4 x_{4i} + \gamma_2 y_{1i} + \varepsilon_{2i}, \tag{17}$$

where y_{1i} and y_{2i} are the i-th observation on the endogenous variables; $x_{1i},\ldots,$ x_{4i} are corresponding observations for the exogenous variables; α_r, γ_r, $r = 1, 2,$

and β_s, $s = 1,\ldots, 4$ are parameters; and ε_{1i} and ε_{2i} are uncorrelated error terms with mean zero and constant variance σ_i^2. For the typical set of assumptions, the system (16)–(17) is overidentified, reflecting a situation often encountered in practice. For the sake of simplicity, cross-equation correlation is not incorporated in the error terms, although this could be added in a straightforward way. In the simulations, the exogenous variables are generated as independent uniform random numbers between 0 and 1 and kept fixed in all experiments.

The values for the endogenous variables are obtained by solving the reduced form of the model (16)–(17):

$$
y_{1i} = \frac{\alpha_1 + \gamma_1\alpha_2}{1 - \gamma_1\gamma_2} + \frac{\beta_1}{1 - \gamma_1\gamma_2}x_{1i} + \frac{\beta_2}{1 - \gamma_1\gamma_2}x_{2i}
$$
$$
+ \frac{\gamma_1\beta_3}{1 - \gamma_1\gamma_2}x_{3i} + \frac{\gamma_1\beta_4}{1 - \gamma_1\gamma_2}x_{4i} + \frac{\varepsilon_{1i} + \gamma_1\varepsilon_{2i}}{1 - \gamma_1\gamma_2}
$$

(18)

and

$$
y_{2i} = \frac{\alpha_2 + \gamma_2\alpha_1}{1 - \gamma_1\gamma_2} + \frac{\gamma_2\beta_1}{1 - \gamma_1\gamma_2}x_{1i} + \frac{\gamma_2\beta_2}{1 - \gamma_1\gamma_2}x_{2i}
$$
$$
+ \frac{\beta_3}{1 - \gamma_1\gamma_2}x_{3i} + \frac{\beta_4}{1 - \gamma_1\gamma_2}x_{4i} + \frac{\gamma_2\varepsilon_{1i} + \varepsilon_{2i}}{1 - \gamma_1\gamma_2} \ .
$$

(19)

Clearly, different degrees of endogeneity can be obtained by manipulating the values of the parameters and the characteristics of the error terms. In these experiments, all parameters were set equal to 1, except γ_2, which was set equal to 0.1. This induced sufficient endogeneity to create a significant degree of simultaneity bias for the OLS estimator. For example, the bias of the OLS estimate for γ_1 was approximately −0.4 over all sample sizes considered (for a true value of 1.0). In comparison, the bias of the consistent IV estimator was −0.07 for samples of size $N = 48$ but only −0.0004 for samples of size $N = 900$. However, in terms of mean squared error (MSE), the OLS estimator was superior to the IV estimator for $N = 48$ and $N = 81$ due to its smaller sampling variance, but it was significantly inferior for the larger data sets. Specifically, for $N = 900$ the MSE of the OLS estimator was 0.17, compared to 0.006 for the IV estimator.

Four different error distributions were considered: the standard normal, the lognormal, the uniform, and the chi-squared distribution with one degree of freedom. In all cases, the generated error terms were transformed, using standard procedures, to obtain a zero mean and unit variance.

The second context considered is the model in which the endogeneity consists solely of the presence of a spatially lagged dependent variable. Formally, this is expressed as

$$y_1 = \rho W y_1 + \alpha_1 \iota + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_1,$$

where y_1, x_1, x_2, and ε_1 are N by 1 vectors of values of the corresponding terms defined in (16); ι is an N by 1 vector of ones; α_1, β_1, and β_2 are parameters; ρ is a spatial autoregressive coefficient; and W is a known spatial weights matrix. In the simulations, ρ was allowed to vary from 0.1 to 0.9 to assess the effect of spatial lag autocorrelation on the empirical Type I errors of test statistics. The exogenous variables are generated as before as uniform random variates, and $\beta_1 = \beta_2 = 1$. The observations on the dependent variable, y_1, are obtained from the usual reduced form solution,

$$y_1 = (I - \rho W)^{-1}(\alpha_1 \iota + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_1). \tag{20}$$

The power of the test statistics is compared under the alternative hypothesis of a spatial autoregressive error process $\varepsilon_1 = \lambda W \varepsilon_1 + \xi_1$, with ξ_1 as an error vector whose terms are i.i.d. with mean zero and unit variance. The same four distributions were considered for ξ_1 as before: the normal, the lognormal, the uniform, and the chi-squared with one degree of freedom. The autoregressive parameter λ took values from 0.1 to 0.9. This is implemented in the simultaneous system (16)–(17) by first generating $\varepsilon_1 = (I - \lambda W)^{-1} \xi_1$ and subsequently substituting the values for ε_1 in the reduced form equations (18) and (19).

Five different spatial configurations were considered: an irregular lattice structure corresponding to first-order contiguity of the 48 U.S. states; and four regular lattice structures for a rook-type contiguity in square 9 by 9 ($N = 81$), 11 by 11 ($N = 121$), 30 by 30 ($N = 900$), and 40 by 40 ($N = 1600$) lattices. All spatial weights matrices were used in row-standardized form.

Under the null hypothesis in the nonspatial model, 20,000 replications were generated, yielding a two standard deviation range for the empirical Type I error from 0.0469 to 0.0531 around $\alpha = 0.05$ and from 0.0086 to 0.0114 around $\alpha = 0.01$. For the other cases, 10,000 replications were used, yielding sufficient precision for the purposes of this study (2SD range for the empirical Type I error from 0.0456 to 0.0544 around $\alpha = 0.05$ and 0.008 to 0.012 around $\alpha = 0.01$). In the interest of space, only the more salient findings are listed here in some detail. A complete set of results for all cases considered is available from the authors.

Size of the Tests in the Presence of Systems Endogeneity

The first case considered pertains to the models (16)–(17) where an endogenous systems variable but no spatially lagged dependent variable is contained in the specification. The empirical rejection frequencies under the null hypothesis of no spatial error autocorrelation for the IV-error test (IV-ERR), Moran's I test using IV residuals (I-IV), the LM test for error autocorrelation using OLS residuals (LM-OLS), and Moran's I test applied to OLS residuals (I-OLS) are presented in table 1 for the four error distributions and four sample sizes (the results for $N = 1600$ are qualitatively similar to those for $N = 900$). The shaded cells in

TABLE 1. Empirical Rejection Frequencies under the Null Hypothesis, Endogeneity Only (20,000 replications)

Sample	α = 0.05				α = 0.01			
	IV-ERR	I-IV	LM-OLS	I-OLS	IV-ERR	I-IV	LM-OLS	I-OLS
N = 48								
Normal		0.0498		0.0482		0.0114		0.0103
Lognormal						0.0095		0.0113
Uniform		0.0527		0.0479		0.0110		0.0095
Chi-Squared						0.0106		0.0111
N = 81								
Normal	0.0499	0.0515		0.0474	0.0103	0.0107	0.0090	0.0094
Lognormal						0.0100		
Uniform	0.0520			0.0480				0.0098
Chi-Squared								
N = 121								
Normal	0.0480	0.0508	0.0474	0.0510	0.0089	0.0098	0.0087	0.0101
Lognormal						0.0089		0.0093
Uniform	0.0505		0.0498		0.0110		0.0099	0.0112
Chi-Squared								0.0088
N = 900								
Normal	0.0524		0.0499	0.0499	0.0099	0.0101	0.0106	0.0109
Lognormal	0.0485	0.0492		0.0469	0.0100	0.0105	0.0094	0.0096
Uniform	0.0514	0.0518	0.0510	0.0524	0.0103	0.0102	0.0110	
Chi-Squared	0.0476	0.0477			0.0087	0.0095	0.0091	0.0093

the table correspond to empirical rejection frequencies that fell outside the two standard deviation (2SD) range for the nominal rejection level. Clearly, the empirical Type I errors for the IV-ERR test do not fall within a 2SD interval of the theoretical size of the test in the smallest sample size, under-rejecting the null for all four distributions. However, for the normal distribution, the test performs properly for $N \geq 81$, and for all four distributions this is the case for $N = 900$. Moreover, one could argue that this under-rejection is not a serious problem since the correct null is not rejected and no alternative model is estimated. The relation between the sample size and the rejection frequency of the other (ad hoc) tests is less regular. The I-IV test has the proper size for the normal distribution (for $\alpha = 0.05$) in the small samples, but not for $N = 900$. The results are slightly better at $\alpha = 0.01$. Surprisingly, Moran's I test based on OLS residuals has the proper size in all cases with an underlying normal distribution. In other words, when normality is present and a spatially lagged dependent variable is absent, ignoring endogeneity in the form of system feedbacks does not seem to affect the *size* of Moran's I test. The rejection frequencies of the test based on the LM-OLS statistic, on the other hand, only lie within the appropriate 2SD range for $N \geq 121$ at $\alpha = 0.05$. However, for $\alpha = 0.01$ this is the case for all samples considered except $N = 48$. In the other cases, the null is under-rejected. The LM-OLS test also tends to be more sensitive to non-normality than the I-OLS test (again, under-rejecting the null), which confirms earlier results in the standard linear regression case (Anselin and Rey 1991).

In spite of its slight discrepancies at the tail of the distribution (reflected in the under-rejection reported above), the distribution of the IV-ERR statistic under the null is remarkably close to the normal, as illustrated in figure 1 for $N = 48$. In the larger sample sizes, the fit to the normal distribution is almost perfect, which clearly demonstrates the asymptotic nature of the test.

Size of the Tests in the Presence of a Spatially Lagged Dependent Variable

The results in terms of the size of the various tests are strikingly different when spatial interaction is the source of endogeneity in the model (in the form of a spatially lagged dependent variable). The IV-ERR test is the only test that properly corrects for the complex interaction between spatial lag and spatial error dependence (see, also, Anselin et al. 1996), but only for very large sample sizes. As shown in table 2 for normal errors, and apart from a few exceptions for $N = 48$, the test significantly over-rejects the null hypothesis for large values of ρ and under-rejects for small values. Even with $N = 1600$, the rejection rates still exceed the 2SD bound for $\rho > 0.5$, although to a lesser extent. The results are qualitatively similar for non-normal errors.

The performance of the ad-hoc tests is totally unacceptable when a spatial lag is present, as demonstrated by the results for the I-IV test (for normal errors) in table 3. The test significantly over-rejects the null hypothesis in all samples, and its performance does not improve as the sample size increases (stabilizing
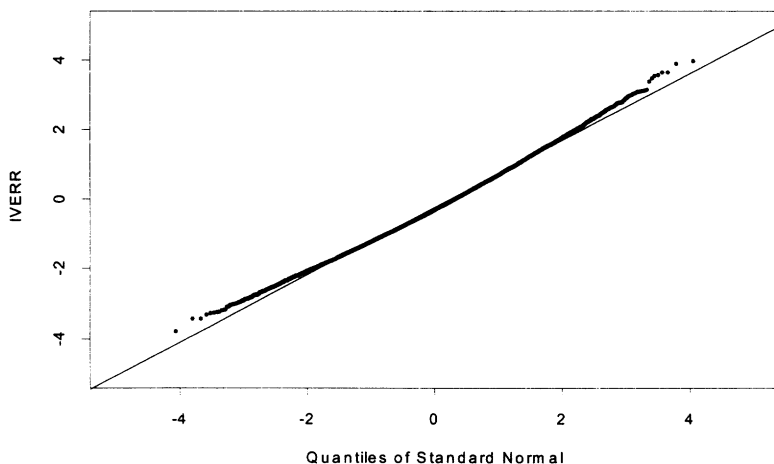
FIGURE 1. QQ-Norm Plot for IV-ERR, N = 48

around a 50 percent rejection rate for $N = 900$). The other tests are even worse in this respect.

Clearly, of the tests considered, the only one that appears to have reasonable empirical Type I errors in the presence of a spatially lagged dependent variable is the IV-ERR test, which applies a correction factor to Moran's I. A comparison of the results in table 1 to those in tables 2 and 3 clearly illustrates the difference in

TABLE 2. Rejection Frequencies under the Null Hypothesis
IV-ERR Test in the Spatial Lag Model (10,000 Replications)

ρ	$N = 48$	$N = 81$	$N = 121$	$N = 900$	$N = 1600$
0.1	0.0640	0.0251	0.0313	0.0484	0.0493
0.2	0.0742	0.0332	0.0387	0.0469	0.0464
0.3	0.0823	0.0476	0.0493	0.0505	0.0491
0.4	0.0898	0.0644	0.0650	0.0528	0.0520
0.5	0.0915	0.0878	0.0843	0.0578	0.0520
0.6	0.0887	0.1165	0.1092	0.0607	0.0551
0.7	0.0766	0.1417	0.1400	0.0638	0.0549
0.8	0.0582	0.1521	0.1694	0.0677	0.0582
0.9	0.0324	0.1248	0.1607	0.0642	0.0649

**TABLE 3. Rejection Frequencies under the Null Hypothesis
I-IV Test in the Spatial Lag Model (10,000 Replications)**

ρ	$N = 48$	$N = 81$	$N = 121$
0.1	0.3329	0.5031	0.5170
0.2	0.3289	0.4987	0.5215
0.3	0.3172	0.4953	0.5251
0.4	0.3030	0.4928	0.5214
0.5	0.2794	0.4788	0.5251
0.6	0.2515	0.4643	0.5212
0.7	0.2142	0.4348	0.5102
0.8	0.1689	0.3717	0.4832
0.9	0.1174	0.2760	0.3885

the nature of the endogeneity between a system feedback variable and a spatially lagged dependent variable. As the reduced form (20) indicates, each value of the dependent variable is related to all values of the error term. This is not the case in models that contain only endogenous variables that represent system feedbacks. This suggests that the correction term A (equation 11), which is only included in the IV-ERR statistic, plays a crucial role in obtaining the proper size. Also, in contrast to what is sometimes suggested in the literature, the spatially lagged dependent variable cannot be viewed as simply another endogenous variable, but the spatial nature of the endogeneity must be explicitly accounted for.

Finally, from table 2, it is seen that the empirical Type I errors of the IV-ERR test generally exceed the theoretical size, even in moderately sized samples such as $N = 121$. A rejection of the null in small samples may therefore be spurious and a further size correction of the statistic may be necessary.

Power of the Tests

The frequency with which the null hypothesis is rejected for increasing degrees of spatial error autocorrelation is reported in table 4 for both IV-based and OLS-based tests. In this instance, the seeming acceptability of using OLS residuals in the presence of endogenous regressors has disappeared, and both IV tests show significantly superior power. The power is essentially unaffected by non-normal distributions (not reported in table 4). Of the two, the I-IV test performs marginally better, particularly for the smaller samples and smaller values of λ, but both tests demonstrate good power. Even in the smallest data set, the null is rejected in roughly half the cases by both tests for $\lambda \geq 0.5$, while for $N = 121$ this is already achieved for $\lambda > 0.2$ and in over 95 percent of the cases for $\lambda \geq 0.5$. By contrast, the 50 percent mark is never reached for the OLS based tests with $N = 48$, and even for $N = 121$ the rejection frequency tops out at 78 percent (for the I-OLS test with $\lambda = 0.8$).

**TABLE 4. Power against Spatial Autoregressive Error
Endogeneity Only (10,000 Replications)**

λ	N = 48		N = 81		N = 121	
	IV-ERR	I-IV	IV-ERR	I-IV	IV-ERR	I-IV
0.1	0.0480	0.0805	0.0673	0.0959	0.0871	0.1164
0.2	0.0900	0.1527	0.1669	0.2258	0.2712	0.3272
0.3	0.1767	0.2682	0.3506	0.4309	0.5659	0.6231
0.4	0.3101	0.4282	0.5826	0.6630	0.8174	0.8554
0.5	0.4810	0.5928	0.7839	0.8369	0.9500	0.9625
0.6	0.6458	0.7364	0.9035	0.9271	0.9878	0.9907
0.7	0.7651	0.8297	0.9467	0.9571	0.9943	0.9947
0.8	0.8253	0.8626	0.9575	0.9643	0.9935	0.9940
0.9	0.8094	0.8489	0.9439	0.9509	0.9885	0.9891
λ	LM-OLS	I-OLS	LM-OLS	I-OLS	LM-OLS	I-OLS
0.1	0.0494	0.0773	0.0443	0.0566	0.0548	0.0696
0.2	0.0704	0.1119	0.0612	0.0879	0.1023	0.1326
0.3	0.1021	0.1640	0.1033	0.1539	0.1930	0.2384
0.4	0.1488	0.2310	0.1710	0.2390	0.3153	0.3781
0.5	0.2068	0.3127	0.2548	0.3452	0.4624	0.5318
0.6	0.2716	0.3989	0.3473	0.4451	0.5975	0.6664
0.7	0.3252	0.4692	0.4180	0.5262	0.6926	0.7590
0.8	0.3349	0.4952	0.4251	0.5442	0.7133	0.7827
0.9	0.2490	0.4038	0.2938	0.4144	0.5442	0.6332

In other words, there is some evidence that ignored endogeneity significantly lowers the power of OLS-based tests against spatial error autocorrelation. While in this instance an OLS estimator may still be superior to an IV estimator in terms of mean squared error, especially in small samples, this attractive property does not extend to diagnostic tests based on least squares residuals.

CONCLUDING REMARKS

As in any simulation study, the generality of the conclusions drawn from this series of experiments is limited by their design. Only a small set of aspects related to endogeneity and the interaction between the size and power of the tests has been considered. Many issues remain to be explored, such as the number of endogenous variables, the degree of identifiability, inter-equation correlation, the choice of instruments, and others that affect the properties of estimators.

The results of the Monte Carlo simulations suggest that the asymptotic form of Moran's I test based on IV residuals, which is derived in this paper, achieves

the proper size in medium-sized samples ($N = 81$) with normally distributed error terms, and in slightly larger samples ($N = 121$) for the other distributions considered. These sample sizes are typical of the number of counties in many states, or the number of census tracts in medium-sized metropolitan areas, which is encouraging for empirical research. However, there is an indication that the test may under-reject the null hypothesis when it is true and have less power in the presence of weak spatial error autocorrelation in the smallest sample ($N = 48$). On the other hand, the test seems to have good overall power, and, of those considered, is the only test that behaves properly when a spatially lagged dependent variable is included in the model. In addition, the test is easy to compute and asymptotically equivalent to the familiar Lagrange multiplier test for spatial error autocorrelation when no spatially lagged variables are present. As a corollary, it was also shown that Moran's I test and the Lagrange multiplier test for spatial error autocorrelation are asymptotically equivalent in the classic case when only exogenous regressors are present, without requiring the assumption of normally distributed errors.

The other tests considered have been used in practice, but are ad hoc in the sense that they are not based on formal properties of the corresponding statistics. The tests based on a Moran's I statistic formulated in terms of IV residuals has surprisingly good properties, especially when the error distribution is normal. The formal basis for these results remains to be determined. However, the attractive properties of this test hold only in the purely endogenous case. When spatially lagged dependent variables are included in the model, the test is unreliable in the sense that the null hypothesis is rejected too frequently when no spatial error autocorrelation is present.

OLS based tests, especially Moran's I test, seem to have the proper size when the errors are normally distributed. However, ignoring endogeneity results in a significant loss in power. In other words, while there is little to lose by ignoring endogeneity when no spatial error autocorrelation is present, the use of OLS-based tests in general is ill-advised since it will tend to under-reject the null when a problem is present. This is important to note, since the generally lower MSE of OLS estimators relative to IV estimators in small samples has often led to the use of OLS in practice (e.g., in the estimation of regional econometric models), even though it is inconsistent in the presence of endogenous regressors.

In sum, the asymptotic properties of Moran's I test based on IV residuals are reflected in reasonably sized samples. While the ad hoc version based on the usual mean and variance approximations also performed well in the simulations, and even achieved slightly higher power, the reason for this is not well understood. Since this test is easy to compute and is the only procedure of the ones considered that has acceptable properties in the presence of a spatially lagged dependent variable, it should be used in practice whenever instrumental variables estimation is carried out for cross-sectional data. Further work is needed in order to assess the extent to which these conclusions hold in a wider range of contexts

and to understand more fully the large-sample properties of the ad hoc version of the test.

REFERENCES

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

Anselin, L. 1988b. Model validation in spatial econometrics: A review and evaluation of alternative approaches. *International Regional Science Review* 11: 279–316.

Anselin, L. 1988c. *Spatial econometrics: Methods and models.* Dordrecht: Kluwer Academic.

Anselin, L. 1990. Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics* 20: 141–63.

Anselin, L., and A. Bera. 1997. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of applied economic statistics*, eds. A. Ullah and D. Giles. New York: Marcel Dekker (forthcoming).

Anselin, L., and R. Florax. 1995. Small sample properties of tests for spatial dependence in regression models: Some further results. In *New directions in spatial econometrics*, eds. L. Anselin and R. Florax. Berlin: Springer-Verlag.

Anselin, L., and S. Hudak. 1992. Spatial econometrics in practice, a review of software options. *Regional Science and Urban Economics* 22: 509–36.

Anselin, L., and S. Rey. 1991. Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23: 112–31.

Anselin, L., A. Bera, R. Florax, and M. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26: 77–104.

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.

Burridge, P. 1980. On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B* 42: 107–8.

Cliff, A., and J. K. Ord. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4: 267–84.

Cliff, A., and J. K Ord. 1973. *Spatial autocorrelation.* London: Pion.

Cliff, A., and J. K. Ord. 1981. *Spatial processes: Models and applications.* London: Pion.

Cressie, N. 1993. *Statistics for spatial data.* New York: Wiley.

Davidson, R., and J. G. MacKinnon. 1993. *Estimation and inference in econometrics.* New York: Oxford University Press.

Florax, R., and S. Rey. 1995. The impacts of misspecified spatial interaction in linear regression models. In *New directions in spatial econometrics*, eds. L. Anselin and R. Florax. Berlin: Springer-Verlag.

Griffith, D. 1983. The boundary value problem in spatial statistical analysis. *Journal of Regional Science* 23: 377–87.

Holtz-Eakin, D. 1994. Public-sector capital and the productivity puzzle. *Review of Economics and Statistics* 76: 12–21.

Hordijk, L. 1974. Spatial correlation in the disturbances of a linear interregional model. *Regional and Urban Economics* 4: 117–40.

Judge, G., W. E. Griffiths, R. Carter Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The theory and practice of econometrics* (2nd edition). New York: Wiley.

Kelejian, H. H., and I. R. Prucha. 1995. A generalized moments estimator for the autoregressive parameter in a spatial model. College Park: University of Maryland Department of Economics, Working Paper No. 95-03.

Kelejian, H. H., and D. P. Robinson. 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science* 72: 297–312.

Kelejian, H. H., and D. P. Robinson. 1995. A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. College Park: University of Maryland Department of Economics, Working Paper No. 95-14.

Kelejian, H. H., and D. P. Robinson. 1997. Infrastructure productivity estimation and its underlying econometric specifications: A sensitivity analysis. *Papers in Regional Science* 76: 115–31.

King, M. L. 1981. A small sample property of the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society B* 43: 263–64.

King, M. L. 1987. Testing for autocorrelation in linear regression models: A survey. In *Specification analysis in the linear model*, eds. M. King and D. Giles. London: Routledge and Kegan Paul.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23.

Schmidt, P. 1976. *Econometrics*. New York: Marcel Dekker.

Sen, A. 1976. Large sample-size distribution of statistics used in testing for spatial autocorrelation. *Geographical Analysis* 9: 175–84.

Sen, A. 1990. Distribution of spatial autocorrelation statistics. In *Spatial statistics: Past, present, and future*, ed. D. Griffith. Ann Arbor, MI: IMAGE.

Terui, N., and M. Kikuchi. 1994. The size-adjusted critical region of Moran's *I* test statistics for spatial autocorrelation and its application to geographical areas. *Geographical Analysis* 26: 213–27.

Tiefelsdorf, M., and B. Boots. 1995. The exact distribution of Moran's *I*. *Environment and Planning A* 27: 985–99.

Weibull, J. 1976. An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics* 6: 357–79.

APPENDIX A

The asymptotic distribution of $N^{1/2}I*$ is derived by applying a central limit theorem and using a series of results on the limiting properties of the numerator and denominator of $I*$ (equation 8). To begin, consider more closely both the numerator and denominator of $N^{1/2}I*$,

$$N^{1/2}I* = (N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon})/\hat{s}_1(\hat{\varepsilon}'\hat{\varepsilon}/N). \qquad (A.1)$$

Note that the limit of the first term in the denominator $\lim_{N\to\infty} \hat{s}_1 = s_1$ by virtue of assumption 4. The second element converges to the error variance, or plim $(\hat{\varepsilon}'\hat{\varepsilon}/N) = \sigma_\varepsilon^2$ (see the section on IV estimation). Consequently, the denominator in (A.1) converges in probability to a nonstochastic constant $s_1\sigma_\varepsilon^2$. Apart from this constant, the asymptotic distribution of $N^{1/2}I*$ in (A.1) is therefore completely determined by the asymptotic distribution of the numerator term on the right-hand side.

The 2SLS estimator $\hat{\delta}$ [equation (5)] may be expressed as

$$\hat{\delta} = \delta + \Delta_N, \qquad (A.2)$$

where $N^{1/2}\Delta_N$ is $O_p(1)$ in light of the limiting result in equation (6), and thus plim(Δ_N) = 0. This can be exploited to establish a relation between the IV residuals $\hat{\varepsilon}$ and the error terms ε. Substituting (A.2) into equation (7) yields $\hat{\varepsilon} = y_1 - Z_1(\delta + \Delta_N)$, or, using (2),

$$\hat{\varepsilon} = \varepsilon - Z_1\Delta_N. \qquad (A.3)$$

Now, consider more closely the asymptotic properties of the numerator on the right-hand side of (A.1) by substituting (A.3) into the spatial cross-products,

$$N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon} = N^{-1/2}(\varepsilon - Z_1\Delta_N)'W_N(\varepsilon - Z_1\Delta_N),$$

or

$$N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon} = N^{-1/2}\varepsilon'W_N\varepsilon - 2N^{-1/2}(\varepsilon'W_NZ_1)\Delta_N + N^{-1/2}\Delta_N'(Z_1'W_NZ_1)\Delta_N. \qquad (A.4)$$

After some manipulation of the powers of N, it easily follows that the second term on the right-hand side of (A.4) equals $-2(\varepsilon'W_NZ_1/N)N^{1/2}\Delta_N$, and the third term equals $N^{-1/2}(N^{1/2}\Delta_N)'(Z_1'W_NZ_1/N)(N^{1/2}\Delta_N)$. Since $N^{1/2}\Delta_N$ is $O_p(1)$ and bounded in probability, and plim $(Z_1'W_NZ_1/N) = S_4$ [assumption 9(d)], the probability limit of the third term is zero (due to the presence of $N^{-1/2}$). As a result, taking probability limits of the difference of terms in (A.4) yields

$$\text{plim}\{(N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}) - [(N^{-1/2}\varepsilon'W_N\varepsilon) - (2(\varepsilon'W_NZ_1/N)N^{1/2}\Delta_N)]\}$$
$$= \text{plim}[N^{-1/2}\Delta_N'(Z_1'W_NZ_1)\Delta_N] = 0. \qquad (A.5)$$

Using assumption 9(c), plim $(\varepsilon'W_NZ_1/N) = S_3$, thus

$$\text{plim}\{(N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}) - [(N^{-1/2}\varepsilon'W_N\varepsilon) - (2(\varepsilon'W_NZ_1/N)N^{1/2}\Delta_N)]\}$$
$$= \text{plim}\{(N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}) - [(N^{-1/2}\varepsilon'W_N\varepsilon) - (2S_3N^{1/2}\Delta_N)]\} = 0. \qquad (A.6)$$

By substituting equation (2) for y_1 in equation (5), it follows readily that

$$\Delta_N = \hat{\delta} - \delta = (Z_1'P_XZ_1)^{-1}Z_1'P_X\varepsilon ,$$

or, with $\hat{Z}_1 = P_XZ_1 = X(X'X)^{-1}X'Z_1 = X\hat{\Pi}$ (using the notation of assumption 8), and thus $Z_1'P_XZ_1 = \hat{\Pi}'X'X\hat{\Pi}$,

$$\Delta_N = (\hat{\Pi}'X'X\hat{\Pi})^{-1}\hat{\Pi}'X'\varepsilon ,$$

and, thus

$$N^{1/2}\Delta_N = [\hat{\Pi}'(N^{-1}X'X)\hat{\Pi}]^{-1}\hat{\Pi}'(N^{-1/2}X'\varepsilon) .$$

Consequently, the following probability limit is obtained

$$\text{plim } \{(N^{1/2}\Delta_N) - [\Pi'Q_X\Pi]^{-1}\Pi'(N^{-1/2}X'\varepsilon)\} = 0 \qquad (A.7)$$

where plim $\hat{\Pi} = \Pi$ due to assumption 8, and plim $(N^{-1}X'X) = Q_X$ in view of assumption 7.
Substituting (A.7) and the result in (A.6) into (A.5) yields

$$\text{plim } \{(N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}) - [(N^{-1/2}\varepsilon'W_N\varepsilon) - 2S_3[\Pi'Q_X\Pi]^{-1}\Pi'(N^{-1/2}X'\varepsilon)]\} = 0$$

or,

$$\text{plim } \{(N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}) - N^{-1/2}(\varepsilon'W_N\varepsilon - 2S_3[\Pi'Q_X\Pi]^{-1}\Pi'X'\varepsilon)\} = 0 . \qquad (A.8)$$

Simplify notation by setting

$$G' = -2S_3[\Pi'Q_X\Pi]^{-1}\Pi'X' , \qquad (A.9)$$

which is a 1 by N vector with elements g_i, and

$$\Psi = N^{-1/2}(\varepsilon'W_N\varepsilon + G'\varepsilon) .$$

Then (A.8) can also be written as

$$\text{plim } (N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon} - \Psi) = 0 . \qquad (A.10)$$

Using a familiar asymptotic result, equation (A.10) implies that if the limiting distribution for Ψ can be established, this will also be the limiting distribution of $N^{-1/2}\hat{\varepsilon}'W_N\hat{\varepsilon}$. This limiting distribution is obtained by expressing Ψ as a sum of random variables and applying the central limit theorem for m-dependent variables given in Schmidt (1976: 258). In the typical fashion, this will yield $\Psi \xrightarrow{D} N(0, \sigma_\Psi^2)$, where the asymptotic variance σ_Ψ^2 must be derived from the terms that constitute Ψ. To obtain these terms, first note that

$$\varepsilon'W_N\varepsilon + G'\varepsilon = \sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_i\varepsilon_j w_{Nij} + \sum_{i=1}^{N}g_i\varepsilon_i , \qquad (A.11)$$

where g_i are the elements of the vector G' in (A.9), and the other notation is as before. Some further manipulation yields

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (\varepsilon_i \varepsilon_j w_{Nij}) + \sum_{i=1}^{N} g_i \varepsilon_i = \sum_{i=1}^{N} \varepsilon_i \left(\sum_{j=1}^{N} \varepsilon_j w_{Nij} + g_i \right)$$

$$= \sum_{i=1}^{N} \varepsilon_i (w_{Ni} \varepsilon + g_i) ,$$

or

$$\Psi \equiv N^{-1/2} \sum_{i=1}^{N} v_i ,$$

where

$$v_i = \varepsilon_i (w_{Ni} \varepsilon + g_i) . \tag{A.12}$$

It must now be demonstrated that Ψ as defined in (A.12) satisfies the four conditions for the central limit theorem for m-dependent variables as described in Schmidt (1976).

CONDITION 1: $E[v_i] = 0$

This condition is satisfied since $w_{Nii} = 0$ [assumption 4(b)] and thus $\varepsilon_i(w_{Ni} \varepsilon)$ in (A.12) only contains cross-products of the form $\varepsilon_i w_{Nij} \varepsilon_j$, with $i \neq j$. Since assumption 6 specifies the error terms to be i.i.d. with zero means, $E[\varepsilon_i w_{Nij} \varepsilon_j] = 0$ when $i \neq j$, and also $E[\varepsilon_i g_i] = 0$, therefore $E[v_i] = 0$.

CONDITION 2: $(v_1, \ldots, v_i)$ *is independent of* $(v_{i+q}, \ldots, v_N)$ *for all* q *greater than a given constant.*

This condition is satisfied by the structure of the weights matrices imposed under assumption 4(a) (where the finite constant λ_2 plays the role of q) and the independence of the error terms in assumption 6. Therefore, two sequences of v_i that are further than λ_2 apart will have no error terms ε_i in common.

CONDITION 3: $E[|v_i|^3] < \lambda_3$, $\forall i$, *where* λ_3 *is a finite constant.*

The boundedness of the absolute third moments can be demonstrated by considering the following inequality, obtained from (A.12):

$$E[|v_i|^3] \leq E[|\varepsilon_i|^3] \ E[|w_{Ni} \varepsilon + g_i|^3] ,$$

or, after expanding the terms in the third power,

$$E[|v_i|^3] \leq E[|\varepsilon_i|^3] \ E[|w_{Ni} \varepsilon|^3 + 3|w_{Ni} \varepsilon|^2 |g_i| + 3|w_{Ni} \varepsilon||g_i|^2 + |g_i|^3] . \tag{A.13}$$

Several assumptions are used to obtain the boundedness of the powers involving spatial weights and error terms in (A.13). Assumption 1 ensures that individual spatial weights are bounded by c_w, and assumption 4(a) requires that there are, at most, λ_2 nonzero elements in each row of the spatial weights matrix. Consequently, the term $w_{Ni} \varepsilon$ involves, at most, λ_2 cross-product terms of the form $w_{Nij} \varepsilon_j$ whose absolute third moments are bounded by $|c_w^3| |E|\varepsilon_i|^3$, which is finite in light of assumption 6. Since the third absolute moments are bounded, all lower order absolute moments will be bounded as well. Furthermore, the elements g_i are bounded because Q_X, Π, and S_3 are assumed to

be finite matrices in assumptions 7–9. This establishes bounds on the right-hand side in inequality (A.13) and hence the boundedness of $E[|v_i|^3]$.

CONDITION 4: $\lim\limits_{N \to \infty} Var\left[N^{-1/2} \sum\limits_{i=1}^{N} v_i\right] = \sigma_\psi^2$ *exists.*

Using (A.11), σ_ψ^2 can be expressed as

$$\sigma_\psi^2 = \lim_{N \to \infty} Var\left[N^{-1/2}\left(\sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i\varepsilon_j w_{Nij} + \sum_{i=1}^{N} g_i\varepsilon_i\right)\right]. \tag{A.14}$$

Since $w_{Nii} = 0$ [assumption 4(b)], the double sum over the spatial cross product terms in (A.14) can be simplified to

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i\varepsilon_j w_{Nij} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \varepsilon_i\varepsilon_j(w_{Nij} + w_{Nji}) \tag{A.15}$$

The variance of (A.15) consists of the sum of the variances of $\varepsilon_i\varepsilon_j(w_{Nij} + w_{Nji})$ and the covariances between $\varepsilon_i\varepsilon_j(w_{Nij} + w_{Nji})$ and $\varepsilon_r\varepsilon_s(w_{Nrs} + w_{Nsr})$. Since the error terms are i.i.d. (assumption 6),

$$E[\varepsilon_i\varepsilon_j\varepsilon_r\varepsilon_s] = 0$$

unless two pairs of indices are equal ($i = r$ and $j = s$, or $i = s$ and $j = r$). Given that $j > i + 1$ in (A.15), all covariances are zero, and the variance for each term is

$$E[(\varepsilon_i\varepsilon_j)^2](w_{Nij} + w_{Nji})^2 = E[\varepsilon_i^2]E[\varepsilon_j^2](w_{Nij} + w_{Nji})^2 = \sigma_\varepsilon^4(w_{Nij} + w_{Nji})^2,$$

and consequently

$$Var\left[N^{-1/2} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i\varepsilon_j w_{Nij}\right] = N^{-1}\sigma_\varepsilon^4 \sum_{i=1}^{N} \sum_{j=i+1}^{N} (w_{Nij} + w_{Nji})^2. \tag{A.16}$$

Note that the summation indices in (A.15) result in half the sum over all i, j [as in equation (4)]. Therefore, from equation (4) and the definition of s_2 in assumption 4, it follows that

$$\lim_{N \to \infty} N^{-1}\sigma_\varepsilon^4 \sum_{i=1}^{N} \sum_{j=i+1}^{N} (w_{Nij} + w_{Nji})^2 = \sigma_\varepsilon^4 s_2/2 \tag{A.17}$$

Since the errors are i.i.d., the variance of the second term in (A.14) is simply $N^{-1} \sum\limits_{i=1}^{N} g_i^2\sigma_\varepsilon^2$, or $\sigma_\varepsilon^2 G'G/N$. Using (A.9), it follows that

$$\lim_{N \to \infty} G'G/N = \lim_{N \to \infty} 4S_3(\Pi'Q_X\Pi)^{-1}\Pi'(X'X/N)\Pi(\Pi'Q_X\Pi)^{-1}S_3',$$

or, since $\lim\limits_{N \to \infty} X'X/N = Q_X$ due to assumption 7, it follows that

$$\lim_{N \to \infty} G'G/N = 4S_3(\Pi'Q_X\Pi)^{-1}S_3',$$

and thus

$$\lim_{N \to \infty} \sigma_\varepsilon^2 G'G/N = 4\sigma_\varepsilon^2 S_3(\Pi' Q_X \Pi)^{-1} S_3' . \qquad (A.18)$$

All covariance terms between the first and second term in (A.14) are zero since $E[\varepsilon_i \varepsilon_j \varepsilon_r] = 0$ unless $i = j = r$, which is ruled out because $w_{Nii} = 0$. Consequently, the limit of the variance of Ψ equals the sum of (A.17) and (A.18), or

$$\sigma_\Psi^2 = \sigma_\varepsilon^4 s_2/2 + 4\sigma_\varepsilon^2 S_3(\Pi' Q_X \Pi)^{-1} S_3' .$$

Since the four conditions required by Schmidt's central limit theorem are satisfied, the limiting distribution of Ψ can be established as $N(0, \sigma_\Psi^2)$, and, due to (A.10),

$$N^{-1/2} \hat{\varepsilon}' W_N \hat{\varepsilon} \xrightarrow{D} N(0, \sigma_\Psi^2) ,$$

and thus also,

$$N^{1/2} I* \xrightarrow{D} N[0, \sigma_\Psi^2/(s_1^2 \sigma_\varepsilon^4)] .$$

APPENDIX B

In the case of no spatially lagged dependent variables, $Z_1 = (X_1, Y_2)$, so equation (12) simplifies to plim $(N^{-1}\varepsilon' W_N Z_1) =$ plim $(N^{-1}\varepsilon' W_N X_1, N^{-1}\varepsilon' W_N Y_2)$. It was shown in footnote 3 that plim $(N^{-1}\varepsilon' W_N X_1) = 0$, so for A to be zero in equation (11), it must still be demonstrated that plim $(N^{-1}\varepsilon' W_N Y_2) = 0$. For ease of exposition, consider the case where there is only one endogenous regressor, y_2, in Y_2, so that Y_2 simplifies to an N by 1 vector. The extension to multiple endogenous variables is straightforward.

The setup described for the general model in the second section of this paper implied the existence of additional equations in which the endogenous system feedback variables Y_2 are determined as functions of all excluded exogenous variables X_2. In general, each endogenous variable may therefore be expressed as a function of exogenous variables in a reduced form. Assuming such a reduced form exists and is linear, y_2 can be expressed as

$$y_2 = R\pi + \xi_2, \tag{B.1}$$

where R is a matrix of observations on exogenous variables, π is the associated vector of parameters, and ξ_2 is a vector of error terms with elements ξ_{2i}. Consistent with assumption 7, it is assumed that the elements of R are bounded, and $N^{-1}R'R \to Q_R$ where Q_R^{-1} exists. In line with assumption 6 for the error terms ε_i, assume that the error terms ξ_{2i} have mean zero and a finite variance σ_ξ^2. Furthermore, the error terms are assumed to be independently distributed over the spatial units so that $E[\xi_{2i}\xi_{2j}] = 0$ for $i \neq j$. This explicitly excludes the presence of a spatially lagged dependent variable in the equation for y_2.

From (B.1) it follows that

$$\text{plim } (N^{-1}\varepsilon' W_N y_2) = \text{plim } (N^{-1}\varepsilon' W_N R\pi) + \text{plim } (N^{-1}\varepsilon' W_N \xi_2).$$

Since R is exogenous and its elements are bounded, it can be shown by means of the same approach used in footnote 3 that plim $(N^{-1}\varepsilon' W_N R\pi) = 0$ and thus

$$\text{plim } (N^{-1}\varepsilon' W_N y_2) = \text{plim } (N^{-1}\varepsilon' W_N \xi_2). \tag{B.2}$$

It can be established that plim $(N^{-1}\varepsilon' W_N \xi_2) = 0$ in the same manner as elsewhere in this paper by relying on Chebychev's inequality together with $E[N^{-1}\varepsilon' W_N \xi_2] = 0$ and $\lim_{N\to\infty} Var[N^{-1}\varepsilon' W_N \xi_2] = 0$. The first result readily follows by considering the cross products involved and using the same rationale as in condition 1 of appendix A:

$$E[\varepsilon' W_N \xi_2] = E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_i\xi_{2j}w_{Nij}\right] = 0 \tag{B.3}$$

since the error terms ε_i and ξ_{2j} are independent for $i \neq j$, and $w_{Nii} = 0$. The second result follows by considering the same simplification of the spatial cross-products as utilized in (A.15) together with the independence of the error terms (as in A.16), such that

$$Var\left[N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_i\xi_{2j}w_{Nij}\right] = N^{-2}\sigma_\varepsilon^2\sigma_\xi^2\sum_{i=1}^{N}\sum_{j=i+1}^{N}(w_{Nij} + w_{Nji})^2,$$

Also, using the definition of s_2 from assumption 4 (analogous to the result used in A.17), it follows that

$$\lim_{N \to \infty} N^{-2} \sigma_\varepsilon^2 \sigma_\xi^2 \sum_{i=1} \sum_{j=i+1} (w_{Nij} + w_{Nji})^2 = \lim_{N \to \infty} N^{-1} \sigma_\varepsilon^2 \sigma_\xi^2 (s_2/2) = 0,$$

or

$$\lim_{N \to \infty} Var[N^{-1} \varepsilon' W_N \xi_2] = 0. \tag{B.4}$$

The results of (B.2), (B.3), and (B.4) together with Chebychev's inequality establish that $\text{plim}\, (N^{-1} \varepsilon' W_N y_2) = 0$.