Properties of Tests for Spatial Dependence in Linear Regression Models

Based on a large number of Monte Carlo simulation experiments on a regular lattice, we compare the properties of Moran's I and Lagrange multiplier tests for spatial dependence, that is, for both spatial error autocorrelation and for a spatially lagged dependent variable. We consider both bias and power of the tests for six sample sizes, ranging from twenty-five to 225 observations, for different structures of the spatial weights matrix, for several underlying error distributions, for misspecified weights matrices, and for the situation where boundary effects are present. The results provide an indication of the sample sizes for which the asymptotic properties of the tests can be considered to hold. They also illustrate the power of the Lagrange multiplier tests to distinguish between substantive spatial dependence (spatial lag) and spatial dependence as a nuisance (error autocorrelation).

1. INTRODUCTION

The recent surge of interest in geographic information systems (GIS) in the field of geographical analysis has resulted in an increased attention to so-called exploratory spatial analysis (ESA) as a special case of exploratory data analysis (EDA; see Tukey 1977), or, in other words, to "letting the data speak for themselves" (Gould 1981; see also Haining 1990). In spite of the seeming attractiveness of such a notion, very few of the standard exploratory data analysis techniques are equipped to deal with the distinguishing characteristics of spatial data, that is, spatial dependence and spatial heterogeneity. For example, a routine type of EDA yields a series of simple regressions and associated measures of fit ($R^2$), as illustrated in Haining (1990). However, as demonstrated in Anselin...
(1988a), Griffith (1988a), and Anselin and Griffith (1988), this measure of fit is biased in the presence of spatial error autocorrelation, a common feature in cross-sectional data. Even in an ESA approach it will therefore be important to check for the validity of the assumption of independence, which is at the basis of the standard interpretation of $R^2$.

Testing for possible misspecification of the assumed model is standard in confirmatory data analysis, that is, the situation where the model is formulated first and then validated or falsified by means of observed data [termed model-driven spatial data analysis in Anselin (1988a)]. In fact, one of the central themes in the discipline of econometrics, as argued by, for example, Hendry (1980), Malinvaud (1981), and Godfrey (1988) is the development of misspecification tests that can indicate deviations from the various assumptions that underlie classical regression analysis.

Deviations from the standard assumption of independence in regression analysis are typically considered in the context of time series data. However, as argued in Anselin and Griffith (1988) cross-sectional or spatial dependence can be just as much a problem in applied empirical work in geographical analysis. In fact, many have argued that this notion of dependence is at the core of geographical analysis [for example, the so-called first law of geography in Tobler (1979)]. It is especially relevant when cross-sectional data are collected for contiguous and aggregate spatial units (for example, data for states, counties, or census tracts). As is well known, this type of dependence is often referred to as spatial autocorrelation (Cliff and Ord 1973) or network autocorrelation (Doreian 1982), although the notion of autocorrelation is more limited than that of dependence. Spatial dependence is somewhat similar to the case of block dependencies in cross-sectional survey data (King and Evans 1986), but differs in its emphasis on exploiting the specific form of the spatial structure of the dependence in the design of tests and estimators.

The spatial dependence occurs in two different forms. In one, it affects the error terms only and is mostly considered to be a nuisance which needs to be eliminated. It is thus a special form of a nonspherical disturbance (King 1987; Anselin 1988a). For example, such spatial error autocorrelation could follow from the poor match between the spatial extent of the phenomenon of interest (for instance, a labor or housing market) and the administrative units for which data are available [for an extensive discussion, see Arbia (1989)]. As in the general case of nonspherical disturbances, spatial error autocorrelation does not cause ordinary least squares (OLS) estimates to be biased, but it alters their efficiency (variance). Consequently, inference (that is, interpretation of significance) that is based on the standard t-tests and measures of fit ($R^2$) will be biased. Moreover, the presence of spatial error dependence affects the validity of a number of standard misspecification tests, such as tests for heteroskedasticity (Anselin and Griffith 1988) and tests for structural stability of the regression coefficients (Anselin 1990a).

In its second form, the spatial dependence is given a substantive interpretation, in that a variable of interest at one location is jointly determined by its values at other locations. For example, this would be the case in a spatial price equilibrium model (Haining 1984), or when a spatial diffusion process is studied (O'Loughlin 1986; O'Loughlin and Anselin 1991). The correct model then should include a spatial autoregressive term, that is, a weighted sum of the values of the dependent variable at other locations (a so-called spatially lagged dependent variable). In mathematical sociology and anthropology, this is called the effects model of network autocorrelation (Doreian 1980); in geography, regional science, and political science it is called a mixed regressive spatial autoregressive model (Anselin 1988a). In contrast to the well-known result for the time series equivalent of this model, OLS is no longer a consistent estimator, whether the error terms are
correlated or not. This is due to the special character of the two-dimensional and two-directional dependence in space (Ord 1975; Cliff and Ord 1981; Anselin 1988a; Griffith 1988a).

An important issue in empirical spatial analysis is how one can detect the presence of these spatial effects, and, moreover, how one can distinguish between spatial dependence as a nuisance and a substantive spatial process (Doreian 1980; Anselin and Griffith 1988). A well-known test for spatial autocorrelation in the regression error term was developed by Cliff and Ord (1972). It is based on a generalization of Moran’s $I$ statistic to regression residuals and achieves asymptotic normality. King (1981) showed that Moran’s $I$ test is locally best invariant (LBI) for both spatial autoregressive and spatial moving average errors in general, and uniformly most powerful invariant (UMPI) in a special case. Burridge (1980) demonstrated the similarity of Moran’s $I$ test for regression residuals to a Lagrange multiplier (LM) statistic. This LM statistic is included within a general framework for testing spatial dependence and spatial heterogeneity (heteroskedasticity) in Anselin (1988b), following the approach of Bera and Jarque (1982). Anselin (1988b) also derived a Lagrange multiplier statistic to test for the presence of a spatially lagged dependent variable, as a special case of an omitted variable test. An attractive feature of both the LM and Moran tests is that they are based on estimates obtained from ordinary least squares estimation (OLS), that is, when the null hypothesis of no spatial autocorrelation is assumed to be valid. Other familiar asymptotic approaches, such as the likelihood ratio test are based on the estimates for the alternative model with spatial autocorrelation. Unlike the situation for time series, these alternative models almost always require maximum likelihood estimation, involving a nonlinear likelihood function (Ord 1975; Griffith 1988a; Anselin 1988a). Not only is such estimation more complex, but the necessary software to carry it out is not (yet) widely available [for some recent efforts in developing the appropriate software, see Griffith (1988b, 1989) and Anselin (1990b)].

In contrast to the vast literature on serial correlation for time series data [for example, the review in King (1987)], there are only a few published simulation studies of spatial effects in a linear regression. As illustrated in Table 1, only two of these, Bartels and Hordijk (1977) and Brandsma and Ketellapper (1979) focus explicitly on the power of tests for spatial error autocorrelation. Other studies address the influence of boundary effects (Griffith and Amrhein 1983; Griffith 1985, 1988a), and the choice of the spatial weights matrix (Stetzer 1982; Anselin 1986a). In Anselin and Griffith (1998) and Anselin (1990a) the effect of spatial error autocorrelation on standard tests for heteroskedasticity and structural stability is assessed. All studies point to the significance of spatial effects and stress the importance of properly specifying the form of the dependence or heterogeneity. Unfortunately, few general conclusions can be drawn from them, since they consider only a limited number of situations, as summarized in Table 2. Moreover, in all but a few, the number of replications that is used in the experiment is small (less than five hundred). Since the empirical rejection frequency ($q$) is a proportion, its sample standard deviation is the square root of $q(1 - q)/R$, where $R$ is the number of replications in the Monte Carlo experiment. For a $q$ of 0.05, this sample standard deviation is fairly large, unless the number of replications exceed one thousand. For example, for $R = 50$, the associated standard deviation would be 0.03, which implies that all rejection frequencies between 0.02 and 0.08 would lie within one standard deviation of 0.05 (the standard deviation is 0.02 for one hundred replications, 0.0097 for five hundred, 0.0069 for one thousand and 0.0031 for five thousand replications). The large standard deviation of the empirical frequencies in most previous studies does not allow for very precise conclusions to be formulated.
TABLE 1
Selected Simulation Studies of Spatial Effects in a Linear Regression

<table>
<thead>
<tr>
<th>Study</th>
<th>Issue</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartels and Hordijk (1977)</td>
<td>Tests for spatial error autocorrelation; Moran for OLS residuals, BLUS, RELUS</td>
<td>Moran for OLS residuals most power</td>
</tr>
<tr>
<td>Brandsma and Ketellapper (1979)</td>
<td>Tests for spatial error autocorrelation; Moran for OLS residuals, BLUS, LUS, Likelihood Ratio</td>
<td>Moran for OLS residuals most power</td>
</tr>
<tr>
<td>Griffith and Amrhein (1983); Griffith (1985, 1988a)</td>
<td>Boundary effects for spatial autocorrelation coefficient</td>
<td>Significant boundary effect; no satisfactory remedy</td>
</tr>
<tr>
<td>Stetzer (1982)</td>
<td>Choice of spatial weight matrix in space-time forecasting</td>
<td>Importance of weight choice; problems less for large N; presence of spatial autocorrelation magnifies problem</td>
</tr>
<tr>
<td>Anselin (1986a)</td>
<td>Non-nested tests on spatial weight matrix</td>
<td>Poor power of tests, but fairly robust to non-normality</td>
</tr>
<tr>
<td>Anselin and Griffith (1988)</td>
<td>Effect of spatial error autocorrelation on tests against heteroskedasticity</td>
<td>Significant effect, but not equal for all tests; tendency to over-reject under $H_0$; positive autocorrelation decreases power; LM-based pretest approach performs well</td>
</tr>
<tr>
<td>Anselin (1990a)</td>
<td>Effect of spatial error autocorrelation on tests for structural stability in single regression and SUR</td>
<td>Size and power affected; difference positive and negative spatial autocorrelation; ML and LM-based pretest perform well in larger samples</td>
</tr>
</tbody>
</table>

In this paper we present the results of the first extensive Monte Carlo simulation study of the properties of tests for spatial dependence. We stress the importance of this empirical approach since analytical attempts at determining exact distributions, critical value approximations, or other indicators of small sample performance are very difficult to apply in the case of spatial dependence, as we point out in more detail in the next section. Earlier empirical illustrations of the tests can be found in Anselin (1988b), for a regression using the familiar Irish county data [see also Cliff

TABLE 2
Selected Technical Aspects of Previous Simulation Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>R</th>
<th>N</th>
<th>Weight</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartels and Hordijk (1977)</td>
<td>100</td>
<td>39</td>
<td>irregular (2)</td>
<td>0, .1, .3, .5, .7, .9</td>
</tr>
<tr>
<td>Brandsma and Ketellapper (1979)</td>
<td>100</td>
<td>39</td>
<td>irregular (1)</td>
<td>0, .1, .3, .5, .7, .9</td>
</tr>
<tr>
<td>Griffith (1983), Griffith (1985)</td>
<td>50</td>
<td>64</td>
<td>rook</td>
<td>+/− .1, .5, .9</td>
</tr>
<tr>
<td>Griffith (1988a)</td>
<td>500</td>
<td>64</td>
<td>rook</td>
<td>+/− .1, .5, .9</td>
</tr>
<tr>
<td>Stetzer (1982)</td>
<td>200</td>
<td>49,196</td>
<td>various (9)</td>
<td>.95</td>
</tr>
<tr>
<td>Anselin (1986a)</td>
<td>200</td>
<td>24</td>
<td>irregular (2)</td>
<td>.25, .75</td>
</tr>
<tr>
<td>Anselin and Griffith (1988)</td>
<td>1000</td>
<td>25,50,75</td>
<td>queen</td>
<td>0, +/− .1, .5, .9</td>
</tr>
<tr>
<td>Anselin (1990a)</td>
<td>500</td>
<td>25,50,75</td>
<td>queen</td>
<td>0, +/− .1, .5, .9</td>
</tr>
<tr>
<td>Anselin (1990a)</td>
<td>300</td>
<td>25,50,75</td>
<td>queen</td>
<td>0, +/− .1, .5, .9</td>
</tr>
</tbody>
</table>

R is the number of replications; N is the sample size; weight is the type of spatial weights matrix with the number of different spatial layouts in parentheses; lattices are irregular, or regular, with rook or queen contiguity; autocorrelation is the range of values for the spatial autocorrelation coefficient.
and Ord (1981) for examples using the same data set), and in O'Loughlin and Anselin (1991), in a spatial analysis of interstate conflict between African countries. The latter in particular illustrates the type of issue confronted when testing for spatial dependence. For a regression of an index of total conflict on a number of country characteristics (for a cross-section of forty-two African states), they report the results for the three tests (O'Loughlin and Anselin 1991, Table 4): a Moran's I z-value of 2.80 (as a normal variate); a value for the LM test for error dependence (LM error) of 3.36; and a value for the LM test for a spatial lag (LM lag) of 6.15. As elaborated below, both LM tests are distributed as a $\chi^2$ variate with one degree of freedom, with critical values of 3.84 ($p = 0.05$) and 6.63 ($p = 0.01$). The LM test for a spatial lag is highly significant, but so is clearly Moran's I for spatial error autocorrelation. However, the LM test for spatial error autocorrelation is not significant (for $p = 0.05$). The question is then whether the LM error test has lower power than Moran's I (that is, is not able to pick up spatial error autocorrelation), or whether Moran's I is unable to distinguish between spatial error dependence and a spatial lag (since LM lag is highly significant). This issue cannot be resolved without a maximum likelihood estimation of the respective alternative models and a formal model discrimination [see Anselin (1988c); in O'Loughlin and Anselin (1991) it follows that the model with a spatial lag is the preferred specification in this instance]. Since we have no theoretical guidance with respect to the relative power of the three tests, both alternative models need to be estimated, instead of only one (the relevant one).

The general objectives of the paper are two-fold. In the first place, we wish to assess the extent to which the asymptotic properties of the tests (Moran and LM) are reflected in a wide range of situations, for different sample sizes, alternative spatial structures, and in the presence of nonstandard error distributions. Secondly, we want to gain insight into the relative power of the Moran versus the Lagrange multiplier tests. Specifically, we are interested in the power of these tests to discriminate between spatial error autocorrelation and a spatial lag, under a variety of circumstances with respect to sample size, form of spatial dependence, and boundary effects. In the design of our experiments, we have tried to cover a wide range of situations, in order to remedy the current gap in the literature with respect to empirical evidence on the properties of tests for spatial dependence.

In the remainder of the paper, we first briefly outline in formal terms the Moran's I and Lagrange multiplier tests. Next we describe the design of the simulation experiments and summarize the main results. In our conclusions we formulate some recommendations and guidelines to follow when applying tests for spatial autocorrelation in empirical work.

### 2. TESTS FOR SPATIAL DEPENDENCE IN A SINGLE EQUATION LINEAR REGRESSION

The null hypothesis for all tests is the standard linear regression model:

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

(1)

with $y$ as an $N$-by-1 vector of observations on the dependent variable, $X$ as a nonstochastic $N$-by-$K$ matrix of observations on the explanatory variables, $\beta$ as a $K$-by-1 vector of coefficients, and $\epsilon$ as an $N$-by-1 vector of error terms. Typically, the $\epsilon$ are assumed to be i.i.d., but of course for the alternative of spatial error autocorrelation, this is no longer the case. The error term can then be specified to follow a spatial autoregressive or a spatial moving average form. Without loss of generality we will concentrate on the autoregressive form only, since the tests for
both alternatives are the same (Burridge 1980). A spatial autoregressive model in the error term is then

$$\epsilon = \lambda W\epsilon + \mu$$

(2)

where $\lambda$ is the spatial autoregressive coefficient, $W$ is an $N$-by-$N$ spatial weights matrix and $\mu$ is an $N$-by-1 vector of i.i.d. errors. The spatial matrix $W$ expresses the strength of potential interaction between each observation and its neighbors. It can be based on simple binary contiguity coefficients ($w_{ij} = 1$ if $i$ and $j$ have a common border), or in general on any meaningful indicator, such as length of the common border, inverse distance, etc. [see Cliff and Ord (1973, 1981) for an extensive discussion]. Its main function is to construct a weighted average of neighboring values, the so-called spatial lag, $\sum_{j} w_{ij} \cdot \epsilon_j$, $\forall j \in J_i$, where $J_i$ is the set of neighbors for $i$. This terminology is somewhat misleading since there is no lag operator involved in the same sense as in time series analysis, but instead a weighted average is applied, similar to the notion of a distributed lag. In space, there is no unique concept that specifies which locations are considered to be lagged with respect to a given position, hence the rather awkward way of constructing a spatially lagged variable.

In regression analysis, the weights matrix is typically used in row-standardized form, that is, the row elements sum to one. As a consequence, the matrix $W$ will be asymmetric and its eigenvalues will be less than one in absolute value. Obviously, the proper choice of a spatial weights matrix is an important problem in spatial analysis, and one which has not yet obtained a satisfactory solution (Anselin 1988a). In our analysis, we will therefore assume the matrix to be known and nonstochastic, although we will consider three different specifications for its structure.

The second alternative hypothesis is that of a mixed regressive spatial autoregressive model, with a spatial lag in the dependent variable, $Wy$:

$$y = \rho Wy + X\beta + \mu$$

(3)

where $\rho$ is a spatial autoregressive coefficient, $\mu$ is an $N$-by-1 vector of i.i.d. errors, and the other notation is as before. This specification is difficult to distinguish from the spatial error dependence, since the latter can also be expressed as an equation with a spatial lag, similar to the common factor model in time series. Clearly, (2) can also be written as

$$\epsilon = (I - \lambda W)^{-1}\mu$$

(4)

which, after substitution in (1) and premultiplying all terms by $(I - \lambda W)$ yields

$$y = \lambda Wy + X\beta - \lambda WX\beta + \mu$$

(5)

The discrimination between alternatives (1)-(2) and (3) could be based on tests for coefficient restrictions in (5), that is, on whether the product of the estimates for $\lambda$ and $\beta$ equals the estimates for $\lambda \beta$. This is the so-called common factor hypothesis outlined in Burridge (1981), Blommestein (1983), Bivand (1984), and Anselin (1988a). In spatial analysis this approach is not very practical, due to the need for nonlinear optimization to estimate (5) (Ord 1975; Griffith 1988a; Anselin 1988a).

As argued before, due to the added complexity of maximum likelihood estimation and the lack of software available to the general user, we ideally want to base all tests on the results of a simple OLS regression of (1).
As pointed out in the introduction, the first class of tests is based on the application of Moran's $I$ to the OLS residuals of (1) suggested by Cliff and Ord (1972, 1973, 1981):

$$I = (N/S)e'W e / e'e$$  \hspace{1cm} (6)

where $e$ are the OLS residuals, $N$ is the number of observations and $S$ is the sum of all elements in the spatial weights matrix. For a row-standardized $W$ this simplifies to $N$, so that $N/S = 1$ and can be ignored. The resulting form,

$$I = e'W e / e'e$$  \hspace{1cm} (7)

shows a striking similarity to the familiar Durbin-Watson statistic (Durbin and Watson 1950), with $W$ taking the place of the $A$ band matrix in the Durbin-Watson test (for example, King and Evans 1985). For normal error terms the distribution of the standardized Moran statistic is shown to be asymptotically normal. In order to carry out an operational test, both the expected value and the variance of $I$ are needed. They are derived in Cliff and Ord (1981, pp. 202-203):

$$E[I] = N \cdot \text{tr}(MW)/(N - K) \cdot S$$

with $\text{tr}$ as the matrix trace operator, $N$, $W$, $K$, and $S$ as before, and $M = I - X(X'X)^{-1}X'$. The variance is

$$V[I] = \left\{N^2/[S^2(N - K)(N - K + 2)]\right\} \cdot \left\{S_1 + 2\text{tr}[(MW)^2] - \text{tr}B - 2[\text{tr}(MW)]^2/(N - K)\right\}$$

with $S_1 = 1/2 \Sigma_i \Sigma_j (w_{ij} + w_{ji})^2$ and $B = (XX)^{-1}X(W + W')^2X$. As an alternative to this approach, a test could be based on a randomization principle (Cliff and Ord 1981, pp. 45-46). The randomization assumption is that each value can equally likely be observed at each location. Consequently, an empirical distribution for the Moran's $I$ can be constructed from many permutations of the observed values [see, for example, Boots and Kanaroglou (1988) for an application to the residuals of a logit model]. This distribution can also be approximated by a normal, but with slightly different expressions for mean and variance (see Cliff and Ord 1981, pp. 45-46). However, in a regression context, the randomization approach is inappropriate, since regression residuals are correlated by construction: $e = Me$, where the idempotent matrix $M$ (as above) is not diagonal. Although it is sometimes applied to regression residuals (see, for example, Cliff and Ord 1981, p. 210), it really should not be. In Anselin and Rey (1990), we show how the randomization approach with Moran's $I$ for regression residuals is unreliable: its empirical rejection frequency is significantly different from the nominal level, and its distribution under the null hypothesis of no spatial dependence does not conform to the normal.\footnote{Due to space limitations, we did not include the results for the randomization approach in the current paper. A detailed discussion can be found in Anselin and Rey (1990)}

For a normal error term, the exact distribution of the Moran statistic (7) can be derived for any finite sample, using the same approach as for the Durbin-Watson test. However, in the spatial case, the critical values will depend not only on $X$ (as for the D-W test), but also on the asymmetric matrix $W$, which is clearly not very practical. Approximations of the critical values can be based on the approaches
discussed by Evans and King (1985), but they will also depend on both $X$ and $W$, and thus suffer from the same problem.

The Lagrange multiplier tests for spatial dependence of Burridge (1980) and Anselin (1988b) fall within the domain of maximum likelihood inference and thus have attractive asymptotic properties. These are based on various central limit theorems for dependent series, which can be applied to spatial processes, provided that the form of spatial dependence satisfies certain conditions, as discussed in more detail in Anselin (1988a, chap. 5). These conditions are satisfied for most spatial processes with a sparse weights matrix that is row-standardized (which is the standard approach). Both LM tests are asymptotically distributed as $\chi^2$ with 1 degree of freedom. They are fairly easy to calculate and require only some additional matrix traces and weighted cross-products. The LM test for spatial error autocorrelation is

$$\text{LM}_{\text{ERR}} = \left[ N \cdot e'Ve / e'e \right]^2 \left[ \text{tr}(WW + W^2) \right]^{-1}$$

which is nothing but a scaled Moran's $I$. The LM test for a spatially lagged dependent variable is

$$\text{LM}_{\text{LAG}} = \left[ N \cdot e'Wy / e'e \right]^2 \left[ N(WXb)'M(WXb) / e'e + \text{tr}(WW + W^2) \right]^{-1}$$

where $b$ is the OLS estimate for $\beta$ in (1), and the other notation is as before.

There are no results on an exact distribution for the LM statistics. Approximations could be based on the expansions outlined in Harris (1985) for the score statistic in general, but its application is very tedious. Moreover, the simplifications that allowed Honda (1988) to derive a size correction for the heteroskedastic case do not hold for the spatial tests. As a result, insight into the small sample properties of these tests can only be gained from an empirical approach based on Monte Carlo simulations. The design of these experiments is outlined next.

3. DESIGN OF THE EXPERIMENTS

A number of Monte Carlo simulation experiments were carried out to compare the Moran's $I$ and LM tests in a variety of empirical settings. In all, we consider 1,486 different situations. We study the distribution of the tests under the null hypothesis of no spatial dependence and the power of the tests for different values of the spatial autoregressive coefficients $\lambda$ (for error dependence) or $\rho$ (for a spatial lag), for a range of sample sizes. The spatial layout of the data set is a system of regular square lattices of increasing size, centered at the same origin. The smallest lattice is of order 5-by-5 (for $N = 25$), and the others have dimensions increasing in all directions by 1, centered at the same origin, that is, the central cell of the 5-by-5 lattice (a 7-by-7 lattice, a 9-by-9 lattice, etc.), up to a 15-by-15 lattice ($N = 225$). This results in a total of six different sample sizes.

We also focus on the sensitivity of the size and power of the tests to the choice of the spatial weights matrix $W$. Specifically, we consider three different binary weights matrices, all in row-standardized form: $W_1$, first-order contiguity according to the $\text{rook}$ criterion (that is, the cells immediately above, below, to the right, and to the left, for a total of four neighboring cells); $W_2$, first-order contiguity according to the $\text{queen}$ criterion (that is, the eight cells immediately surrounding
the central cell); and $W_3$, a distance-based first-order contiguity (that is, all cells within a distance of three are considered to be contiguous, for a maximum of twenty-eight neighboring cells). The first two matrices correspond to a criterion where the existence of common borders is most important, whereas the third matrix is based on a notion of interaction as a function of distance in all directions. The choice of these matrices is not arbitrary since they represent an increasing connectedness of each cell with the other cells in the system. As discussed in detail in Anselin (1988a, chap. 51), a limited extent of interconnectedness (that is, not increasing with sample size) is one of the requirements for the central limit theorems on dependent spatial series to hold. In other words, everything else being the same, the shorter the range, the more likely should it be for the asymptotic properties to be achieved in smaller samples.

The choice of a regular lattice rather than an irregularly shaped layout was a conscious one. The relative merits of both approaches are discussed at length in Haining (1986) and Anselin (1986b). For this particular application, we agree with Haining (1986) and take a regular lattice approach, to control as much as possible for spatial layout. Whereas at first sight such a layout may seem to differ from the shapes of many administrative units encountered in socioeconomic research, it is not all that unrealistic. In fact, there is a considerable body of analysis where data from irregular units are first transformed into regular units or grids [see Bronars and Jansen (1987) for a recent example of a study where county employment data are transformed to square grids]. Also, this is routinely the case in a raster-based GIS, for instance, as used with remotely sensed data. It should be noted that this very transformation is likely to induce spatial dependence.

The specification under the null hypothesis is a simple linear regression:

$$ y = \alpha + \beta x + \mu $$

(10)

with $\alpha = \beta = 1$, $x$ as an $N$-by-1 vector of uniform random variates (between 0 and 10) and $\mu$ as a standard i.i.d. normal error term. The same 225-by-1 $x$ vector is used in all replications, with the smaller sample sizes taken as increasing subsets of this vector. For each situation five thousand replications are analyzed. A nominal Type I error of 0.05 is used throughout. As pointed out in the introduction, the associated sample standard deviation in each simulation run is 0.0031. In other words, two standard deviations of a sample proportion equal to the nominal level of 0.05 would be less than 0.01, which we considered to be small enough to be able to adequately test significant deviations between the empirical rejection frequency and the nominal Type I error (as a standard test on a sample proportion).

A vector of spatially autoregressive error terms is obtained from the standard normal variates $\mu$ as

$$ \epsilon = (I - \lambda W)^{-1} \mu $$

for values of $\lambda$ ranging from -0.9 to 0.9 and for the three spatial weights matrices. Values for the $N$-by-1 vector $y$ are then found as

$$ y = \alpha + \beta x + \epsilon $$

(11)

Note that the number of contiguous cells will be less than the maximum for locations close to the boundaries of the lattice. This is especially important for the smaller sample sizes.

All random numbers are based on the uniform and normal generators in Gauss386 (Aptech Systems) and generated on a 25 Mhz IBM PS/2/70 with a 80387 numerical coprocessor. All computations are carried out in Gauss386, using the SpaceStat software described in Anselin (1990b).
Observations for the second alternative of a spatial lag are constructed as

\[ y = (I - \rho W)^{-1}(\alpha + \beta x + \mu) \]  

(12)

again, with \( \rho \) as the spatial autoregressive coefficient ranging from \(-0.9\) to \(0.9\) and for the three spatial weights matrices.

In order to assess boundary effects, the alternatives (11) and (12) are generated in two ways. In the first, there are no boundary effects. In other words, each sample is considered to be an isolated region. In technical terms, this means that the \( W \) matrix used to generate the dependent variates is also the one used in the tests. In the second approach, boundary effects are created by generating the dependent variates for the largest sample size only \((N = 225)\). Following the procedure outlined in Griffith (1988a), the observations for the smaller samples are simply taken as subsets of the largest set and thus correctly include spillovers across the boundaries. However, the spatial weights matrix used in the tests ignore these boundary effects and are the same as in the nonboundary case. In addition, we also analyze the consequences of misspecifying the spatial weights matrix in the tests. In order to achieve this, we generate the spatial dependence using \( W_1 \) and carry out the tests using \( W_3 \). This represents a situation where the contiguity criterion for the test is based on a distance measure, whereas the actual dependence is purely a border effect (or, in other words, in the test too many cells are considered to be contiguous to each observation).

A final issue we study is the effect of nonstandard error terms on the distribution of the statistics under the null hypothesis. We consider two non-normal error distributions: the lognormal and the exponential, both with variance 1. In addition, we also assess the effect of the presence of heteroskedasticity in the form of two and four spatial regimes with different error variance. These are referred to below as HET1 and HET2. In the first, we make a distinction between the upper half of cells in each lattice (with error variance 1) and the lower half (with error variance 4). In the second case, we distinguish four regimes, northeast (error variance 1), northwest (error variance 2), southwest (error variance 3), and southeast (error variance 4). These spatial regimes are very similar to positive spatial autocorrelation although they pertain to spatial heterogeneity. This is similar to the problem of discriminating true from apparent contagion in the analysis of spatial point patterns. As shown in Anselin (1990a), tests for structural instability for regimes of this type are seriously affected by the presence of spatial autocorrelation.

4. EMPIRICAL RESULTS

In this section, we present a summary of the results of the Monte Carlo experiments. The complete set of results, with extensive tables and figures, can be found in Anselin and Rey (1990). In order not to overburden the reader, only the most representative of the figures and tables are presented here. We consider five issues in turn. The first pertains to the properties of the tests under the null hypothesis of no spatial dependence, that is, whether size of the test, the nominal Type I error \(0.05\) in the simulations, and the asymptotic distributions (respectively the normal and the chi square with one degree of freedom) are reflected in the empirical results. The second and third issues relate to the performance of the tests when spatial dependence is present, that is, their relative power to reject the null hypothesis when it is not true, for both spatial error dependence and spatial lag

\footnote{We used the procedures outlined in Naylor et al. (1966), pp. 81–82 and pp. 99–101, based on the same uniform random numbers and generated by means of a Gauss386 routine.}
TABLE 3
Size of Tests, Normal Error Model (z values given below empirical rejection frequencies)

<table>
<thead>
<tr>
<th>N</th>
<th>M1</th>
<th>LM Err</th>
<th>LM Lag</th>
<th>M1</th>
<th>LM Err</th>
<th>LM Lag</th>
<th>M1</th>
<th>LM Err</th>
<th>LM Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.049</td>
<td>0.038</td>
<td>0.052</td>
<td>0.049</td>
<td>0.020</td>
<td>0.047</td>
<td>0.048</td>
<td>0.002</td>
<td>0.021</td>
</tr>
<tr>
<td>49</td>
<td>0.052</td>
<td>0.045</td>
<td>0.050</td>
<td>0.045</td>
<td>0.034</td>
<td>0.048</td>
<td>0.049</td>
<td>0.015</td>
<td>0.035</td>
</tr>
<tr>
<td>81</td>
<td>0.056</td>
<td>0.054</td>
<td>0.053</td>
<td>0.052</td>
<td>0.042</td>
<td>0.052</td>
<td>0.052</td>
<td>0.027</td>
<td>0.042</td>
</tr>
<tr>
<td>121</td>
<td>0.049</td>
<td>0.044</td>
<td>0.046</td>
<td>0.054</td>
<td>0.047</td>
<td>0.051</td>
<td>0.049</td>
<td>0.031</td>
<td>0.043</td>
</tr>
<tr>
<td>169</td>
<td>0.051</td>
<td>0.050</td>
<td>0.048</td>
<td>0.049</td>
<td>0.045</td>
<td>0.053</td>
<td>0.050</td>
<td>0.037</td>
<td>0.049</td>
</tr>
<tr>
<td>225</td>
<td>0.050</td>
<td>0.047</td>
<td>0.048</td>
<td>0.055</td>
<td>0.051</td>
<td>0.046</td>
<td>0.047</td>
<td>0.042</td>
<td>0.042</td>
</tr>
</tbody>
</table>

TABLE 4
Size of Tests, Alternative Error Models (W1) (z values given below empirical rejection frequencies)

<table>
<thead>
<tr>
<th>N</th>
<th>Lognormal M1</th>
<th>LM Err</th>
<th>LM Lag</th>
<th>Exponential M1</th>
<th>LM Err</th>
<th>LM Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0382</td>
<td>0.018</td>
<td>0.0554</td>
<td>0.0410</td>
<td>0.0300</td>
<td>0.0538</td>
</tr>
<tr>
<td>49</td>
<td>0.0390</td>
<td>0.0348</td>
<td>0.0568</td>
<td>0.0398</td>
<td>0.0332</td>
<td>0.0488</td>
</tr>
<tr>
<td>81</td>
<td>0.0466</td>
<td>0.0438</td>
<td>0.0524</td>
<td>0.0444</td>
<td>0.0398</td>
<td>0.0542</td>
</tr>
<tr>
<td>121</td>
<td>0.0436</td>
<td>0.0418</td>
<td>0.0488</td>
<td>0.0442</td>
<td>0.0398</td>
<td>0.0464</td>
</tr>
<tr>
<td>169</td>
<td>0.0440</td>
<td>0.0430</td>
<td>0.0528</td>
<td>0.0456</td>
<td>0.0422</td>
<td>0.0518</td>
</tr>
<tr>
<td>225</td>
<td>0.0498</td>
<td>0.0484</td>
<td>0.0532</td>
<td>0.0430</td>
<td>0.0424</td>
<td>0.0508</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>HET1 M1</th>
<th>LM Err</th>
<th>LM Lag</th>
<th>HET2 M1</th>
<th>LM Err</th>
<th>LM Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0758</td>
<td>0.0642</td>
<td>0.0344</td>
<td>0.0648</td>
<td>0.0516</td>
<td>0.0374</td>
</tr>
<tr>
<td>49</td>
<td>0.0824</td>
<td>0.0726</td>
<td>0.0528</td>
<td>0.0700</td>
<td>0.0638</td>
<td>0.0498</td>
</tr>
<tr>
<td>81</td>
<td>0.0860</td>
<td>0.0812</td>
<td>0.0672</td>
<td>0.0670</td>
<td>0.0670</td>
<td>0.0612</td>
</tr>
<tr>
<td>121</td>
<td>0.0826</td>
<td>0.0756</td>
<td>0.0642</td>
<td>0.0644</td>
<td>0.0608</td>
<td>0.0546</td>
</tr>
<tr>
<td>169</td>
<td>0.0908</td>
<td>0.0924</td>
<td>0.0616</td>
<td>0.0724</td>
<td>0.0678</td>
<td>0.0554</td>
</tr>
<tr>
<td>225</td>
<td>0.0884</td>
<td>0.0848</td>
<td>0.0610</td>
<td>0.0680</td>
<td>0.0672</td>
<td>0.0582</td>
</tr>
</tbody>
</table>

dependence. The fourth and fifth issues deal with the importance of the weights matrix, either in the form of a border effect, or in the form of misspecified weights.

Distribution of Test Statistics under the Null Hypothesis

In Tables 3 and 4, we report the empirical rejection frequencies under the null hypothesis of no spatial dependence (i.e., the standard regression model) of the Moran’s I test (normal assumption) and the two Lagrange multiplier tests (LM error and LM lag), for a nominal rejection probability of \( p = 0.05 \). We also list the results of a test on the equality of the nominal and empirical rejection frequencies.\(^5\) The results for the normal error model are listed in Table 3, while Table 4 contains

\(^5\)This test is a simple test of the null hypothesis that the sample proportion equals 0.05.
the figures for the alternative error models: lognormal, exponential, and the two forms of heteroskedasticity (HET1 and HET2). An obvious feature is the difference in rejection found using $W_1$ and $W_2$ on one hand and $W_3$ on the other hand, particularly for the Lagrange multiplier tests: both significantly under-reject for $W_3$, but less so for the other weights matrices. While the LM error test significantly under-rejects for smaller samples with $W_1$ and $W_2$ ($N = 25, 121$ and $N = 25, 49, 81$, respectively), this is not the case for the LM lag test, which does not have an empirical rejection frequency that is significantly different from 0.05 for these weights matrices. The Moran's $I$ statistic based on the normal assumption has the best overall performance in terms of empirical size (no rejections at $p = 0.05$). However, for a lognormal and exponential error distribution (in Table 4, for $W_1$ only) this desirable property is not maintained, and Moran's $I$ significantly under-rejects the null hypothesis for four out of the six sample sizes. The two LM statistics are affected differently: while the LM error test significantly under-rejects for both lognormal and exponential errors (except for $N = 225$, lognormal), the LM lag test is much more robust to these deviations from normality (only one significant rejection, $N = 49$, lognormal). A somewhat similar result is obtained for the heteroskedastic errors, where LM lag is slightly less affected than the other tests. Clearly, the presence of heteroskedasticity rather than dependence can cause the tests to give a misleading indication.

We next focused on more extensive characteristics of the distribution of the test statistics under the null. For the three weights matrices in the normal error case, and for the four nonstandard error distributions, the Moran's $I$ statistic was tested for normality, using the Kiefer and Salmon (1983) approach.\footnote{The detailed values of this test and associated marginal probabilities are given in Anselin and Rey (1990).}

For $W_1$, the
empirical distribution of Moran's $I$ cannot be differentiated from the normal. However, a striking difference occurs for $W_a$ and $W_b$, where the normal is strongly rejected, for all sample sizes. This is illustrated in Figure 1, which shows the differences between the empirical frequency distributions for the three spatial weights matrices (for $N = 25$). The non-normal and heteroskedastic errors have a similar effect, and the normality of Moran's $I$ is not maintained (to a lesser extent for HET2).

We also assessed the extent to which the Lagrange multiplier statistics conformed to a chi-square distribution, by means of a Kolmogorov-Smirnov goodness-of-fit test. Overall, we find an acceptable performance for all but the smallest sample sizes, with less sensitivity to the choice of the weights matrix for LM lag than for LM error. LM lag is also robust against the presence of an exponential error distribution, and less affected by the lognormal. However, for large sample sizes both LM tests are robust. On the other hand, heteroskedasticity invalidates the chi-square distribution for both LM tests, though slightly less for LM lag (more for HET1 than HET2).

In sum, under ideal circumstances (normal error) and when $W_1$ is used as the spatial weights matrix, the Moran's $I$ test performs closest to what its asymptotic distribution would indicate. However, its sensitivity to the choice of the weights matrix and to the presence of non-normality of the errors is troublesome. We could speculate that this could be due to the fact that $W_a$ and $W_b$ represent a higher degree of connectedness between the cells, and thus would need larger sample sizes to approach the asymptotic properties of the test [this is related to the mixing conditions of dependent spatial processes; see Anselin (1998a, chap. 5)], but this issue merits further investigation, which is beyond the scope of the current paper. Both LM tests are also sensitive to the weights matrix in terms of size, but less so in terms of overall distribution. The LM lag statistic in particular is less affected and seems fairly robust against exponential and lognormal errors.

**Power of Test Statistics against Spatial Error Dependence**

The complete results for the rejection frequencies for Moran's $I$ and the LM tests obtained in the presence of spatial error dependence are listed in Anselin and Rey (1990). We summarize here a number of interesting characteristics that are reflected in these results. First, the power of the tests varies considerably with the choice of the weights matrix. Specifically, a striking lack of symmetry is obtained between negative and positive values of $\lambda$ for $W_a$ and $W_b$, as illustrated in Figures 2–4 for Moran's $I$. Whereas for $W_1$ power increases with sample size and with the absolute value of $\lambda$ in a more or less symmetric fashion, this is less so for $W_a$ and clearly not the case for $W_b$. For both these weights matrices the power of the test against the negative alternative is considerably lower than for the positive one. Also, the power curve itself is less steep for $W_a$ and $W_b$. For example, for $W_1$ a 50 percent rejection frequency is obtained for values of $\lambda$ of approximately 0.15 for $N = 225$ and 0.5 for $N = 25$; for $W_a$ these values are 0.25 and 0.7; and for $W_b$ they are 0.35 for $N = 225$ while the 50 percent rejection is never reached for $N = 25$! The LM error shows a similar pattern. Its power is slightly higher than that of Moran's $I$ for negative values of $\lambda$, but slightly lower for positive values of $\lambda$, as can be seen from comparing Figure 5 to Figure 2 (for $W_1$). This power difference is smaller for large as well as small sample sizes of $\lambda$, with a peak for lower absolute values.

---

7 Detailed results are given in Anselin and Rey (1990).

8 It should be noted that positive (more spatial grouping than random) and negative (less spatial grouping than random) spatial autocorrelation are qualitatively different concepts, and not just opposites of the same phenomenon. A similar asymmetry of results between negative and positive spatial autocorrelation was also found in Anselin (1990a), for the queen concept of contiguity.
FIG. 2. Power of Moran's I (Normal Error Model, W1)

FIG. 3. Power of Moran's I (Normal Error Model, W2)
Fig. 6. Power of Moran's I (Normal Lag Model, W1)

Fig. 7. Power of LM Lag (Normal Lag Model, W1)
alternative [see Godfrey (1988) for a more elaborate discussion of misspecification tests].

Boundary Effects

We also considered the rejection frequencies when boundary effects are present, but ignored. They are consistently lower than those for the similar cases without boundary effects. The difference in power varies by test and by weights matrix. For example, for $W_1$ it is generally less than 0.2, but for $W_3$ it can be as high as 0.6. For Moran's $I$ ($W_1$), the difference is symmetric between positive and negative values of $\lambda$, with a peak somewhere around 0.5, but for lower absolute values of $\lambda$ with larger sample sizes. For LM lag a similar pattern is found, though slightly less symmetric and with the peak differences occurring for smaller values of the autoregressive coefficient. However, for $W_3$ the relation between $p$, $N$, and the power difference is much less regular. Again, this issue merits further attention from a theoretical perspective.

Misspecification of the Spatial Weights Matrix

A final issue is the effect of using the wrong spatial weights matrix to carry out the tests. In all cases a substantially lower power is obtained when $W_3$ is used compared to when the correct $W_1$ is used [see Anselin and Rey (1990) for details]. We found that the difference in power can be substantial, and is considerably

---

Fig. 5. Power of LM Error (Normal Error Model, W1)

---

10 Detailed results are given in Anselin and Rey (1990).

11 For positive $\lambda$, the peak difference for $N = 25$ is for $\lambda = 0.6$, for $N = 49$ for $\lambda = 0.5$, for $N = 81$ for $\lambda = 0.4$, and for $N = 121$ and 169 for $\lambda = 0.3$. 
Fig. 6. Power of Moran's I (Normal Lag Model, W1)

Fig. 7. Power of LM Lag (Normal Lag Model, W1)
larger than for the boundary effects. The power difference is smaller for positive spatial dependence than for negative, but otherwise shows a fairly symmetric pattern in terms of a decreasing peak with larger values of the sample size. For spatial error dependence, the power difference did not seem to decrease with \( N \), but rather slightly increases, in contrast to the boundary effects. This did not seem to be the case for the spatial lag.

5. CONCLUSIONS

As in any empirical analysis based on Monte Carlo simulations, the degree of generality of the conclusions is limited by the range of situations covered in the design of the experiments. In spatial analysis, the variety of layouts that is possible for the data points is virtually limitless, particularly for irregular lattice structures. Nevertheless, we think that our experiments provide some useful indication of the properties of tests for spatial dependence in regression models.

The most important finding is the sensitivity of the properties of the tests to the choice of the spatial weights matrix. The distribution under the null as well as the rejection frequencies in the presence of spatial dependence varied considerably with the selection of the spatial weights matrix. The most desirable properties were obtained for the first-order contiguity according to the rook criterion \( (W_1) \). With the queen criterion of contiguity \( (W_2) \) and the distance based contiguity \( (W_3) \), Moran’s \( I \) statistic no longer followed the normal distribution under the null hypothesis (although its empirical rejection frequency was not significantly different from the nominal value). Moreover, for \( W_2 \) and \( W_3 \) the power of all three tests was lower and showed a clear asymmetric pattern between negative and positive values of the autoregressive coefficient (most dramatically for \( W_3 \)). In addition, ignoring boundary effects and misspecifying the spatial weights matrix lowered the power of all tests in similar fashion although substantially more so in the latter case. In other words, considerable care needs to be taken when interpreting the results of these tests in applied work. In fact, the asymptotic properties of the tests may be unreliable for a poorly specified weights matrix in a finite data set. Since both boundary effects and misspecified weights matrices lead to lower rejection frequencies, it may be more appropriate in marginal cases to reject the null hypothesis.

We also obtained some insight into the robustness of the tests in conditions that differ from the standard normal error assumption. In the presence of lognormal and exponential errors, both Moran’s \( I \) and the LM error statistic were found to significantly under-reject the null hypothesis. In contrast, the LM lag test seemed fairly robust under these conditions. The presence of heteroskedasticity led to significant over-rejection of the null hypothesis for the two tests against error dependence, but much less for the LM lag statistic. Together with the results in Anselin and Griffith (1988) and Anselin (1990a) there is now ample evidence that heteroskedasticity and spatial autocorrelation each affect the properties of the tests against the other form of misspecification. This issue is important enough to warrant an approach in applied work in which both sources of misspecification are taken into account, for example, following the strategy suggested in Anselin (1988b).

The theoretical distribution of the statistics under the null was fairly well reflected in the small sample simulations. Except for the weights issue discussed above, Moran’s \( I \) followed the normal distribution. The asymptotic chi-square distribution of the LM lag statistic was only rejected for the smallest sample size \((N = 25)\). However, the LM error test did not follow the chi-square distribution in eight out of the eighteen cases and also showed significant under-rejection of the
null hypothesis. A size correction may therefore be necessary in order to obtain reliable inference in practice.

In terms of providing guidance in the choice between spatial error autocorrelation and a spatial lag, the Lagrange multiplier tests are the most appropriate. Moran's I has power against both alternatives and thus cannot be used to discriminate between the two. In contrast, the LM tests had highest power in the case for which they were designed. Moreover, the LM lag test had good power for all but the smallest data set and showed evidence of some degree of robustness. In sum, the simulations provided considerable evidence of the superiority of the Lagrange multiplier approach as a strategy for misspecification tests for spatial dependence. Since these tests are easily implemented with the results of a standard OLS regression, they should be used routinely in applied spatial data analysis.12

LITERATURE CITED


12 The LM tests are implemented in a user-friendly format in the SpaceStat software for spatial data analysis (Anselin 1990b).
Lac Anselin and Serge Rey


