:LENDING NOTES:
:RETURN TO:
:RETURN VIA:

27@.15

ILL record updated to IN PROCESS
Record 5 of 8

ILL pe
CAN YOU SUPPLY ? YES NO COND FUTUREDATE Record 5 of 8
:ILL: 8017102 :Borrower: GZM :ReqDate: 20030725 :NeedBefore: 20030908
:Status: IN PROCESS 20030725 :RecDate: :RenewalReq:
:OCLC: 38042956 :Source: OCLCILL :DueDate: :NewDueDate:
:Lender: *MNP,NUI,UIU,IAY,INU
:CALLNO: Stx HB137.H36 1998
:TITLE: Handbook of applied economic statistics /
:IMPRINT: New York : Marcel Dekker, c1998.
:SERIES: Statistics, textbooks and monographs ; v. 155
:ARTICLE: NEED ESSAY/CHAPTER ONLY: Anselin, Luc and Anil Bera: ''Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics''
:VOL: :NO: :DATE: 1998 :PAGES: 237-289
:VERIFIED: <TN:26697><ODYSSEY:216.54.57.190/GZM>OCLC [Format: Book]
:PATRON: <TN:26697>
:SHIP TO: ILL Borrowing/231 Memorial Library/Univ of Wisconsin Libraries/728 State St./Madison, WI 53706-1494/
:BILL TO: same FEIN 39-6006-492 CIC
:SHIP VIA: Best way :MAXCOST: WILS or $25ifm
:COPYRT COMPLIANCE: CCL
:FAX: 608 262-4649 ARIEL: 128.104.63.200 NOTE: NEW ARIEL IP ADDRESS
:E-MAIL: gzmill@library.wisc.edu
:BILLING NOTES: IFM preferred CIC

CIC

41-29

7

Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics

Luc Anselin
West Virginia University, Morgantown, West Virginia

Anil K. Bera
University of Illinois, Champaign, Illinois

I. INTRODUCTION

Econometric theory and practice have been dominated by a focus on the time dimension. In stark contrast to the voluminous literature on serial dependence over time (e.g., the extensive review in King 1987), there is scant attention paid to its counterpart in cross-sectional data, spatial autocorrelation. For example, there is no reference to the concept nor to its relevance in estimation or specification testing in any of the commonly cited econometrics texts, such as Judge et al. (1982), Greene (1993), or Poirier (1995), or even in more advanced ones, such as Fomby et al. (1984), Amemiya (1985), Judge et al. (1995), and Davidson and MacKinnon (1993) (a rare exception is Johnston 1984). In contrast, spatial autocorrelation and spatial statistics in general are widely accepted as highly relevant in the analysis of cross-sectional data in the physical sciences, such as in statistical mechanics, ecology, forestry, geology, soil science, medical imaging, and epidemiology (for a recent review, see National Research Council 1991).

In spite of this lack of recognition in "mainstream" econometrics, applied workers saw the need to explicitly deal with problems caused by spatial autocorrelation in cross-sectional data used in the implementation of regional and multiregional econometric models. In the early 1970s, the Belgian economist Jean Paelinck coined the term "spatial econometrics" to designate a field of applied econometrics dealing with estimation and specification problems that arose from this. In their classic book *Spatial Econometrics*, Paelinck and Klaassen (1979) outlined five characteristics of the field: (1) the role of spatial interdependence in spatial models; (2) the asymmetry in spatial relations; (3) the importance of explanatory factors located in other spaces; (4) differentiation between ex post and ex ante interaction; and (5) explicit modeling of space (Paelinck and Klaassen 1979, pp. 5–11; see also Hordijk and Paelinck 1976, Paelinck 1982). In Anselin (1988a, p. 7), spatial econometrics is defined more broadly as "the collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models." The latter incorporate regions, location and spatial interaction explicitly and form the basis of most recent empirical work in urban and regional economics, real estate economics, transportation economics, and economic geography. The emphasis on the model as the starting point differentiates spatial econometrics from the broader field of spatial statistics, although they share a common methodological framework. Much of the contributions to spatial econometrics have appeared in specialized journals in regional science and analytical geography, such as the *Journal of Regional Science, Regional Science and Urban Economics, Papers in Regional Science, International Regional Science Review, Geographical Analysis*, and *Environment and Planning A*. Early reviews of the relevant methodological issues are given in Hordijk (1974, 1979), Bartels and Hordijk (1977), Arora and Brown (1977), Paelinck and Klaassen (1979), Bartels and Ketellapper (1979), Cliff and Ord (1981), Blommestein (1983), and Anselin (1980, 1988a, 1988b). More recent collections of papers dealing with spatial econometric issues are contained in Anselin (1992a), Anselin and Florax (1995a), and Anselin and Rey (1997).

Recently, an attention to the spatial econometric perspective has started to appear in mainstream empirical economics as well. This focus on spatial dependence has occurred in a range of fields in economics, not only in urban, real estate, and regional economics, where the importance of location and spatial interaction is fundamental, but also in public economics, agricultural and environmental economics, and industrial organization. Recent examples of empirical studies in mainstream economics that explicitly incorporated spatial dependence are, among others, the analysis of U.S. state expenditure patterns in Case et al. (1993), an examination of recreation expenditures by municipalities in the Los Angeles region in Murdoch et al. (1993), pricing in agricultural markets in LeSage (1993), potential spillovers from public infrastructure investments in Holtz-Eakin (1994), the determination of agricultural land values in Benirschka and Binkley (1994), the choice of retail sales contracts by integrated oil companies in Pinkse and Slade (1995), strategic interaction among local governments in Brueckner (1996), and models of nations' decisions to ratify environmental controls in Beron et al. (1996) and Murdoch et al. (1996). Substantively, this follows from a renewed focus on Marshallian externalities, spatial spillovers, copy-catting, and other forms of behavior where an economic actor

mimics or reacts to the actions of other actors, for example in the new economic geography of Krugman (1991), in theories of endogenous growth (Romer 1986), and in analyses of local political economy (Besley and Case 1995). Second, a number of important policy issues have received an explicit spatial dimension, such as the designation of target areas or enterprise zones in development economics and the identification of underserved mortgage markets in urban areas. A more practical reason is the increased availability of large socioeconomic data sets with detailed spatial information, such as county-level economic information in the REIS CD-ROM (Regional Economic Information System) of the U.S. Department of Commerce, and tract-level data on mortgage transactions collected under the Housing Mortgage Disclosure Act (HMDA) of 1975.

From a methodological viewpoint, spatial dependence is not only important when it is part of the model, be it in a theoretical or policy framework, but it can also arise due to certain misspecifications. For instance, often the cross-sectional data used in model estimation and specification testing are imperfect, which may cause spatial dependence as a side effect. For example, census tracts are not housing markets and counties are not labor markets, but they are used as proxies to record transactions in these markets. Specifically, a mismatch between the spatial unit of observation and the spatial extent of the economic phenomena under consideration will result in spatial measurement errors and spatial autocorrelation between these errors in adjoining locations (Anselin 1988a).

In this chapter, we review the methodological issues related to the explicit treatment of spatial dependence in linear regression models. Specifically, we focus on the specification of the structure of spatial dependence (or spatial autocorrelation), on the estimation of models with spatial dependence and on specification tests to detect spatial dependence in regression models. Our review is organized accordingly into three main sections. We have limited the review to cross-sectional settings for linear regression models and do not consider dependence in space-time nor models for limited dependent variables. Whereas there is an established body of theory and methodology to deal with the standard regression case, this is not (yet) the case for techniques to analyze the other types of models. Both areas are currently the subject of active ongoing research (see, e.g., some of the papers in Anselin and Florax 1995a). Also, we have chosen to focus on a classical framework and do not consider Bayesian approaches to spatial econometrics (e.g., Hepple 1995a, 1995b, LeSage 1997).

In our review, we attempt to outline the extent to which general econometric principles can be applied to deal with spatial dependence. Spatial econometrics is often erroneously considered to consist of a straightforward extension of techniques to handle dependence in the time domain to two dimensions. In this chapter, we emphasize the limitations of such a perspective and stress the need to explicitly tackle the spatial aspects of model specification, estimation, and diagnostic testing.

II. THE PROBLEM OF SPATIAL AUTOCORRELATION

We begin this review with a closer look at the concept of spatial dependence, or its weaker expression, spatial autocorrelation, and how it differs from the more familiar serial correlation in the time domain. While, in a strict sense, spatial autocorrelation and spatial dependence clearly are not synonymous, we will use the terms interchangeably. In most applications, the weaker term autocorrelation (as a moment of the joint distribution) is used and only seldom has the focus been on the joint density as such (a recent exception is the semiparametric framework suggested in Brett and Pinkse 1997).

In econometrics, an attention to serial correlation has been the domain of time-series analysis and the typical focus of interest in the specification and estimation of models for cross-sectional data is heteroskedasticity. Until recently, spatial autocorrelation was largely ignored in this context, or treated in the form of groupwise equicorrelation, e.g., as the result of certain survey designs (King and Evans 1986). In other disciplines, primarily in physical sciences, such as geology (Isaaks and Srivastava 1989, Cressie 1991) and ecology (Legendre 1993), but also in geography (Griffith 1987, Haining 1990) and in social network analysis in sociology and psychology (Dow et al. 1982, Doreian et al. 1984, Leenders 1995), the dependence across "space" (in its most general sense) has been much more central. For example, Tobler's (1979) "first law of geography" states that "everything is related to everything else, but closer things more so," suggesting spatial dependence to be the rule rather than exception. A large body of spatial statistical techniques has been developed to deal with such dependencies (for a recent comprehensive review, see Cressie 1993; other classic references are Cliff and Ord 1973, 1981, Ripley 1981, 1988, Upton and Fingleton 1985, 1989). Useful in this respect is Cressie's (1993) taxonomy of spatial data strucures differentiating between point patterns, geostatistical data, and lattice data. In the physical sciences, the dominant underlying assumption tends to be that of a continuous spatial surface, necessitating the so-called geostatistical perspective rather than discrete observation points (or regions) in space, for which the so-called lattice perspective is relevant. The latter is more appropriate for economic data, since it is to some extent an extension of the ordering of observations on a one-dimensional time axis to an ordering in a two-dimensional space. It will be the almost exclusive focus of our review.

The traditional emphasis in econometrics on heterogeneity in cross-sectional data is not necessarily misplaced, since the distinction between spatial heterogeneity and spatial autocorrelation is not always obvious. More specifically, in a single cross section the two may be observationally equivalent. For example, when a spatial cluster of exceptionally large residuals is observed for a regression model, it cannot be ascertained without further structure whether this is an instance of heteroskedasticity (i.e., clustering of outliers) or spatial autocorrelation (a spatial stochastic process yielding clustered outliers). This problem is known in the literature as "true" contagion versus "apparent" contagion and is a major methodological issue in fields such as epidemiology (see, e.g., Johnson and Kotz 1969, Chapter 9, for a formal distinction between different forms of contagious distributions). The approach taken in spatial econometrics is to impose structure on the problem through the specification of a model, coupled with extensive specification testing for potential departures from the null model. This emphasis on the "model" distinguishes (albeit rather subtly) spatial econometrics from the broader field of spatial statistics (see also Anselin 1988a, p. 10, for further discussion of the distinction between the two). In our review, we deal almost exclusively with spatial autocorrelation. Once this aspect of the model is specified, the heterogeneity may be added in a standard manner (see Anselin 1988a, Chap. 9, and Anselin 1990a).

In this section, we first focus on a formal definition of spatial autocorrelation. This is followed by a consideration of how it may be operationalized in tests and econometric specifications by means of spatial weights and spatial lag operators. We close with a review of different ways in which spatial autocorrelation may be incorporated in the specification of econometric models in the form of spatial lag dependence, spatial error dependence, or higher-order spatial processes.

A. Defining Spatial Autocorrelation

Spatial autocorrelation can be loosely defined as the coincidence of value similarity with locational similarity. In other words, high or low values for a random variable tend to cluster in space (positive spatial autocorrelation), or locations tend to be surrounded by neighbors with very dissimilar values (negative spatial autocorrelation). Of the two types of spatial autocorrelation, positive autocorrelation is by far the more intuitive. Negative spatial autocorrelation implies a checkerboard pattern of values and does not always have a meaningful substantive interpretation (for a formal discussion, see Whittle 1954). The existence of positive spatial autocorrelation implies that a sample contains less information than an uncorrelated counterpart. In order to properly carry out statistical inference, this loss of information must be explicitly acknowledged in estimation and diagnostics tests. This is the essence of the problem of spatial autocorrelation in applied econometrics.

A crucial issue in the definition of spatial autocorrelation is the notion of "locational similarity," or the determination of those locations for which the values of the random variable are correlated. Such locations are referred to as "neighbors," though strictly speaking this does not necessarily mean that they need to be collocated (for a more formal definition of neighbors in terms of the conditional density function, see Anselin 1988a, pp. 16–17; Cressie 1993, p. 414).

More formally, the existence of spatial autocorrelation may be expressed by the following moment condition:

$$\mathrm{Cov}(y_i, y_j) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0 \qquad \text{for } i \neq j \tag{1}$$

where y_i and y_j are observations on a random variable at locations i and j in space, and i, j can be points (e.g., locations of stores, metropolitan areas, measured as latitude and longitude) or areal units (e.g., states, counties or census tracts). Of course, there is nothing *spatial* per se to the nonzero covariance in (1). It only becomes spatial when the pairs of i, j locations for which the correlation is nonzero have a meaningful interpretation in terms of spatial structure, spatial interaction or spatial arrangement of observations.

For a set of N observations on cross-sectional data, it is impossible to estimate the potentially N by N covariance terms or correlations directly from the data. This is a fundamental problem in dealing with spatial autocorrelation and necessitates the imposition of structure. More specifically, in order for the problem to become tractable, it is necessary to impose sufficient constraints on the N by N spatial interaction (covariance) matrix such that a finite number of parameters characterizing the correlation can be efficiently estimated. Note how this contrasts with the situation where repeated observations are available, e.g., in panel data sets. In such instances, under the proper conditions, the elements of the covariance matrix may be estimated explicitly, in a vector autoregressive approach (for a review, see Lütkepohl 1991) or by means of so-called generalized estimating equations (Liang and Zeger 1986, Zeger and Liang 1986, Zeger et al. 1988, Albert and McShane 1995).

In contrast, when the N observations are considered as fixed effects in space, there is insufficient information in the data to estimate the N by N interactions. Increasing the sample size does not help, since the number of interactions increases with N^2, or, in other words, there is an incidental parameter problem. Alternatively, when the locations are conceptualized in a random-effects framework, sufficient constraints must be imposed to preclude that the range of interaction implied by a particular spatial stochastic process increases faster than the sample size as asymptotics are invoked to obtain the properties of estimators and test statistics.

Two main approaches exist in the literature to impose constraints on the interaction. In geostatistics, all *pairs* of locations are sorted according to the distance that separates them, and the strength of covariance (correlation) between them is expressed as a continuous function of this distance, in a so-called variogram or semi-variogram (Cressie 1993, Chap. 2). As pointed out, the geostatistical perspective is seldom taken in empirical economics, since it necessitates an underlying process that is continuous over space. In such an approach, observations (points) are considered to form a sample from an underlying continuous spatial process, which is hard to maintain when the data consist of counties or census tracts. A possible exception may be the study of real estate data, where the locations of transactions may be conceptualized as points and analyzed using a geostatistical framework, as in Dubin (1988, 1992). Such an approach is termed "direct representation" in the literature, since the elements of the covariance (or correlation) matrix are modeled *directly* as functions of distances.

Our main focus in this review will be on the second approach, the so-called lattice perspective. For each data point, a relevant "neighborhood set" must be defined, consisting of those other locations that (potentially) interact with it. For each observation i, this yields a spatial ordering of locations $j \in S_i$ (where S_i is the neighborhood set), which can then be exploited to specify a spatial stochastic process. The covariance structure between observations is thus not modeled directly, but follows from the particular form of the stochastic process. We return to this issue below. First, we review the operational specification of the neighborhood set for each observation by means of a so-called spatial weights matrix.

B. Spatial Weights

A spatial weights matrix is a N by N positive and symmetric matrix W which expresses for each observation (row) those locations (columns) that belong to its neighborhood set as nonzero elements. More formally, $w_{ij} = 1$ when i and j are neighbors, and $w_{ij} = 0$ otherwise. By convention, the diagonal elements of the weights matrix are set to zero. For ease of interpretation, the weights matrix is often standardized such that the elements of a row sum to one. The elements of a row-standardized weights matrix thus equal $w^s_{ij} = w_{ij}/\sum_j w_{ij}$. This ensures that all weights are between 0 and 1 and facilitates the interpretation of operations with the weights matrix as an averaging of neighboring values (see Section II.C). It also ensures that the spatial parameters in many spatial stochastic processes are comparable between models. This is not intuitively obvious, but relates to constraints imposed in a maximum likelihood estimation framework. For the latter to be valid, spatial autoregressive parameters must be constrained to lie in the interval $1/\omega_{\min}$ to $1/\omega_{\max}$, where $\omega_{\min}$ and $\omega_{\max}$ are respectively the smallest (on the real line) and largest eigenvalues of the matrix W (Anselin 1982). For a row-standardized weights matrix, the largest eigenvalue is always $+1$ (Ord 1975), which facilitates the interpretation of the autoregressive coefficient as a "correlation" (for an alternative view, see Kelejian and Robinson 1995). A side effect of row-standardization is that the resulting matrix is likely to become asymmetric (since $\sum_j w_{ij} \neq \sum_i w_{ji}$), even though the original matrix may have been symmetric. In the calculation of several estimators and test statistics, this complicates computational matters considerably.

The specification of which elements are nonzero in the spatial weights matrix is a matter of considerable arbitrariness and a wide range of suggestions have been offered in the literature. The "traditional" approach relies on the geography or spatial arrangement of the observations, designating areal units as "neighbors" when they have a border in common (first-order contiguity) or are within a given distance of each other; i.e., $w_{ij} = 1$ for $d_{ij} \leq \delta$, where d_{ij} is the distance between units i and j, and δ is a distance cutoff value (distance-based contiguity). This geographic approach has been generalized to so-called Cliff-Ord weights that consist of a function of the relative length of the common border, adjusted by the inverse distance between two observations (Cliff and Ord 1973, 1981). Formally, Cliff-Ord weights may be expressed as:

$$w_{ij} = \frac{b_{ij}^{\beta}}{d_{ij}^{\alpha}} \tag{2}$$

where b_{ij} is the share of the common border between units i and j in the perimeter of i (and hence b_{ij} does not necessarily equal b_{ji}), and α and β are parameters. More generally, the weights may be specified to express any measure of "potential interaction" between units i and j (Anselin 1988a, Chap. 3). For example, this may be related directly to spatial interaction theory and the notion of potential, with $w_{ij} = 1/d_{ij}^{\alpha}$ or $w_{ij} = e^{-\beta d_{ij}}$, or more complex distance metrics may be implemented (Anselin 1980, Murdoch et al. 1993). Typically, the parameters of the distance function are set a priori (e.g., $\alpha = 2$, to reflect a gravity function) and not estimated jointly with the other coefficients in the model. Clearly, when they are estimated jointly, the resulting specification will be highly nonlinear (Anselin 1980, Chap. 8, Ancot et al. 1986, Bolduc et al. 1989, 1992, 1995).

Other specifications of spatial weights are possible as well. In sociometrics, the weights reflect whether or not two individuals belong to the same social network (Doreian 1980). In economic applications, the use of weights based on "economic" distance has been suggested, among others, in Case et al. (1993). Specifically, they suggest to use weights (before row standardization) of the form $w_{ij} = 1/|x_i - x_j|$, where x_i and x_j are observations on "meaningful" socioeconomic characteristics, such as per capita income or percentage of the population in a given racial or ethnic group.

It is important to keep in mind that, irrespective of how the spatial weights are specified, the resulting spatial process must satisfy the necessary regularity conditions such that asymptotics may be invoked to obtain the properties of estimators and test statistics. For example, this requires constraints on the extent of the range of interaction and/or the degree of heterogeneity implied by the weights matrices (the so-called mixing conditions; Anselin 1988a, Chap. 5). Specifically, this means that weights must be nonnegative and remain finite, and that they correspond to a proper metric (Anselin 1980). Clearly, this may pose a problem with socioeconomic weights when $x_i = x_j$ for some observation pairs, which may be the case for poorly chosen economic determinants (e.g., when two states have the same percentage in a given racial group). Similarly, when multiple observations belong to the same areal unit (e.g., different banks located in the same county) the distance between them must be set to something other than zero (or $1/d_{ij} \rightarrow \infty$). Finally, in the standard estimation and testing approaches, the weights matrix is taken to be *exogenous*. Therefore, indicators for the socioeconomic weights should be chosen with great care to ensure their exogeneity, unless their endogeneity is considered explicitly in the model specification.

Operationally, the derivation of spatial weights from the location and spatial arrangement of observations must be carried out by means of a geographic information system, since for all but the smallest data sets a visual inspection of a map is impractical (for implementation details, see Anselin et al. 1993a, 1993b, Anselin 1995, Can 1996). A mechanical construction of spatial weights, particularly when based on a distance criterion, may easily result in observations to become "unconnected" or isolated islands. Consequently, the row in the weights matrix that corresponds to these observations will consist of zero values. While not inherently invalidating estimation or testing procedures, the unconnected observations imply a loss of degrees of freedom, since, for all practical purposes, they are eliminated from consideration in any "spatial" model. This must be explicitly accounted for.

C. Spatial Lag Operator

In time-series analysis, values for "neighboring" observations can be easily expressed by means of a backward- or forward-shift operator on the one-dimensional time axis, yielding lagged variables y_{t-k} or y_{t+k}, where k is the desired shift (or lag). By contrast, there is no equivalent and unambiguous *spatial* shift operator. Only on a regular grid structure is there a potential solution, although not as straightforward as in the time domain. Following the so-called rook criterion for contiguity, each grid cell or vertex on a regular lattice, (i, j), has four neighbors: $(i+1, j)$ (east), $(i-1, j)$ (west), $(i, j+1)$ (north), and $(i, j-1)$ (south). Corresponding to this are four spatially shifted variables: $y_{i+1,j}$, $y_{i-1,j}$, $y_{i,j+1}$, and $y_{i,j-1}$, each of which may require its own parameter in a spatial process model. However, the rook criterion is not the only way spatial neighbors may be defined on a regular lattice, nor does the number of neighbors necessarily equal 4. For example, following the queen criterion, each observation has eight neighbors, yielding eight spatially shifted variables; the four for the rook criterion, as well as $y_{i-1,j+1}$, $y_{i-1,j-1}$, $y_{i+1,j+1}$ and $y_{i+1,j-1}$, again each possibly with its own parameter. This notion of a spatial shift operator on a regular lattice has received only limited attention in the literature, mostly with a theoretical focus and primarily in statistical mechanics, in so-called Ising models (for details, see Cressie 1993, pp. 425–426).

On an irregular spatial structure, which characterizes most economic applications, this formal notion of spatial shift is impractical, since the number of shifts would differ by observation, thereby making any statistical analysis extremely unwieldy. Instead, the concept of a *spatial lag operator* is used, which consists of a weighted average of the values at neighboring locations. The weights are fixed and exogenous, similar to a distributed lag in time series. Formally, a spatial lag operator is obtained as the product of a spatial weights matrix W with the vector of observations on a random variable y, or Wy. Each element of the resulting *spatially lagged variable* equals $\sum_j w_{ij} y_j$, i.e., a weighted average of the y values in the neighbor set S_i, since $w_{ij} = 0$ for $j \notin S_i$. Row standardization of the spatial weights matrix ensures that a spatial lag operation yields a smoothing of the neighboring values, since the positive weights sum to one.

Higher-order spatial lag operators are defined in a recursive manner, by applying the spatial weights matrix to a lower-order lagged variable. For example, a second-order spatial lag is obtained as $W(Wy)$, or W^2y. However, in contrast to time series, where such an operation is unambiguous, higher-order spatial operators yield redundant and circular neighbor relations, which must be eliminated to ensure proper estimation and inference (Blommestein 1985, Blommestein and Koper 1992, Anselin and Smirnov 1996).

In spatial econometrics, spatial autocorrelation is modeled by means of a functional relationship between a variable. y, or error term, ε, and its associated spatial lag, respectively Wy for a spatially lagged dependent variable and $W\varepsilon$ for a spatially lagged error term. The resulting specifications are referred to as *spatial lag* and *spatial error* models, the general properties of which we consider next.

D. Spatial Lag Dependence

Spatial lag dependence in a regression model is similar to the inclusion of a serially autoregressive term for the dependent variable (y_{t-1}) in a time-series context. In spatial econometrics, this is referred to as a mixed regressive, spatial autoregressive model (Anselin 1988a, p. 35). Formally,

$$y = \rho Wy + X\beta + \varepsilon \tag{3}$$

where y is a N by 1 vector of observations on the dependent variable, Wy is the corresponding spatially lagged dependent variable for weights matrix W, X is a N by K matrix of observations on the explanatory (exogenous) variables, ε is a N by 1 vector of error terms, ρ is the spatial autoregressive parameter, and β is a K by 1 vector of regression coefficients. The presence of the spatial lag term Wy on the right side of (3) will induce a nonzero correlation with the error term, similar to the presence of an endogenous variable, but different from a serially lagged dependent variable in the time-series case. In the latter model, y_{t-1} is uncorrelated with ε_t, in the absence of serial correlation in the errors. In contrast, $(Wy)_i$ is always correlated with ε_i, irrespective of the correlation structure of the errors. Moreover, the spatial lag for a given observation i is not only correlated with the error term at i, but also with the error terms at all other locations. Therefore, unlike what holds in the time-series case, an ordinary least-squares estimator will not be consistent for this specification (Anselin 1988a, Chap. 6). This can be seen from a slight reformulation of the model:

$$y = (\mathbf{I} - \rho W)^{-1} X\beta + (\mathbf{I} - \rho W)^{-1}\varepsilon \tag{4}$$

The matrix inverse $(\mathbf{I} - \rho W)^{-1}$ is a full matrix, and not triangular as in the time-series case (where dependence is only one-directional), yielding an infinite series

that involves error terms at all locations, $(\mathbf{I} + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots)\varepsilon$. It therefore readily follows that $(Wy)_i$ contains the element ε_i, as well as other ε_j, $j \neq i$. Thus,

$$E[(Wy)_i \varepsilon_i] = E[\{W(\mathbf{I} - \rho W)^{-1}\varepsilon\}_i \varepsilon_i] \neq 0 \tag{5}$$

The spatial dynamics embedded in the structure of the spatial process model (3) determine the form of the covariance between the observations at different locations (i.e., the spatial autocorrelation). For the mixed regressive, spatial autoregressive model this can easily be seen to equal $(\mathbf{I} - \rho W)^{-1}\Omega(\mathbf{I} - \rho W')^{-1}$, where Ω is the variance matrix for the error term ε (note that for a row-standardized spatial weights matrix, $W \neq W'$). Without loss of generality, the latter can be assumed to be diagonal and homoskedastic, or, $\Omega = \sigma^2\mathbf{I}$, and hence, $\mathrm{Var}[y] = \sigma^2(\mathbf{I} - \rho W)^{-1}(\mathbf{I} - \rho W')^{-1}$ The resulting variance matrix is full, implying that each location is correlated with every other location, but in a fashion that decays with the order of contiguity (the powers of W in the series expansion of $(\mathbf{I} - \rho W)^{-1}$).

The implication of this particular variance structure is that the simultaneity embedded in the Wy term must be explicitly accounted for, either in a maximum likelihood estimation framework, or by using a proper set of instrumental variables. We turn to this issue in Section III. When a spatially lagged dependent variable is ignored in a model specification, but present in the underlying data generating process, the resulting specification error is of the omitted variable type. This implies that OLS estimates in the nonspatial model (i.e., the "standard" approach) will be biased and inconsistent.

The interpretation of a significant spatial autoregressive coefficient ρ is not always straightforward. Two situations can be distinguished. In one, the significant spatial lag term indicates true contagion or substantive spatial dependence, i.e., it measures the extent of spatial spillovers, copy-catting or diffusion. This interpretation is valid when the actors under consideration match the spatial unit of observation and the spillover is the result of a theoretical model. For example, this holds for the models of farmers' innovation adoption in Case (1992), state expenditures and tax setting behavior in Case et al. (1993) and Besley and Case (1995), strategic interaction among California cities in the choice of growth controls in Brueckner (1996), and in the median voter model for recreation expenditures of Murdoch et al. (1993). Alternatively, the spatial lag model may be used to deal with spatial autocorrelation that results from a mismatch between the spatial scale of the phenomenon under study and the spatial scale at which it is measured. Clearly, when data are based on administratively determined units such as census tracts or blocks, there is no good reason to expect economic behavior to conform to these units. For example, this interpretation is useful for the spatial autoregressive models of urban housing and mortgage markets in Can (1992), Can and Megbolugbe (1997), and Anselin and Can (1996). Since urban housing and mortgage markets operate at a different spatial scale than census tracts, positive spatial autocorrelation may be expected and will in fact result in the sample containing less information than a truly "independent" sample of observations. The inclusion of a spatially lagged dependent variable in the model specification is a way to correct for this loss of information. In other words, it allows for the proper interpretation of the significance of the exogenous variables in the model (the X), after the spatial effects have been corrected for, or filtered out (see also Getis 1995 for a discussion of alternative approaches to spatial filtering). More formally, the spatial lag model may be reexpressed as

$$(\mathbf{I} - \rho W)y = X\beta + \varepsilon \tag{6}$$

where $(\mathbf{I} - \rho W)y$ is a spatially filtered dependent variable, i.e., with the effect of spatial autocorrelation taken out. This is roughly similar to first differencing of the dependent variable in time series, except that a value of $\rho = 1$ is not in the allowable parameter space for (3) and thus ρ must be estimated explicitly (Section III).

E. Spatial Error Dependence

A second way to incorporate spatial autocorrelation in a regression model is to specify a spatial process for the disturbance terms. The resulting error covariance will be nonspherical, and thus OLS estimates, while still unbiased, will be inefficient. More efficient estimators are obtained by taking advantage of the particular structure of the error covariance implied by the spatial process. Different spatial processes lead to different error covariances, with varying implications about the range and extent of spatial interaction in the model. The most common specification is a spatial autoregressive process in the error terms:

$$y = X\beta + \varepsilon \tag{7}$$

i.e., a linear regression with error vector ε, and

$$\varepsilon = \lambda W\varepsilon + \xi \tag{8}$$

where λ is the spatial autoregressive coefficient for the error lag $W\varepsilon$ (to distinguish the notation from the spatial autoregressive coefficient ρ in a spatial lag model), and ξ is an uncorrelated and (without loss of generality) homoskedastic error term. Alternatively, this may be expressed as

$$y = X\beta + (\mathbf{I} - \lambda W)^{-1}\xi \tag{9}$$

From this follows the error covariance as

$$E[\varepsilon\varepsilon'] = \sigma^2(\mathbf{I} - \lambda W)^{-1}(\mathbf{I} - \lambda W')^{-1} = \sigma^2[(\mathbf{I} - \lambda W)'(\mathbf{I} - \lambda W)]^{-1} \tag{10}$$

a structure identical to that for the dependent variable in the spatial lag model. Therefore, a spatial autoregressive error process leads to a nonzero error covariance between every pair of observations, but decreasing in magnitude with the order of contiguity. Moreover, the complex structure in the inverse matrices in (10)

yields nonconstant diagonal elements in the error covariance matrix, thus inducing heteroskedasticity in ε, irrespective of the heteroskedasticity of ξ (an illuminating numerical illustration of this feature is given in McMillen 1992). We have a much simpler situation for the case of autocorrelation in the time-series context where the model is written as $\varepsilon_t = \lambda\varepsilon_{t-1} + \xi_t$. Therefore, this is a special case of (8) with

$$W = W^T = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix}$$

where each observation is connected to only its immediate past value. As we know, for this case, $\mathrm{Var}(\varepsilon_t) = \sigma^2/(1-\lambda^2)$ for all t. That is, autocorrelation does not induce heteroskedasticity. In a time-series model, heteroskedasticity can come only through ξ_t given the above AR(1) model.

A second complicating factor in specification testing is the great degree of similarity between a spatial lag and a spatial error model, as suggested by the error covariance structure. In fact, after premultiplying both sides of (9) by $(\mathbf{I} - \lambda W)$ and moving the spatial lag term to the right side, a *spatial Durbin model* results (Anselin 1980):

$$y = \lambda Wy + X\beta - \lambda WX\beta + \xi \tag{11}$$

This model has a spatial lag structure (but with the spatial autoregressive parameter λ from (8)) with a well-behaved error term ξ. However, the equivalence between (7)–(8) and (11) imposes a set of nonlinear common factor constraints on the coefficients. Indeed, for (11) to be a proper spatial error model, the coefficients of the lagged exogenous variables WX must equal minus the product of the spatial autoregressive coefficient λ and the coefficients of X, for a total of K constraints (for technical details, see Anselin 1988a, pp. 226–229).

Spatial error dependence may be interpreted as a nuisance (and the parameter λ as a nuisance parameter) in the sense that it reflects spatial autocorrelation in measurement errors or in variables that are otherwise not crucial to the model (i.e., the "ignored" variables spillover across the spatial units of observation). It primarily causes a problem of inefficiency in the regression estimates, which may be remedied by increasing the sample size or by exploiting consistent estimates of the nuisance parameter. For example, this is the interpretation offered in the model of agricultural land values in Benirschka and Binkley (1994).

The spatial autoregressive error model can also be expressed in terms of spatially filtered variables, but slightly different from (6). After moving the spatial lag variable in (11) to the left hand side, the following expression results:

$$(\mathbf{I} - \lambda W)y = (\mathbf{I} - \lambda W)X\beta + \xi \tag{12}$$

This is a regression model with spatially filtered dependent and explanatory variables and with an uncorrelated error term ξ, similar to first differencing of both y and X in time-series models. As in the spatial lag model, $\lambda = 1$ is outside the parameter space and thus λ must be estimated jointly with the other coefficients of the model (see Section III).

Several alternatives to the spatial autoregressive error process (8) have been suggested in the literature, though none of them have been implemented much in practice. A spatial moving average error process is specified as (Cliff and Ord 1981, Haining 1988, 1990):

$$\varepsilon = \gamma W\xi + \xi \tag{13}$$

where γ is the spatial moving average coefficient and ξ is an uncorrelated error term. This process thus specifies the error term at each location to consist of a location-specific part, ξ_i ("innovation"), as well as a weighted average (smoothing) of the errors at neighboring locations, $W\xi$. The resulting error covariance matrix is

$$E[\varepsilon\varepsilon'] = \sigma^2(\mathbf{I} + \gamma W)(\mathbf{I} + \gamma W') = \sigma^2[\mathbf{I} + \gamma(W + W') + \gamma^2 WW'] \tag{14}$$

Note that in contrast to (10), the structure in (14) does not yield a full covariance matrix. Nonzero covariances are only found for first-order ($W + W'$) and second-order (WW') neighbors, thus implying much less overall interaction than the autoregressive process. Again, unless all observations have the same number of neighbors and identical weights, the diagonal elements of (14) will not be constant, inducing heteroskedasticity in ε, irrespective of the nature of ξ.

A very similar structure to (13) is the spatial error components model of Kelejian and Robinson (1993, 1995), in which the disturbance is a sum of two independent error terms, one associated with the "region" (a smoothing of neighboring errors) and one which is location-specific:

$$\varepsilon = W\xi + \psi \tag{15}$$

with ξ and ψ as independent error components. The resulting error covariance is

$$E[\varepsilon\varepsilon'] = \sigma_\psi^2\mathbf{I} + \sigma_\xi^2 WW' \tag{16}$$

where σ_ψ^2 and σ_ξ^2 are the variance components associated with respectively the location-specific and regional error parts. The spatial interaction implied by (16) is even more limited than for (14), pertaining only to the first- and second-order neighbors contained in the nonzero elements of WW'. Heteroskedasticity is implied unless all locations have the same number of neighbors and identical weights, a situation excluded by the assumptions needed for the proper asymptotics in the model (Kelejian and Robinson 1993, p. 301).

In sum, every type of spatially dependent error process induces heteroskedasticity as well as spatially autocorrelated errors, which will greatly complicate specification testing in practice. Note that the "direct representation" approach based on geostatistical principles does not suffer from this problem. For example, in Dubin (1988, 1992), the elements of the error covariance matrix are expressed directly as functions of the distance d_{ij} between the corresponding observations, e.g., $E[\varepsilon_i\varepsilon_j] = \gamma_1 e^{(-d_{ij}/\gamma_2)}$, with γ_1 and γ_2 as parameters. Since $e^{-d_{ii}/\gamma_2} = 1$, irrespective of the value of γ_2, the errors ε will be homoskedastic unless explicitly modeled otherwise.

F. Higher-Order Spatial Processes

Several authors have suggested processes that combine spatial lag with spatial error dependence, though such specifications have seen only limited applications. The most general form is the spatial autoregressive, moving-average (SARMA) process outlined by Huang (1984). Formally, a SARMA(p, q) process can be expressed as

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \cdots + \rho_p W_p y + \varepsilon \tag{17}$$

for the spatial autoregressive part, and

$$\varepsilon = \gamma_1 W_1\xi + \gamma_2 W_2\xi + \cdots + \gamma_q W_q\xi + \xi \tag{18}$$

for the moving-average part, in the same notation as above. For greater generality, a regressive component $X\beta$ can be added to (17) as well. The spatial autocorrelation pattern resulting from this general formulation is highly complex. Models that implement aspects of this form are the second-order SAR specification in Brandsma and Ketellapper (1979a) and higher-order SAR models in Blommestein (1983, 1985).

A slightly different specification combines a first-order spatial autoregressive lag with a first-order spatial autoregressive error (Anselin 1980, Chap. 6; Anselin 1988a, pp. 60–65). It has been applied in a number of empirical studies, most notably in the work of Case, such as the analysis of household demand (Case 1987, 1991), of innovation diffusion (Case 1992), and local public finance (Case et al. 1993, Besley and Case 1995). Formally, the model can be expressed as a combination of (3) with (8), although care must be taken to differentiate the weights matrix used in the spatial lag process from that in the spatial error process:

$$y = \rho W_1 y + X\beta + \varepsilon \tag{19}$$

$$\varepsilon = \lambda W_2\varepsilon + \xi \tag{20}$$

After some algebra, combining (20) and (19) yields the following reduced form:

$$y = \rho W_1 y + \lambda W_2 y - \lambda\rho W_2 W_1 y + X\beta - \lambda W_2 X\eta + \xi \tag{21}$$

i.e., an extended form of the spatial Durbin specification but with an additional set of nonlinear constraints on the parameters. Note that when W_1 and W_2 do not overlap, for example when they pertain to different orders of contiguity, the product $W_2 W_1 = 0$

and (21) reduces to a biparametric spatial lag formulation, albeit with additional constraints on the parameters. On the other hand, when W_1 and W_2 are the same, the parameters ρ and λ are only identified when at least one exogenous variable is included in X (in addition to the constant term) and when the nonlinear constraints are enforced (Anselin 1980, p. 176). When $W_1 = W_2 = W$, the model becomes

$$y = (\rho + \lambda)Wy - \lambda\rho W^2 y + X\beta - \lambda WX\beta + \xi \qquad (22)$$

Clearly, the coefficients of Wy and W^2y alone do not allow for a separate identification of ρ and λ. Using the nonlinear constraints between the β and $-\lambda\beta$ (the coefficients of X and WX) yields an estimate of λ, but this will only be unique when the constraints are strictly enforced. Similarly, an estimate of λ may result in two possible estimates for ρ (one using the coefficient of Wy, the other of W^2y) unless the nonlinear constraints are strictly enforced. This considerably complicates estimation strategies for this model. In contrast, a SARMA(1, 1) model does not suffer from this problem.

In empirical practice, an alternative perspective on the need for higher-order processes is to consider them to be a result of a poorly specified weights matrix rather than as a realistic data generating process. For example, if the weights matrix in a spatial lag model underbounds the true spatial interaction in the data, there will be remaining spatial error autocorrelation. This may lead one to implement a higher-order process, while for a properly specified weights matrix no such process is needed (see Florax and Rey 1995 for a discussion of the effects of misspecified weights). In practice, this will require a careful specification search for the proper form of the spatial dependence in the model, an issue to which we return in Section IV. First, we consider the estimation of regression models that incorporate spatial autocorrelation of a spatial lag or error form.

III. ESTIMATING SPATIAL PROCESS MODELS

Similar to when serial dependence is present in the time domain, classical sampling theory no longer holds for spatially autocorrelated data, and estimation and inference must rely on the asymptotic properties of stochastic processes. In essence, rather than considering N observations as independent pieces of information, they are conceptualized as a single realization of a process. In order to carry out meaningful inference on the parameters of such a process, constraints must be imposed on both heterogeneity and the range of interaction. While many properties of estimators for spatial process models may be based on the same principles as developed for dependent (and heterogeneous) processes in the time domain (e.g., the formal properties outlined in White 1984, 1994), there are some important differences as well. Before covering specific estimation procedures, we discuss these differences in some detail, focusing in particular on the notion of stationarity in space and the distinction between simultaneous and conditional spatial processes. Next, we turn to a review of maximum likelihood and instrumental variables estimators for spatial regression models. We close with a brief discussion of operational implementation and software issues.

A. Spatial Stochastic Processes

As in the time domain, in order to carry out meaningful inference for a spatial process, some degree of equilibrium must be assumed in the sense that the stochastic generating mechanism is taken to work uniformly over space. In a strict sense, a notion of "spatial stationarity" accomplishes this objective since it imposes the condition that any joint distribution of the random variable under consideration over a subset of the locations depends only on the relative position of these observations in terms of their relative orientation (angle) and distance. Even stricter is a notion of isotropy, for which only distance matters and orientation is irrelevant. For practical purposes, the notions of stationarity and isotropy are too demanding and not verifiable. Hence, weaker conditions are typically imposed in the form of stationarity of the first (mean) and second moments (variance, covariance, or spatial autocorrelation). Even weaker requirements follow from the so-called intrinsic hypothesis in geostatistics, which requires only stationarity of the variance of the increments, leading to the notion of a variogram (for technical details, see Ripley 1988, pp. 6–7; Cressie 1993, pp. 52–68).

For stationary processes in the time domain, the careful inspection of autocovariance and autocorrelation functions is a powerful aid in the identification of the model, e.g., following the familiar Box-Jenkins approach (Box et al. 1994). One could transpose this notion to spatial processes and consider spatial autocorrelation functions indexed by order of contiguity as the basis for model identification. However, as Hooper and Hewings (1981) have shown, this is only appropriate for a very restrictive class of spatial processes on regular lattice structures. For applied work in empirical economics, such restrictions are impractical and the spatial dependence in the model must be specified explicitly by means of the spatial lag and spatial error structures reviewed in the previous section. Inference may be based on the asymptotic properties (central limit theorems and laws of large numbers) of so-called dependent and heterogeneous processes, as developed in White and Domowitz (1984) and White (1984, 1994). Central to these notions is the concept of mixing sequences, allowing for a trade-off between the range of dependence and the extent of heterogeneity (see Anselin 1988a, pp. 45–46 for an intuitive extension of this to spatial econometric models). While rigorous proofs of these properties have not been derived for the explicit spatial case, the notion of a spatial weights matrix based on a proper metric is general enough to meet the criteria imposed by mixing conditions. In a spatial econometric approach then, a spatial lag model is considered to be a special case of simultaneity or endogeneity with dependence, and a spatial error model is a special case of a nonspherical error term, both of which can be tackled by means of generally established econometric theory, though not as direct extensions of the time-series analog.

The emphasis on "simultaneity" in spatial econometrics differs somewhat from the approach taken in spatial statistics, where conditional models are often considered to be more natural (Cressie 1993, p. 410). Again, the spatial case differs substantially from the time-series one since in space a conditional and simultaneous approach are no longer equivalent (Brook 1964, Besag 1974, Cressie 1993, pp. 402–410). More specifically, in the time domain a Markov chain stochastic process can be expressed in terms of the joint density (ignoring a starting point to ease notation) as

$$\text{Prob}[z] = \prod_{t=1}^{N} Q_t[z_t, z_{t-1}] \tag{23}$$

where z refers to the vector of observations for all time points, and Q_t is a function that only contains the observation at t and at $t-1$ (hence, a Markov chain). The conditional density for this process is

$$\text{Prob}[z_t|z_1, z_2, \ldots, z_{t-1}] = \text{Prob}[z_t|z_{t-1}] \tag{24}$$

illustrating the lack of memory of the process (i.e., the conditional density depends only on the first-order lag). Due to the one-directional nature of dependence in time, (23) and (24) are equivalent (Cressie 1993, p. 403). An extension of (23) to the spatial domain may be formulated as

$$\text{Prob}[z] = \prod_{i=1}^{N} Q_i[z_j, z_j; j \in S_i] \tag{25}$$

where the z_j only refer to those locations that are part of the neighborhood set S_i of i. A conditional specification would be

$$\text{Prob}[z_i|z_j, j \neq i] = \text{Prob}[z_i|z_j; j \in S_i] \tag{26}$$

i.e., the conditional density of z_i, given observations at all other locations only depends on those locations in the neighborhood set of i. The fundamental result in this respect goes back to Besag (1974), who showed that the conditional specification only yields a proper joint distribution when the so-called Hammersley-Clifford theorem is satisfied, which imposes constraints on the type and range of dependencies in (26). Also, while a joint density specification always yields a proper conditional specification, in range of spatial interaction implied is not necessarily the same. For example, Cressie (1993, p. 409) illustrates how a first-order symmetric spatial autoregressive process corresponds with a conditional specification that includes third-order neighbors (Haining 1990, pp. 89–90). Consequently, it does make a difference whether one approaches a spatially autocorrelated phenomenon by means of (26) versus (25). This also has implications for the substantive interpretation of the model results, as illustrated for an analysis of retail pricing of gasoline in Haining (1984).

In practice, it is often easier to estimate a conditional model, especially for nonnormal distributions (e.g., auto-Poisson, autologistic). Also, a conditional specification is more appropriate when the focus is on spatial prediction or interpolation. For general estimation and inference, however, the constraints imposed on the type and range of spatial interaction in order for the conditional density to be proper are often highly impractical in empirical work. For example, an auto-Poisson model (conditional model for spatially autocorrelated counts) only allows negative autocorrelation and hence is inappropriate for any analysis of clustering in space.

In the remainder, our focus will be exclusively on simultaneously specified models, which is a more natural approach from a spatial econometric perspective (Anselin 1988a, Cressie 1993, p. 410).

B. Maximum Likelihood Estimation

The first comprehensive treatment of maximum likelihood estimation of regression models that incorporate spatial autocorrelation in the form of a spatial lag or a spatial error term was given by Ord (1975). The point of departure is a joint normal density for the errors in the model, from which the likelihood function is derived. An important aspect of this likelihood function is the Jacobian of the transformation, which takes the form $|\mathbf{I} - \rho W|$ and $|\mathbf{I} - \lambda W|$ in respectively the spatial lag and spatial autoregressive error models, with ρ and λ as the autoregressive coefficient and W as the spatial weights matrix. The need for this Jacobian can be seen from expression (4) for the spatial lag model and (12) for the spatial autoregressive error model (for a more extensive treatment, see Anselin 1988a, Chap. 6). In contrast to the time-series case, the spatial Jacobian is not the determinant of a triangular matrix, but of a full matrix. This would complicate computational matters considerably, were it not that Ord (1975) showed how it can be expressed in function of the eigenvalues ω_i of the spatial weights matrix as

$$|\mathbf{I} - \rho W| = \prod_{i=1}^{N} (1 - \rho\omega_i) \tag{27}$$

Using this simplification, under the normality assumption, the log-likelihood function for the spatial lag model (3) follows in a straightforward manner as

$$L = \sum_i \ln(1 - \rho\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - \rho W y - X\beta)'(y - \rho W y - X\beta)}{2\sigma^2} \tag{28}$$

in the same notation as used in Section II. This expression clearly illustrates why, in contrast to the time-series case, ordinary least squares (i.e., the minimization of the last term in (28)) is not maximum likelihood, since it ignores the Jacobian term. From the usual first-order conditions, the ML estimates for β and σ^2 in a spatial lag model are obtained as (for details, see Ord 1975, Anselin 1980, Chap. 4: Anselin 1988a, Chap. 6):

$$\beta_{ML} = (X'X)^{-1}X'(\mathbf{I} - \rho W)y \tag{29}$$

and

$$\sigma^2_{ML} = \frac{(y - \rho Wy - X\beta_{ML})'(y - \rho Wy - X\beta_{ML})}{N} \tag{30}$$

Conditional upon ρ, these estimates are simply OLS applied to the spatially filtered dependent variable and the explanatory variables in (6). Substitution of (29) and (30) in the log-likelihood function yields a concentrated log-likelihood as a nonlinear function of a single parameter ρ:

$$L_c = -\frac{N}{2}\ln\left[\frac{(e_0 - \rho e_L)'(e_0 - \rho e_L)}{N}\right] + \sum_i \ln(1 - \rho\omega_i) \tag{31}$$

where e_0 and e_L are residuals in a regression of y on X and Wy on X, respectively (for technical details, see Anselin 1980, Chap. 4). A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function (31). Based on the framework outlined in Heijmans and Magnus (1986a, 1986b), it can be shown that the resulting estimates have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency. The asymptotic variance matrix follows as the inverse of the information matrix

$$\mathrm{AsyVar}[\rho, \beta, \sigma^2]$$

$$= \begin{bmatrix} \mathrm{tr}[W_A]^2 + \mathrm{tr}[W_A' W_A] + \dfrac{[W_A X\beta]'[W_A X\beta]}{\sigma^2} & \dfrac{(X'W_A X\beta)'}{\sigma^2} & \dfrac{\mathrm{tr}(W_A)}{\sigma^2} \\ \dfrac{X'W_A X\beta}{\sigma^2} & \dfrac{X'X}{\sigma^2} & 0 \\ \dfrac{\mathrm{tr}(W_A)}{\sigma^2} & 0 & \dfrac{N}{2\sigma^4} \end{bmatrix}^{-1} \tag{32}$$

where $W_A = W(\mathbf{I} - \rho W)^{-1}$ to simplify notation. Note that while the covariance between β and the error variance is zero, as in the standard regression model, this is not the case for ρ and the error variance. This lack of block diagonality in the information matrix for the spatial lag model will lead to some interesting results on the structure of specification tests, to which we turn in Section IV. It is yet another distinguishing characteristic between the spatial case and its analog in time series.

Maximum likelihood estimation of the models with spatial error autocorrelation that were covered in Section II.E can be approached by considering them as special cases of general parametrized nonspherical error terms, for which $E[\varepsilon\varepsilon'] = \sigma^2\Omega(\theta)$, with θ as a vector of parameters. For example, from (32) for a spatial autoregressive error term, it follows that

$$\Omega(\lambda) = [(\mathbf{I} - \lambda W)'(\mathbf{I} - \lambda W)]^{-1} \tag{33}$$

As shown in Anselin (1980, Chap. 5), maximum likelihood estimation of such specifications can be carried out as an application of the general framework outlined in Magnus (1978). Most spatial processes satisfy the necessary regularity conditions, although this is not necessarily the case for direct representation models (Mardia and Marshall 1984, Warnes and Ripley 1987, Mardia and Watkins 1989). Under the assumption of normality, the log-likelihood function takes on the usual form:

$$L = -\frac{1}{2}\ln|\Omega(\lambda)| - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - X\beta)'\Omega(\lambda)^{-1}(y - X\beta)}{2\sigma^2} \tag{34}$$

for example, with $\Omega(\lambda)$ as in (33). First-order conditions yield the familiar generalized least-squares estimates for β, conditional upon λ:

$$\beta_{\mathrm{ML}} = [X'\Omega(\lambda)^{-1}X]^{-1}X'\Omega(\lambda)^{-1}y \tag{35}$$

For a spatial autoregressive error process, $\Omega(\lambda)^{-1} = (\mathbf{I} - \lambda W)'(\mathbf{I} - \lambda W)$, so that for known λ, the maximum likelihood estimates are equivalent to OLS applied to the spatially filtered variables in (12). Note that for other forms of error dependence, the GLS expression (35) will involve the inverse of an N by N error covariance matrix. For example, for the spatial moving average errors, as in (13), $\Omega(\gamma)^{-1} = [\mathbf{I} + \gamma(W + W') + \gamma^2 WW']^{-1}$, which does not yield a direct expression in terms of spatially transformed y and X.

Obtaining a consistent estimate for λ is not as straightforward as in the time-series case. As pointed out, OLS does not yield a consistent estimate in a spatial lag model. It therefore cannot be used to obtain an estimate for λ from a regression of residuals e on We, as in the familiar Cochrane-Orcutt procedure for serially autoregressive errors in the time domain. Instead, an explicit optimization of the likelihood function must be carried out. One approach is to use the iterative solution of the first-order conditions in Magnus (1978, p. 283):

$$\mathrm{tr}\left[\left(\frac{\partial\Omega^{-1}}{\partial\lambda}\right)\Omega\right] = e'\left(\frac{\partial\Omega^{-1}}{\partial\lambda}\right)e \tag{36}$$

where $e = y - X\beta$ are GLS residuals. For a spatial autoregressive error process, $\partial\Omega^{-1}/\partial\lambda = -W - W' + \lambda W'W$. Solution of condition (36) can be obtained by numerical means. Alternatively, the GLS expression for β and similar solution of the first-order conditions for σ^2 can be substituted into the log-likelihood function to yield a concentrated log-likelihood as a nonlinear function of the autoregressive parameter λ (for technical details, see Anselin 1980, Chap. 5):

$$L_C = -\frac{N}{2}\ln\left(\frac{u'u}{N}\right) + \sum_i \ln(1 - \lambda\omega_i) \tag{37}$$

with $u'u = y_L'y_L - y_L'X_L[X_L'X_L]^{-1}X_L'y_L$, and y_L and X_L as spatially filtered variables, respectively $y - \lambda Wy$ and $X - \lambda WX$. The Jacobian term follows from $\ln|\Omega(\lambda)| = 2\ln|\mathbf{I} - \lambda W|$ and the Ord simplification in terms of eigenvalues of W.

The asymptotic variance for the ML estimates conforms to the Magnus (1978) and Breusch (1980) general form and is block diagonal between the regression (β) and error variance parameters σ^2 and θ. For example, for a spatial autoregressive error, the asymptotic variance for the regression coefficients is $\text{AsyVar}[\beta] = \sigma^2[X_L'X_L]^{-1}$. The variance block for the error parameters is

$$\text{AsyVar}[\sigma^2, \lambda] = \begin{bmatrix} N/2\sigma^4 & \dfrac{\text{tr}(W_B)}{\sigma^2} \\ \dfrac{\text{tr}(W_B)}{\sigma^2} & \text{tr}(W_B)^2 + \text{tr}(W_B'W_B) \end{bmatrix}^{-1} \tag{38}$$

where, for ease of notation, $W_B = W(\mathbf{I} - \lambda W)^{-1}$. Due to the block-diagonal form of the asymptotic variance matrix, knowledge of the precision of λ does not affect the precision of the β estimates. Consequently, if the latter is the primary interest, the complex inverse and trace expressions in (38) need not be computed, as in Benirschka and Binkley (1994). A significance test for the spatial error parameter can be based on a likelihood ratio test, in a straightforward way (Anselin 1988a, Chap. 8).

Higher-order spatial processes can be estimated using the same general principles, although the resulting log-likelihood function will be highly nonlinear and the use of a concentrated log-likelihood becomes less useful (Anselin 1980, Chap. 6).

The fit of spatial process models estimated by means of maximum likelihood procedures should not be based on the traditional R^2, which will be misleading in the presence of spatial autocorrelation. Instead, the fit of the model may be assessed by comparing the maximized log-likelihood or an adjusted form to take into account the number of parameters in the models, such as the familiar AIC (Anselin 1988b).

C. GMM/IV Estimation

The view of a spatially lagged dependent variable Wy in the spatial lag model as a form of endogeneity or simultaneity suggests an instrumental variable (IV) approach

to estimation (Anselin 1980, 1988a, Chap. 7; 1990b). Since the main problem is the correlation between Wy and the error term in (3), the choice of proper instruments for Wy will yield consistent estimates. However, as usual, the efficiency of these estimates depends crucially on the choice of the instruments and may be poor in small samples. On the other hand, in contrast to the maximum likelihood approach just outlined, IV estimation does not require an assumption of normality.

Using the standard econometric results (for a review, see Bowden and Turkington 1984), and with Q as a P by N matrix ($P \geq K+1$) of instruments (including K "exogenous" variables from X), the IV or 2SLS estimate follows as

$$\beta_{\text{IV}} = [Z'Q(Q'Q)^{-1}Q'Z]^{-1}Z'Q(Q'Q)^{-1}Q'y \tag{39}$$

with $Z = [Wy\ X]$, $\text{AsyVar}(\beta_{\text{IV}}) = \sigma^2[Z'Q(Q'Q)^{-1}Q'Z]^{-1}$, and $\sigma^2 = (y - Z\beta_{\text{IV}})'(y - Z\beta_{\text{IV}})/N$.

Clearly, this approach can also be applied to models where other endogenous variables appear in addition to the spatially lagged dependent variable, as in a simultaneous equation context, provided that the instrument set is augmented to deal with this additional endogeneity. It also forms the basis for a bootstrap approach to the estimation of spatial lag models (Anselin 1990b). Moreover, it is easily extended to deal with more complex error structures, e.g., reflecting forms of heteroskedasticity or spatial error dependence (Anselin 1988a, pp. 86–88). The formal properties of such an approach are derived in Kelejian and Robinson (1993) for a general methods of moments estimator (GMM) in the model $y = \rho Wy + X\beta + \varepsilon$ with spatial error components, $\varepsilon = W\xi + \psi$. The GMM estimator takes the form

$$\beta_{\text{GMM}} = [Z'Q(Q'\hat{\Omega}Q)^{-1}Q'Z]^{-1}Z'Q(Q'\hat{\Omega}Q)^{-1}Q'y \tag{40}$$

where $\hat{\Omega}$ is a consistent estimate for the error covariance matrix. The asymptotic variance for β_{GMM} is $[Z'Q(Q'\hat{\Omega}Q)^{-1}Q'Z]^{-1}$. For the spatial error components model, Kelejian and Robinson (1993, pp. 302–304) suggest an estimate for $\hat{\Omega} = \hat{\varphi}_1\mathbf{I} + \hat{\varphi}_2WW'$, with $\hat{\varphi}_1$ and $\hat{\varphi}_2$ as the least-squares estimates in an auxilliary regression of the squared IV residuals $(y - Z\beta_{\text{IV}})$ on a constant and the diagonal elements of WW'.

A particularly attractive application of GLS-IV estimation in spatial lag models is a special case of the familiar White heteroskedasticity-consistent covariance estimator (White 1984, Bowden and Turkington 1984, p. 91). The estimator is as in (40), but $Q'\hat{\Omega}Q$ is estimated by $Q'\tilde{\Omega}Q$, where $\tilde{\Omega}$ is a diagonal matrix of squared IV residuals, in the usual fashion. This provides a way to obtain consistent estimates for the spatial autoregressive parameter ρ in the presence of heteroskedasticity of unknown form, often a needed feature in applied empirical work.

A crucial issue in instrumental variables estimation is the choice of the instruments. In spatial econometrics, several suggestions have been made to guide the selection of instruments for Wy (for a review, see Anselin 1988a, pp. 84–86; Land and Deane 1992). Recently, Kelejian and Robinson (1993 p. 302) formally demonstrated the consistency of β_{GMM} in the spatial lag model with instruments consisting of first-order and higher-order spatially lagged explanatory variables (WX, W^2X, etc.).

An important feature of the instrumental variables approach is that estimation can easily be carried out by means of standard econometric software, provided that the spatial lags can be computed as the result of common matrix manipulations (Anselin and Hudak 1992). In contrast, the maximum likelihood approach requires specialized routines to implement the nonlinear optimization of the log-likelihood (or concentrated log-likelihood). We next turn to some operational issues related to this.

D. Operational Implementation and Illustration

To date, none of the widely available econometric software packages contain specific routines to implement maximum likelihood estimation of spatial process models or to carry out specification tests for spatial autocorrelation in regression models. This lack of attention to the analysis of the lattice data structures that are most relevant in empirical economics contrasts with a relatively large range of software for spatial data analysis in the physical sciences, geared to point patterns and geostatistical data. Examples of these are the GSLIB library (Deutsch and Journel 1992) and the recent S+Spatialstats add-on to the S-PLUS statistical software (MathSoft 1996). While the latter does include some analyses for lattice data, estimation is limited to maximum likelihood of spatial error models with autoregressive or moving-average structures. However, the spatial lag model is not covered and specification diagnostics are totally absent.

The only self-contained software package specifically geared to spatial econometric analysis in SpaceStat (Anselin 1992b, 1995). It contains both maximum likelihood and instrumental variables estimators for spatial lag and error models, as well as ways to estimate heteroskedastic specifications and a wide range of diagnostics for spatial effects. In addition, SpaceStat also includes extensive features to carry out exploratory spatial data analysis as well as utilities to create and manipulate spatial weights matrices and interface with geographic information systems.

There are two major practical issues that must be resolved to implement the estimation of spatial lag and spatial error models. The first is the need to construct spatially lagged variables from observations on the dependent variable or residual term. This is relevant for both instrumental variables (IV, 2SLS, GMM) as well as maximum likelihood estimation. In principle, the lag can be computed as a simple matrix multiplication of the spatial weights matrix W with the vector of observations, say Wy. This is straightforward to implement in most econometric software packages that contain matrix algebra routines (specific examples for Gauss, Splus, Limdep, Rats and Shazam are given in Anselin and Hudak 1992, Table 2, p. 514). In practice, however, the size of the matrix that can be manipulated by econometric software is severely limited and insufficient for most empirical applications, unless sparse matrix routines can be exploited (avoiding the need to store a full N by N matrix). This is increasingly the case in state-of-the-art matrix algebra packages (e.g., Matlab, Gauss), but still fairly uncommon in application-oriented econometric software; hence, the computation of spatial lags will typically necessitate some programming effort on the part of the user (the construction of spatial lags based on sparse spatial weights formats in SpaceStat is discussed in Anselin 1995). Once the spatial lagged dependent variables are computed, IV estimation of the spatial lag model can be carried out with any standard econometric package.

The other major operational issue pertains only to maximum likelihood estimation. It is the need to manipulate large matrices of dimension equal to the number of observations in the asymptotic variance matrices (32) and (38) and in the Jacobian term (27) of the log-likelihoods (31) and (37). In contrast to the time-series case, the matrix W is not triangular and hence a host of computational simplifications are not applicable. The problem is most serious in the computation of the asymptotic variance matrix of the maximum likelihood estimates. The inverse matrices in both $W_A = W(\mathbf{I} - \rho W)^{-1}$ of (32) and $W_B = W(\mathbf{I} - \lambda W)^{-1}$ of (38) are full matrices which do not lend themselves to the application of sparse matrix algorithms. For low values of the autoregressive parameters, a power expansion of $(\mathbf{I} - \rho W)^{-1}$ or $(\mathbf{I} - \lambda W)^{-1}$ may be a reasonable approximation to the inverse, e.g., $(\mathbf{I} - \rho W)^{-1} = \sum_k \rho^k W^k +$ error, with $k = 0, 1, \ldots, K$, such that $\rho^K < \delta$, where δ is a sufficiently small value. However, this will involve some computing effort in the construction of the powers of the weights matrices and is increasingly burdensome for higher values of the autoregressive parameter. In general, for all practical purposes, the size of the problem for which an asymptotic variance matrix can be computed is constrained by the largest matrix inverse that can be carried out with acceptable numerical precision in a given software/hardware environment. In current desktop settings, this typically ranges from a few hundred to a few thousand observations. While this makes it impossible to compute asymptotic t-tests for all the parameters in spatial models with very large numbers of observations, it does not preclude asymptotic inference. In fact, as we argued in Section III.B, due to the block diagonality of the asymptotic variance matrix in the spatial error case, asymptotic t-statistics can be constructed for the estimated β coefficients without knowledge of the precision of the autoregressive parameter λ (see also Benirschka and Binkley 1994, Pace and Barry 1996). Inference on the autoregressive parameter can be based on a likelihood ratio test (Anselin 1988a, Chap. 6). A similar approach can be taken in the spatial lag model. However, in contrast to the error case, asymptotic t-tests can no longer be constructed for the estimated β coefficients, since the asymptotic variance matrix (32) is not block diagonal. Instead, likelihood ratio tests must be considered explicitly for any subset of coefficients of interest (requiring a separate optimization for each specification; see Pace and Barry 1997).

With the primary objective of obtaining consistent estimates for the parameters in spatial regression models, a number of authors have suggested ways to manipulate popular statistical and econometric software packages in order to maximize the log-likelihoods (28) and (37). Examples of such efforts are routines for ML estimation of the spatial lag and spatial autoregressive error model in Systat, SAS, Gauss, Limdep, Shazam, Rats and S-PLUS (Bivand 1992, Griffith 1993, Anselin and Hudak 1992, Anselin et al. 1993b). The common theme among these approaches is to find a way to convert the log-likelihoods for the spatial models to a form amenable for use with standard nonlinear optimization routines. Such routines proceed incrementally, in the sense that the likelihood is built up from a sum of elements that correspond to individual observations. At first sight, the Jacobian term in the spatial models would preclude this. However, taking advantage of the Ord decomposition in terms of eigenvalues, pseudo-observations can be constructed for the elements of the Jacobian. Specifically, each term $1 - \rho\omega_i$ is considered to correspond to a pseudo-variable ω_i, and is summed over all "observations." For example, for the spatial lag model, the log-likelihood (ignoring constant terms) can be expressed as

$$L = \sum_i \left[\ln(1 - \rho\omega_i) - \frac{\ln(\sigma^2)}{2} - \frac{(y_i - \rho\{Wy\}_i - x_i\beta)^2}{2\sigma^2} \right] \tag{41}$$

which fits the format expected by most nonlinear optimization routines. Examples of practical implementations are listed in Anselin and Hudak (1992, Table 10, p. 533) and extensive source code for various econometric software packages is given in Anselin et al. (1993b).

One problem with this approach is that the asymptotic variance matrices computed by the routines tend to be based on a numerical approximation and do not necessarily correspond to the analytical expressions in (32) and (38). This may lead to slight differences in inference depending on the software package that is used (Anselin and Hudak 1992, Table 10, p. 533). An alternative approach that does not require the computation of eigenvalues is based on sparse matrix algorithms to efficiently compute the determinant of the Jacobian at each iteration of the optimization routine. While this allows the estimation of models for very large data sets (tens of thousands of observations), for example, by using the specialized routines in the Matlab software, this does not solve the asymptotic variance matrix problem. Inference therefore must be based on likelihood ratio statistics (for details and implementation, see Pace and Barry 1996, 1997).

To illustrate the various spatial models and their estimation, the results for the parameters in a simple spatial model of crime estimated for 49 neighborhoods in Columbus, Ohio, are presented in Table 1. The model and results are based on Anselin (1988a, pp. 187–196) and have been used in a number of papers to benchmark different estimators and specification tests (e.g., McMillen 1992, Getis 1995, Anselin et al. 1996, LeSage 1997). The data are also available for downloading via the internet from http://www.rri.wvu.edu/spacestat.htm. The estimates reported in Table 1 include OLS in the standard regression model, OLS (inconsistent), ML, IV, and heteroskedastic-robust IV for the spatial lag model, and ML for the spatial error

Table 1 Estimates in a Spatial Model of Crime[a]

	OLS	Lag-OLS	Lag-ML	Lag-IV	Lag-GIVE	Err-ML
Constant	68.629	38.783	45.079	43.963	46.667	59.893
	(4.73)	(9.32)	(7.18)	(11.23)	(7.61)	(5.37)
ρ		0.549	0.431	0.453	0.419	
		(0.153)	(0.118)	(0.191)	0.139)	
Income	−1.597	−0.886	−1.032	−1.010	−1.185	−0.941
	(0.334)	(0.358)	(0.305)	(0.389)	(0.434)	(0.331)
Housing	−0.274	−0.264	−0.266	−0.266	−0.234	−0.302
value	(0.103)	(0.092)	(0.088)	(0.092)	(0.173)	(0.090)
λ						0.562
						(0.134)
R^2	0.552	0.652		0.620	0.633	
Log-lik	−187.38		−182.39			−183.38

[a]Data are for 49 neighborhoods in Columbus, Ohio, 1980. Dependent variable is per capita residential burglaries and vehicle thefts. Income and housing values are in thousand dollars. A first-order contiguity spatial weights matrix was used to construct the spatial lags.

model. The spatial lags for the exogenous variables (WX) were used as instruments in the IV estimation. In addition to the estimates and their standard errors, the fit of the different specifications estimated by ML is compared by means of the maximized log-likelihood. For OLS and the IV estimates, the R^2 is listed. However, this should be interpreted with caution, since R^2 is inappropriate as a measure of fit when spatial dependence is present. All estimates were obtained by means of the SpaceStat software.

A detailed interpretation of the results in Table 1 is beyond the scope of this chapter, but a few noteworthy features may be pointed out. The two spatial models provide a superior fit relative to OLS, strongly suggesting the presence of spatial dependence. Of the two, the spatial lag model fits better, indicating it is the preferred alternative. Given the lack of an underlying behavioral model (unless one is willing to make heroic assumptions to avoid the ecological fallacy problem), the results should be interpreted as providing consistent estimates for the coefficients of income and housing value after the spatial dependence in the crime variable is filtered out. The most affected coefficient (besides the constant term) pertains to the income variable, and is lowered by about a third while remaining highly significant. The estimates for the autoregressive coefficient vary substantially between the inconsistent and biased OLS and the consistent estimates, but the Lag-IV coefficient has a considerably higher standard error. In some instances, OLS can thus yield "better" estimates in an MSE sense relative to IV. Diagnostics in the Lag-ML model indicate strong remaining presence of heteroskedasticity (the spatial Breusch-Pagan test from Anselin

1988a, p. 123, yields a highly significant value of 25.35, $p < 0.00001$). The robust Lag-GIVE estimates support the importance of this effect: the estimate for the autoregressive parameter is quite close to the ML value while obtaining a significantly smaller standard error relative to both OLS and the nonrobust IV. Moreover, the estimate for the Housing variable is no longer significant. This again illustrates the complex interaction between heterogeneity and spatial dependence.

IV. TESTS FOR SPATIAL DEPENDENCE

As it happened in the mainstream econometrics literature, the initial stages of development in spatial econometrics were characterized by an emphasis on estimation. As discussed in the last section, Cliff and Ord (1973) and others formulated the maximum likelihood approach which goes to back to work of Whittle (1954). In mainstream econometrics, the test for serial correlation developed by Durbin and Watson (1950, 1951) was the first explicit specification test for the regression model. It has gained widespread acceptance since its inception. However, routine testing for other specifications (such as homoskedasticity, normality, exogeneity, and functional form) did not take prominence until the early eighties. A major breakthrough was the rediscovery of the Rao (1947) score (RS) test (known as the Lagrange multiplier test in econometrics). The RS test became very popular due to its computational ease compared to the other two asymptotically "equivalent" test procedures, namely the likelihood ratio (LR) and Wald (W) tests (see Godfrey 1988 and Bera and Ullah 1991).

In a similar fashion, the origins of specification testing in spatial econometrics can be traced back to Moran's (1950a, 1950b) test for autocorrelation. This test laid in obscurity until it was revived by Cliff and Ord (1972). It received further impetus by Burridge (1980) as an RS test. However, the early spatial econometrics literature on testing was dominated by the Wald and LR tests (for example, see Brandsma and Ketellapper 1979a, 1979b; Anselin 1980). Since the latter require the estimation of the alternative model by means of nonlinear optimization (as discussed in Section III), the advantages of basing a test on the least-squares regression of the null model, offered by the RS test, were quickly realized. During the last 15 years, a number of such tests were developed (see Anselin 1988a, 1988c).

Although mainstream econometrics and spatial econometrics literature went through similar developments in terms of specification testing, the implementation of the tests in spatial models turns out to be quite different from the standard case. For example, most of the RS specification tests cannot be written in the familiar "NR^2" form (where R^2 is a coefficient of determination) nor they can be computed by running any artificial regression. In addition, the interaction between spatial lag dependence and spatial error dependence in terms of specification testing is stronger and more complex than its standard counterpart. There are, however, some common

threads. As in the standard case, most of the tests for dependence in the spatial model can be constructed based on the OLS residuals. In our discussion we will emphasize the similarities and the differences between specification testing in spatial econometric models and the standard case.

We start the remainder of the section with a discussion of Moran's I statistic and stress its close connection to the familiar Durbin-Watson test. Moran's I was not developed with any specific kind of dependence as the alternative hypothesis, although it has been found to have power against a wide range of forms of spatial dependence. We next consider a test developed in the same spirit by Kelejian and Robinson (1992). This is followed by a focus on tests for specific alternative hypothesis in the form of either spatial lag or spatial error dependence. Tests for these two kinds of autocorrelations are not independent even asymptotically, and their separate applications when other or both kinds of autocorrelations are present will lead to unreliable inference. Therefore, it is natural to discuss a test for joint lag and error autocorrelations. However, the problem with such a test is that we cannot make any specific inference regarding the exact nature of dependence when the joint null hypothesis is rejected. One approach to deal with this problem is to test for spatial error autocorrelation after estimating a spatial lag model, and vice versa. This, however, requires ML estimation, and the simplicity of tests based on OLS residuals is lost. We therefore consider a recently developed set of diagnostics in which the OLS-based RS test for error (lag) dependence is adjusted to take into account the local presence of lag (error) dependence (Anselin et al. 1996). We then provide a brief review of the small-sample properties of the various tests. Finally, the section is closed into a discussion of implementation issues and our illustrative example of the spatial model of crime.

A. Moran's I Test

Moran's (1950a, 1950b) I test was originally developed as a two-dimensional analog of the test of significance of the serial correlation coefficient in univariate time series. Cliff and Ord (1972, 1973) formally presented Moran's I statistics as

$$I = \frac{N}{S_o}\left(\frac{e'We}{e'e}\right) \tag{42}$$

where $e = y - X\tilde{\beta}$ is a vector of OLS residuals, $\tilde{\beta} = (X'X)^{-1}X'y$, W is the spatial weights matrix, N is the number of observations, and S_o is a standardization factor equal to the sum of the spatial weights, $\sum_i \sum_j w_{ij}$. For a row-standardized weights matrix W, S_o simplifies to N (since each row sum equals 1) and the statistic becomes

$$I = \frac{e'We}{e'e} \tag{43}$$

Moran did not derive the statistic from any basic principle; instead, it was suggested as a simple test for correlation between nearest neighbors which generalized one of his earlier tests in Moran (1948). Consequently, the test could be given different interpretations. The first striking characteristic is the similarity between Moran's I and the familiar Durbin-Watson (DW) statistic

$$DW = \frac{e'Ae}{e'e} \tag{44}$$

where

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

Therefore, both statistics equal the ratio of quadratic forms in OLS residuals and they differ only in the specification of the interconnectedness between the observations (neighboring locations). It is well known that the DW test is a uniformly most powerful (UMP) test for one sided alternatives with error distribution $\varepsilon_t = \lambda\varepsilon_{t-1} + \xi_t$ (see, e.g., King 1987). Similarly Moran's I possesses some optimality properties. More precisely, Cliff and Ord (1972) established a link between the LR and I tests. If we take the alternative model as (8), i.e.,

$$\varepsilon = \lambda W\varepsilon + \xi$$

then the LR statistic for testing H_0: $\lambda = 0$ against the alternative H_a: $\lambda = \lambda_1$, when ε and σ^2 are known, is proportional to

$$\frac{\varepsilon' W\varepsilon}{\varepsilon'(\mathbf{I} + \lambda_1^2 G)\varepsilon} \tag{45}$$

where G is a function of W. Therefore, I approaches the LR statistic as $\lambda_1 \to 0$, and it can be shown to be consistent for H_0: $\lambda = 0$ against H_a: $\lambda \neq 0$. As we discuss later, Burridge (1980) also showed that I is equivalent to the RS test for $\lambda = 0$ in (8) (or $\gamma = 0$ in the spatial moving average process (13)) with an unscaled denominator. Since we know that the LR and RS tests are asymptotically equivalent under the null and local alternatives, Cliff and Ord's result regarding asymptotic equivalence of I and LR becomes very apparent. King and Hillier (1985) derived the locally best invariant (LBI) test for the wider problem of testing H_0: $\lambda = 0$ against H_a: $\lambda > 0$ when the covariance matrix of the regression disturbance is of the known form $\sigma^2\Omega(\lambda)$ (as in our (10)), and showed the test to be identical to the one-sided version of the RS test. Combining this result with that of Burridge (1980), we can conclude that Moran's I must be an LBI test, which was demonstrated by King (1981).

In practice the test is implemented on the basis of an asymptotically normal standardized z-value, obtained by subtracting the expected value and dividing by the standard deviation. One advantage of statistic like I is that under $H_0 : \lambda = 0$ and normality of ε, $e'e$ is distributed as central χ^2. Cliff and Ord (1972) exploited this to derive the first two moments as

$$E(I) = \frac{\mathrm{tr}(MW)}{N-K} \tag{46}$$

and

$$V(I) = \frac{\mathrm{tr}(MWMW') + \mathrm{tr}(MW)^2 + \{\mathrm{tr}(MW)\}^2}{(N-K)(N-K+2)} - [E(I)]^2 \tag{47}$$

where $M = \mathbf{I} - X(X'X)^{-1}X'$, and W is a row-standardized weights matrix.

It is possible to develop a finite-sample-bound test for I following Durbin and Watson (1950, 1951). However, for I, we need to make the bounds independent of not only X but also of the weight matrix W. This poses some difficulties. Tiefelsdorf and Boots (1995), using the results of Imhof (1961) and Koerts and Abrahamse (1968), showed that exact critical values of I can be computed by numerical integration. They first expressed I in terms of the eigenvalues $\gamma_1, \gamma_2, \ldots, \gamma_{N-K}$ of MW, other than the K zeros, and $N-K$ independent $N(0,1)$ variables $\delta_1, \delta_2, \ldots, \delta_{N-K}$; more specifically,

$$I = \sum_{i=1}^{N-K} \gamma_i \delta_i^2 \Big/ \sum_{i=1}^{N-K} \delta_i^2 \tag{48}$$

Then

$$\Pr(I \leq I_0 | H_0) = \Pr\left(\sum_{i=1}^{N-K} (\gamma_i - I_0)\delta_i^2 \leq 0 | H_0\right) \tag{49}$$

Note that $\sum_{i=1}^{N-K}(\gamma_i - I_0)\delta_i^2$ is a weighted sum of $(N-K)$ χ_1^2 variables. Imhof's method simplifies the probability in (49) to

$$\Pr(I \leq I_0 | H_0) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\{a(u)\}}{ub(u)\, du} \tag{50}$$

where

$$a(u) = \frac{1}{2}\sum_{i=1}^{N-K} \arctan\{(\gamma_i - I_0)u\}$$

$$b(u) = \prod_{i=1}^{N-K} [1 + (\gamma_i - I_0)^2 u^2]^{1/4}$$

The integral in (50) can be evaluated by numerical integration (for more on this, see Tiefelsdorf and Boots 1995).

It is instructive to note that the computation of exact critical values of the DW statistic involves the same calculations as for Moran's I except that the γ_i is the eigenvalues of MA, where A is the fixed matrix given by in (44). Even with the recent dramatic advances in computer technology, it will take some time for practitioners to use the above numerical integration technique to implement Moran's I test.

B. Kelejian-Robinson Test

The test developed by Kelejian and Robinson (1992) is in the same spirit of Moran's I in the sense that it is not based on an explicit specification of the generating process of the disturbance term. At the same time the test does not require the model to be linear or the disturbance term to be normally distributed. Although the test does not attempt to identify the cause of spatial dependence, Kelejian and Robinson (1992) made the following assumption about spatial autocorrelation:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} = Z_{ij}\alpha \tag{51}$$

where Z_{ij} is 1 by q vector which can be constructed from the independent variables X, α is q by 1 vector of parameters, and i, j are contiguous in the sense that they are neighbors in a general spatial "ordering" of the observations. The null hypothesis of no spatial correlation can be tested by $H_0 : \alpha = 0$ in (51).

For a given sample of size N, let C denote h_N by 1 vector σ_{ij}'s which are not zero for $i < j$. Therefore, a test for $\alpha = 0$ can be achieved by running a regression of C on the observation matrix Z which is of dimension h_N by q consisting of Z_{ij} values. Since we do not observe the elements of C, they are replaced by the cross product of OLS residuals, e_ie_j. The resulting h_N by 1 vector is denoted by $\hat{C}$. The test is based on $\hat{\gamma} = (Z'Z)^{-1}Z'\hat{C}$ and is given by

$$KR = \frac{\hat{\gamma}'Z'Z\hat{\gamma}}{\tilde{\sigma}^4} \tag{52}$$

where $\tilde{\sigma}^4$ is a consistent estimator of σ^4. For example, we can use $[e'e/N]^2$ or $(\hat{C} - Z\hat{\gamma})'(\hat{C} - Z\hat{\gamma})/h_N$ for $\tilde{\sigma}^4$. Under $H_0 : \alpha = 0$, $KR \xrightarrow{\mathcal{D}} \chi^2_q$ (central chi-square with q degrees of freedom), where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Putting $\hat{\gamma} = (Z'Z)^{-1}Z'\hat{C}$, KR can be expressed as

$$KR = \frac{\hat{C}'Z(Z'Z)^{-1}Z'\hat{C}}{\tilde{\sigma}^4} \tag{53}$$

Since for the implementation of the test we need the distribution only under the null hypothesis, it is legitimate to replace σ^4 by a consistent estimate under $\alpha = 0$.

Note that under H_0, $\hat{C}'\hat{C}/h_N \xrightarrow{P} \sigma^4$, where $\xrightarrow{P}$ means convergence in probability. Therefore, an asymptotically equivalent form of the test is

$$h_N \cdot \frac{\hat{C}'Z(Z'Z)^{-1}Z'\hat{C}}{\hat{C}'\hat{C}} \tag{54}$$

which has the familiar NR^2 form. Here R^2 is the uncentered coefficient of determination of $\hat{C}$ on Z and h_N is the sample size of this regression.

It is also not difficult to see an algebraic connection between KR and Moran's I. From (43)

$$\begin{aligned} I^2 &= \frac{(e'We)^2}{(e'e)^2} = \frac{1}{N^2\tilde{\sigma}^4}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}e_ie_j\right)^2 \\ &= \sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} \frac{W_{kl}W_{mn}(e_ke_l)(e_me_n)}{N^2\tilde{\sigma}^4} \end{aligned} \tag{55}$$

Using (53), we can write

$$KR = \sum_{i=1}^{h_N}\sum_{j=1}^{h_N} \frac{p_{ij}\hat{C}_i\hat{C}_j}{\tilde{\sigma}^4} \tag{56}$$

where p_{ij} are the elements of $Z(Z'Z)^{-1}Z'$. Given that $\hat{C}_i$'s contain terms like e_ke_l, $k < l$, it appears that the I^2 and KR statistics have similar algebraic structure.

C. Tests for Spatial Error Autocorrelation

In contrast to the earlier two tests, the alternative hypothesis is now stated explicitly through the data generating process of ε as in (8), i.e.,

$$\varepsilon = \lambda W\varepsilon + \xi$$

and we test $\lambda = 0$. All three general principles of testing, namely LR, W, and RS can be applied. Out of the three, the RS test as described in Rao (1947) is the most convenient one to use since it requires estimation only under the null hypothesis. That is, the RS test can be based on the OLS estimation of the regression model (7). Silvey (1959) derived the RS test using the Lagrange multiplier(s) of a constrained optimization problem.

Burridge (1980) used Silvey's form to test $\lambda = 0$, although the Rao's score form, namely

$$RS = d'(\tilde{\theta})\mathcal{I}(\tilde{\theta})^{-1}d(\tilde{\theta}) \tag{57}$$

is more popular and much easier to use. In (57), $d(\theta) = \partial L(\theta)/\partial\theta$ is the score vector, $\mathcal{I}(\theta) = -E[\partial^2 L(\theta)/\partial(\theta)\partial(\theta)']$ is the information matrix, $L(\theta)$ is the log-likelihood function, and $\tilde{\theta}$ is the *restricted* (under the tested hypothesis) maximum likelihood estimator of the parameter vector θ. For the spatial error autocorrelation model $\theta = (\beta', \sigma^2, \lambda)'$ and the log-likelihood function is given in (34). The test is essentially based on the score with respect to λ, i.e., on

$$d_\lambda = \left.\frac{\partial L}{\partial \lambda}\right|_{\lambda=0} = \frac{\varepsilon' W \varepsilon}{\sigma^2} \tag{58}$$

We can immediately see the connection of this to Moran's I statistic. After computing $\mathcal{I}(\theta)$ under H_0, from (36), we have the test statistic

$$RS_\lambda = \frac{\tilde{d}_\lambda^2}{T} = \frac{[e'We/\tilde{\sigma}^2]^2}{T} \tag{59}$$

where $T = \mathrm{tr}[(W' + W)W]$. Therefore, the test requires only OLS estimates, and under H_0, $RS_\lambda \xrightarrow{\mathcal{D}} \chi_1^2$. It is interesting to put $W = W^T$ (Section II.E) and obtain $T = N - 1$ and $RS_\lambda = (N-1)\tilde{\lambda}^2$ where $\tilde{\lambda} = \sum_t e_t e_{t-1} / \sum_t e_{t-1}^2$ in the time-series context. Burridge (1980) derived the RS test (59) using the estimates of the Lagrange multiplier following Silvey (1959). The Lagrangian function for this problem is

$$L^R(\theta, \mu) = L(\theta) - \mu\lambda \tag{60}$$

where μ is the associated Lagrange multiplier. From the first-order conditions, we have

$$\left.\frac{\partial L^R(\theta, \mu)}{\partial \lambda}\right|_{\tilde{\theta}, \tilde{\mu}} = 0$$

i.e.,

$$\tilde{\mu} = \left.\frac{\partial L(\theta)}{d\lambda}\right|_{\tilde{\theta}} = \tilde{d}_\lambda \tag{61}$$

and this results in the same statistic RS_λ.

A striking feature of the RS test is its invariance to different alternatives (for details, see Bera and McKenzie 1986). The RS test uses the slope $\partial L/\partial\theta$ at $\theta = \tilde{\theta}$, and there may be many likelihood functions (models) which have the same slope at $\tilde{\theta}$. If we specify the alternative hypothesis as a spatial moving-average process (13) and test $H_0 : \gamma = 0$, we obtain the same Rao's score statistic RS_λ. Therefore, RS_λ will be locally optimal for both autoregressive and moving-average alternatives. But this also means that when the null hypothesis is rejected, the test does not provide any guidance regarding the nature of the disturbance process, even when other aspects of the spatial model are resolved. This also raises the question whether RS_λ will be inferior to other asymptotically equivalent tests such as LR and W, with respect to power, since it does not use the precise information contained in the alternative hypothesis. In the context of the standard regression model, Monte Carlo results of Godfrey (1981) and Bera and McKenzie (1986) suggest that there is no setback in the performance of RS test compared to the LR test. In Section IV.G, we discuss the finite sample performance of RS_λ and other tests.

Computationally, the W and LR tests are more demanding since they require ML estimation under the alternative, and the explicit forms of the tests are more complicated. For instance, let us consider the W test which can be computed using the ML estimate $\hat{\lambda}$ by maximizing (34) with respect to β, σ^2, and λ. We can write the W statistic as (Anselin 1988a, p. 104)

$$WS_\lambda = \frac{\hat{\lambda}^2}{\widehat{\mathrm{AsyVar}}(\hat{\lambda})} \tag{62}$$

where AsyVar $(\hat{\lambda})$ can be obtained from (38) as

$$\mathrm{AsyVar}(\hat{\lambda}) = \left[\mathrm{tr}(W_B^2) + \mathrm{tr}(W_B' W_B) - \frac{\{\mathrm{tr}(W_B)\}^2}{N}\right]^{-1} \tag{63}$$

For implementation we need to replace λ by $\hat{\lambda}$ in the above expression. In the standard time-series regression case the results are much simpler. For example, AsyVar $[\sigma^2, \lambda]$ is a diagonal matrix and $\mathrm{AsyVar}(\hat{\lambda})$ is simply $(1 - \lambda^2)/(N - 1)$. Therefore the Wald test statistic can be simply written as

$$WS_\lambda^T = \frac{(N-1)\hat{\lambda}^2}{1 - \hat{\lambda}^2} \tag{64}$$

Note that under $\lambda = 0$, the asymptotic variance $(1 - \lambda^2)/(N - 1)$ reduces to $1/(N - 1)$, the expression for $\mathrm{AsyVar}(\hat{\lambda})$ used in the time series case to test the significance of λ.

The LR statistic can be easily obtained using the concentrated log-likelihood function L_C in (37). We can write

$$LR_\lambda = 2[\hat{L}_C - \tilde{L}_C] \tag{65}$$

where the "hat" denotes that the quantities are evaluated at the unrestricted ML estimates $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{\lambda}$. It is easy to see that LR_λ reduces to (Anselin 1988a, p. 104)

$$LR_\lambda = N[\ln \tilde{\sigma}^2 - \ln \hat{\sigma}^2] + 2\sum_{i=1}^{N} \ln(1 - \hat{\lambda}\omega_i) \tag{66}$$

The appearance of the last term in (66) differentiate the spatial dependence situation from the serial correlation case for time-series data.

Finally, for higher-order spatial processes, it is easy to generalize the RS statistic (59). For example, if we consider a qth-order spatial autoregressive model

$$\varepsilon = \lambda_1 W_1 \varepsilon + \lambda_1 W_2 \varepsilon + \cdots + \lambda_q W_q \varepsilon + \xi \tag{67}$$

and test $H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_q = 0$, the RS statistic will be given by

$$RS_{\lambda_1 \ldots \lambda_q} = \sum_{l=1}^{q} \frac{[e' W_l e/\tilde{\sigma}^2]^2}{T_l} \tag{68}$$

where $T_l = \mathrm{tr}[W_l' W_l + W_l^2]$, $l = 1, 2, \ldots, q$. Under the null of no spatial dependence, $RS_{\lambda_1 \ldots \lambda_q} \xrightarrow{\mathcal{D}} \chi^2_q$. Therefore, the test statistic for higher-order dependence is simply the sum of corresponding individual tests. The same test statistic will result when a moving average model as in (18) is taken as the alternative instead of (67). As expected, the Wald and LR tests in this context will be more complicated as they require ML estimation of $\lambda_1, \lambda_2, \ldots, \lambda_q$.

D. Tests for Spatial Lag Dependence

In this section, we consider tests on the null hypothesis $H_0 : \rho = 0$ in (3) using the log-likelihood function (26). Once again the RS test is the easiest one to use, and Anselin (1988c) derived it explicitly (his equation (32)). The score with respect to ρ is

$$d\rho = \left.\frac{\partial L}{\partial \rho}\right|_{\rho=0} = \frac{\varepsilon' W y}{\sigma^2} \tag{69}$$

The inverse of the information matrix is given in (30). The complicating feature of this matrix is that even under $\rho = 0$, it is not block diagonal; the (ρ, β) term is equal to $(X'WX\beta)/\sigma^2$, obtained by putting $\rho = 0$; i.e., $W_A = W$. This absence of block diagonality causes two problems. First, as we mentioned in Section II, the presence of spatial dependence implies that a sample contains less information than an independent counterpart. This can now be easily demonstrated using (30). In the absence of dependence ($\rho = 0$ in (3)), the ML estimate of β will have variance $\sigma^2(X'X)^{-1}$ which is the inverse of the information. But when $\rho \neq 0$, to compute the variance of the ML estimate of β we need to add a positive-definite part to $\sigma^2(X'X)^{-1}$ due to absence of block diagonality. Second, to obtain the asymptotic variance of d_ρ, even under $\rho = 0$ from (30), we cannot ignore one of the off-diagonal terms. This was not the case for d_λ in Section IV.C. Asymptotic variance of d_λ was obtained just using the (2, 2) element of (36) (see (59)). For the spatial lag model, asymptotic variance of d_ρ is obtained from the reciprocal of the (1,1) element of

$$\begin{bmatrix} \mathrm{tr}[W^2 + W'W] + [WX\beta]'[WX\beta]/\sigma^2 & (X'WX\beta)'/\sigma^2 \\ (X'WX\beta)/\sigma^2 & (X'X)/\sigma^2 \end{bmatrix}^{-1} \tag{70}$$

Since under $\rho = 0$, $W_A = W$ and $\mathrm{tr}(W) = 0$, the expression is $T_1 = [(WX\beta)' M(WX\beta) + T\sigma^2]/\sigma^2$, where T is given in (59). Therefore, the RS statistic is given by

$$RS_\rho = \frac{\tilde{d}_\rho^2}{\tilde{T}_1} = \frac{[e'Wy/\tilde{\sigma}^2]^2}{\tilde{T}_1} \tag{71}$$

where in $\tilde{T}_1$, β, and σ^2 are replaced by $\tilde{\beta}$ and $\tilde{\sigma}^2$, respectively. Under $H_0 : \rho = 0$, $RS_\rho \xrightarrow{\mathcal{D}} \chi^2_1$, the Wald and LR tests will require maximization of the log-likelihood function (26) or (29). Let $\hat{\rho}$ be the ML estimate of ρ. To get the asymptotic variance of $\hat{\rho}$, we need the $(1, 1)$ element of (30). Since the Wald test requires estimation under the alternative hypothesis (i.e., $\rho \neq 0$), the $(1, 3)$ element $\mathrm{tr}(W_A)/\sigma^2$ will also be nonzero and the resulting expression will more complicated than T_1 given above (Anselin 1988a, p. 104). The LR statistic will have the same form as in (66) except for the last term:

$$LR_\rho = N[\ln \tilde{\sigma}^2 - \ln \hat{\sigma}^2] + 2\sum_{i=1}^{N} \ln(1 - \hat{\rho}\omega_i) \tag{72}$$

If ML estimation is already performed, LR_ρ is much easier to compute than its Wald counterpart. Under $\rho = 0$ both Wald and LR statistics will be asymptotically distributed as χ^2_1.

E. Testing in the Possible Presence of Both Spatial Error and Lag Autocorrelation

The test described in the Sections IV.C and IV.D can be termed as *one-directional* tests in the sense that they are designed to test a single specification assuming correct specification for the rest of the model. For example, we discussed RS_λ, WS_λ, and LR_λ statistics for the null hypothesis $H_0 : \lambda = 0$ assuming that $\rho = 0$. Because of the nature of the information matrix, these tests will not be valid even asymptotically, when $\rho \neq 0$. For instance, we noted that under the null, $H_0 : \lambda = 0$ all the three statistics are asymptotically distributed as *central* χ^2 with one degree of freedom. This result is valid only when $\rho = 0$. To evaluate the effects of nonzero ρ on RS_λ, WS_λ, and LR_λ, let us write the model when both the spatial error and lag autocorrelation are present:

$$\begin{aligned} y &= \rho W_1 y + X\beta + \varepsilon \\ \varepsilon &= \lambda W_2 \varepsilon + \xi \qquad \xi \sim N(0, \sigma^2 I) \end{aligned} \tag{73}$$

where W_1 and W_2 are spatial weights matrices associated with the spatially lagged dependent variable and the spatial autoregressive disturbances, respectively. Recall from Section II.F that for model (73) to be identified, it is necessary that $W_1 \neq W_2$ or that the matrix X contain at least one exogenous variable in addition to the constant term. An alternative specification of spatial moving-average error process for ε as in (13),

$$\varepsilon = \lambda W_2 \xi + \xi \tag{74}$$

has no such problems and it also leads to identical results in terms of test statistics discussed here. Using the results of Davidson and MacKinnon (1987) and Saikkonen (1989), we evaluate the impact of local presence of ρ on the asymptotic null distribution of RS_λ, LR_λ, and WS_λ. Let $\rho = \delta/\sqrt{N}$, $\delta < \infty$, then it can be shown that under $H_0 : \lambda = 0$, all three tests asymptotically converge to a *noncentral* χ_1^2, with noncentrality parameter

$$R_\rho = \frac{\delta^2 T_{12}^2}{N T_{22}} \tag{75}$$

where $T_{ij} = \mathrm{tr}[W_i W_j + W_i' W_j]$, $j = 1, 2$ (note that $T_{12} = T_{21}$). Therefore, the tests will reject the null of error autocorrelation even when $\lambda = 0$ due to the local presence of the lag dependence. In a similar way we can express the asymptotic distributions of RS_ρ, LR_ρ, and WS_ρ. Under $\rho = 0$ and local presence of error dependence, say, $\lambda = \tau/\sqrt{N}$, $\tau < \infty$. In this case the distributions remain χ_1^2, but with a noncentrality parameter

$$R_\lambda = \frac{\tau^2 T_{12}^2 \sigma^2}{N\, D} \tag{76}$$

where $D = (W_1 X\beta)' M (W_1 X\beta) + T_{11}\sigma^2$. Therefore, again we will have unwanted "power" due to the presence of local error dependence. In the noncentrality parameters R_ρ and R_λ, the crucial quantity is $T_{12}/\sqrt{N}$, which can be interpreted as the covariance between the scores d_λ and d_ρ. Note that if $T_{12} = 0$, then both R_ρ and R_λ vanish, and local presence of one kind of dependence cannot affect the test for the other one. The trace term $T_{12} = \mathrm{tr}[W_1 W_2 + W_1' W_2]$, which will only be zero when the nonzero elements in each row/column of the weights matrices W_1 and W_2 do not overlap. In other words, this will be the case when the pattern of spatial dependence in the lag term and in the error term pertain to a completely different set of neighbors for each observation. However, in the typical case where $W_1 = W_2$ (or overlap to any extent) then the noncentrality parameter will not vanish.

For valid statistical inference there is a need to take account of possible lag dependence while we test for error dependence, and vice versa. In Anselin (1988c) two different approaches are suggested. One is to test jointly for $H_0 : \lambda = \rho = 0$ in (73) using the RS principle so that the test can be implemented with OLS residuals (see Anselin 1988c). The resulting joint test statistic is given by

$$RS_{\lambda\rho} = \tilde{E}^{-1}\left[(\tilde{d}_\lambda)^2 \frac{\tilde{D}}{\tilde{\sigma}^2} + (\tilde{d}_\rho)^2 T_{22} - 2\tilde{d}_\lambda \tilde{d}_\rho T_{12}\right] \tag{77}$$

where $E = (D/\sigma^2) T_{22} - (T_{12})^2$. Note that this joint test not only depends on d_λ and d_ρ but also on their interaction factor with a coefficient T_{12}. Expression (77) appears to be somewhat complicated but can be computed quite easily using only OLS residuals. Also the expression simplifies greatly when the spatial weights matrices W_1 and W_2 are assumed to be the same which is the case in most applications. Under $W_1 = W_2 = W$, $T_{11} = T_{21} = T_{22} = T = \mathrm{tr}[(W' + W)W]$, and (77) reduces to

$$RS_{\lambda\rho} = \frac{\tilde{d}_\lambda^2}{T} + \frac{(\tilde{d}_\lambda - \tilde{d}_\rho)^2}{\tilde{\sigma}^{-2}(\tilde{D} - T\tilde{\sigma}^2)} \tag{78}$$

Under $H_0 : \lambda = \rho = 0$, $RS_{\lambda\rho}$ will converge to a central χ^2 with *two* degrees of freedom. Because of this two degrees of freedom, the statistic will result in loss of power compared to the proper one-directional test when only one of the two forms of misspecification is present. To see this consider the presence of only $\lambda = \tau/\sqrt{N}$, with $\rho = 0$. In this case the noncentrality parameter for both RS_λ and $RS_{\lambda\rho}$ is the same $\tau^2 NT$. Due to the higher degrees of freedom of the joint test $RS_{\lambda\rho}$, we can expect some loss of power (Dasgupta and Perlman 1974). Another problem with $RS_{\lambda\rho}$ is that since it is an omnibus test, if the null hypothesis is rejected, it is not possible to infer whether the misspecification is due to lag or error dependence.

A second approach is to carry out an RS test for one form of misspecification in a model where the other form is unconstrained. For example, this consists of testing the null hypothesis $H_0 : \lambda = 0$ in the presence of ρ, i.e., based on the residuals of a maximum likelihood estimation of the spatial lag model. The resulting statistic $RS_{\lambda|\rho}$ is given as

$$RS_{\lambda|\rho} = \frac{\hat{d}_\rho^2}{T_{22} - (T_{21A})^2 \widehat{\mathrm{Var}}(\hat{\rho})} \tag{79}$$

where the "hat" denotes quantities are evaluated at the maximum likelihood estimates of the parameters of the model $Y = \rho W_1 y + X\beta + \xi$ obtained by means of nonlinear optimization. In (79) T_{21A} stands for $\mathrm{tr}[W_2 W_1 A^{-1} + W_2' W_1 A^{-1}]$, with $A = I - \hat{\rho} W_1$. Under $H_0 : \lambda = 0$, $RS_{\lambda/\rho}$ will converge to a central χ^2 with one degree of freedom. Similarly, an RS test can be developed for $H_0 : \rho = 0$ in the presence of error dependence (Anselin et al. 1996). This test statistic can be written as

$$RS_{\rho|\lambda} = \frac{[\hat{\varepsilon}' B' B W_1 y]^2}{H_\rho - H_{\theta\rho} \widehat{\mathrm{Var}}(\hat{\theta}) H_{\theta\rho}'} \tag{80}$$

where $\hat{\varepsilon}$ is a vector of residuals in the ML estimation of the null model with spatial AR errors, $y = X\beta + (\mathbf{I} - \lambda W_2)^{-1}\xi$ with $\theta = (\beta', \lambda, \sigma^2)'$, and $B = \mathbf{I} - \hat{\lambda} W_2$. The terms in the denominator of (80) are

$$H_\rho = \text{tr } W_1^2 + \text{tr}(BW_1B^{-1})'(BW_1B^{-1}) + \frac{(BW_1X\beta)'(BW_1X\beta)}{\sigma^2}$$

$$H'_{\theta\rho} = \begin{bmatrix} \dfrac{(BX)'BW_1X\beta}{\sigma^2} \\ \text{tr}(W_2B^{-1})'BW_1B^{-1} + \text{tr } W_2W_1B^{-1} \\ 0 \end{bmatrix}$$

and $\widehat{\text{Var}}(\hat{\theta}$ is the estimated variance-covariance matrix for the parameter vector θ.

It is also possible to obtain the W and LR statistics in the above three cases, though these will involve the estimation of a spatial model with two parameters, requiring considerably more complex nonlinear optimization. In contrast, $RS_{\lambda|\rho}$ and $RS_{\rho|\lambda}$ are theoretically valid statistics that have the potential to identify the possible source(s) of misspecification and can be derived from the results of the maximization of the log-likelihood functions (32) and (26). However, this is clearly more computationally demanding than tests based on OLS residuals. We now turn to an approach that accomplishes carrying out the tests without maximum likelihood estimation of λ and ρ.

F. Robust Test in the Presence of Local Misspecification

It is not possible to robustify tests in the presence of *global* misspecification (i.e., λ and ρ taking values far away from zero). However, using the general approach of Bera and Yoon (1993), Anselin et al. (1996) suggested tests which are robust to *local* misspecifications, as defined in the previous subsection. The idea is to adjust the one-directional score tests RS_λ and RS_ρ by taking account of the noncentrality parameters R_ρ and R_λ, given in (75) and (76), so that under the null the resulting test statistics have *central* χ_1^2 distributions.

The modified test for $H_0 : \lambda = 0$ in the local presence of ρ is given by

$$RS_\lambda^* = \frac{[\tilde{d}_\lambda - T_{12}\tilde{\sigma}^2\tilde{D}^{-1}\tilde{d}_\rho]^2}{T_{22} - (T_{12})^2\tilde{\sigma}^2\tilde{D}} \tag{81}$$

When $W_1 = W_2 = W$, RS_λ^* becomes

$$RS_\lambda^* = \frac{[\tilde{d}_\lambda - T\tilde{\sigma}^2\tilde{D}^{-1}\tilde{d}_\rho]^2}{T(1 - T\tilde{\sigma}^2\tilde{D})} \tag{82}$$

Comparing RS_λ^* in (81) and RS_λ in (59), it is clear that the adjusted test modifies RS_λ by correcting for the presence of ρ through $\tilde{d}_\rho$ and T_{12}, where the latter quantity represents the covariance between d_λ and d_ρ. Under H_0: $\lambda = 0$ (and $\rho = \delta/\sqrt{N}$), RS_λ^* converges to a *central* χ_1^2 distribution; i.e., RS_λ^* has the same asymptotic distribution as RS_λ based on the correct specification. This therefore produces asymptotically the correct size in the presence of local lag dependence. Also as noted for RS_λ^*,

we only need OLS estimation thus circumventing direct estimation of the nuisance parameter ρ. However, there is a price to be paid for robustification and simplicity in estimation. Consider the case when there is no lag dependence ($\rho = 0$), but only spatial error dependence through $\lambda = \tau/\sqrt{N}$. Under this setup, the noncentrality parameters of RS_λ and RS_λ^* are respectively τ^2T_{22}/N and $\tau^2(T_{22} - T_{12}^2\sigma^2D^{-1})/N$. Since $\tau^2T_{12}^2\sigma^2D^{-1}/N \geq 0$, the asymptotic power of RS_λ^* will be less than that of RS_λ when there is no lag dependence. The above quantity can be regarded as a cost of robustification. Once again, note its dependence on T_{12}. It is also instructive to compare RS_λ^* with Anselin's $RS_{\lambda|\rho}$ in (79). It is readily seen that $RS_{\lambda|\rho}$ does not have the mean correction factor. $RS_{\lambda|\rho}$ uses the restricted ML estimator of ρ (under $\lambda = 0$) for which $\hat{d}_\rho = 0$. We may view $RS_{\lambda|\rho}$ as the spatial version of Durbin's h statistic, which can also be derived from the general RS principle. Unlike Durbin's h, however, $RS_{\lambda|\rho}$ cannot be computed using the OLS residuals.

In a similar way, the adjusted score test for $H_0 : \rho = 0$, in the presence of local misspecification involving spatial-dependent error process can be expressed as

$$RS_\rho^* = \frac{[\tilde{d}_\rho - T_{12}T_{22}^{-1}\tilde{d}_\lambda]^2}{\tilde{\sigma}^{-2}\tilde{D} - (T_{12})^2T_{22}^{-1}} \tag{83}$$

Under $W_1 = W_2 = W$, the above expression simplifies to

$$RS_\rho^* = \frac{[\tilde{d}_\rho - \tilde{d}_\lambda]^2}{\tilde{\sigma}^{-2}\tilde{D} - T} \tag{84}$$

All our earlier discussion of RS_λ^* also applies to RS_ρ^*.

Finally, consider the relationship among the five statistics RS_λ, RS_ρ, RS_λ^*, RS_ρ^*, and $RS_{\lambda\rho}$ given in (59), (71), (82), (84), and (78) respectively. $RS_{\lambda\rho}$ is not the sum of RS_λ and RS_ρ; i.e., there is no additivity of the score tests along the lines discussed in Bera and Jarque (1982) and Bera and McKenzie (1987). From (77), it is clear that additivity follows only if $T_{12} = 0$ or $T = 0$ for the case of $W_1 = W_2$, i.e., when d_λ and d_ρ are asymptotically uncorrelated. In that case also $RS_\lambda^* = RS_\lambda$ and $RS_\rho^* = RS_\rho$ (see (81), (59), (83), and (71)). Hence, for $T = 0$, the conventional one-directional tests RS_λ and RS_ρ are asymptotically valid in the presence of local misspecification. However, as noted earlier $T > 0$ and $T_{12} > 0$ when W_1 and W_2 have some overlap in the neighbor structure. Under these circumstances (which are the most common situation encountered in practice), the following very intriguing result is obtained:

$$RS_{\lambda\rho} = RS_\lambda^* + RS_\rho = RS_\lambda + RS_\rho^* \tag{85}$$

i.e., the two-directional test for λ and ρ can be decomposed into the sum of the adjusted one-directional test of one type of alternative and the unadjusted form for the other. By construction, under $\lambda = \rho = 0$, RS_λ^* and RS_ρ are asymptotically independently distributed, which cannot be said about RS_λ and RS_ρ. By applying all the unadjusted and adjusted tests and exploiting the result (85), it is possible to identify the exact nature of dependence in practice (Anselin et al. 1996). Finally, we should mention that because of the complexity of the Wald and LR tests, it is not possible to derive their adjusted versions that would be valid under local misspecification. Of course, it is not computationally prohibitive to obtain these tests after the joint estimation of both λ and ρ.

G. Small Sample Properties

We have covered a number of procedures for testing spatial dependence. For ease of implementation, we have emphasized Rao's score test which in many cases can be computed based on the OLS residuals. As we indicated, all these tests are of asymptotic nature; i.e., their justification derives from the presence of very large samples. That is, however, not the case in most applications. The small sample performance of the above tests both in terms of size and power is of major concern to practitioners.

There are only a few papers on the finite sample properties of tests on spatial dependence compared to the vast literature on those for testing for serial correlation for time-series data as summarized in King (1987). Bartels and Hordijk (1977) studied the behavior of Moran's I. However, their focus was on the performance of different residuals, and they found that OLS residuals give the best results. Brandsma and Ketellapper (1979b) included the LR test (LR_λ) in their study, but it performed poorly compared to I. Both these studies were quite limited in terms of a small number of replications, few sample sizes, the use of only one type (irregular) weights matrix and the narrow range of alternative values for the autocorrelation coefficient. A first extensive set of Monte Carlo simulations was carried out by Anselin and Rey (1991), who compared Moran's I to RS_λ and RS_ρ for different weights matrices and error distributions. In terms of size, the small sample distributions of the statistics corresponded close to their theoretical counterparts, except for the smallest size ($N = 25$). In terms of power, Moran's I had power against both kinds of dependence, spatial lag and error autocorrelations. RS_λ and RS_ρ had highest power for their respective designated alternatives. These tests were found to possess superior performance, but they fall short of providing a good strategy for identifying the exact nature of dependence.

Anselin and Florax (1995b) provide the most comprehensive set of simulation results to date. They carried out experiments for both regular (rook and queen) and nonregular weight matrices, single- and multidirectional alternatives, and for different error distributions, and included all the tests discussed earlier except the Wald and LR tests. The results are too extensive to discuss in detail, and here we provide only a brief summary of the main findings. First, the earlier results of Anselin and Rey (1991) were confirmed on the power of I against any form of dependence and the optimality of the RS tests against the alternatives for which they were designed. Second, the specification of the spatial weights matrix impacted the performance of

all tests, with a higher power obtained in the rook case. Third, as in Anselin and Rey (1991), higher powers were achieved by lag tests relative to tests against error dependence. This is important, since the consequences of ignoring lag dependence are more serious. Fourth, the KR statistic did not perform well. For example, when the errors were generated as lognormal, it significantly over rejected the true null hypothesis in all configurations. There are two possible explanations. One is its higher degrees of freedom. Another is that its power depends on the degree of autocorrelation in the explanatory variables which substitute for the weights matrix (compare (55) and (56)). It is interesting to note that White's (1980) test for heteroskedasticity which is very similar to KR encounters problems of the same type. Fifth, the most striking result is that the adjusted tests RS^*_λ and $RS\rho^*$ performed remarkably well. They had reasonable empirical sizes, remaining within the confidence interval in all cases. In terms of power they performed exactly the way they were supposed to. For instance, when the data were generated under $\rho > 0$, $\lambda = 0$, although RS_ρ had the most power, the powers of RS^*_ρ was very close to that of RS_ρ. That is, the price paid for adjustments that were not needed turned out to be small. The real superiority of RS^*_ρ was revealed when $\lambda > 0$ and $\rho = 0$. It yielded low rejection frequencies even for $\lambda = 0.9$. The correction for error dependence in RS^*_ρ worked in the right direction when no lagged dependence was present for all configurations. When $\rho > 0$, the power function of RS^*_ρ was seen to be almost unaffected by the values of λ, even for those far away from zero (global misspecification). For these alternatives $RS_{\lambda\rho}$ also had good power, but could not point to the correct alternative when only one kind of dependence is present. RS^*_λ also performed well though not as spectacularly as RS^*_ρ. The adjusted tests thus seem more appropriate to test for lag dependence in the presence of error correlation than for the reverse case. Again, this is important since ignoring lag dependence has more severe consequences. Based on these results Anselin and Florax (1995b) suggested a simple decision rule. When RS_ρ is more significant than RS_λ, and RS^*_ρ is significant while RS^*_λ is not, a lag dependence is the likely alternative. In a similar way presence of error dependence can be identified through RS^*_λ. Finally, the finite-sample performance of tests against higher-order dependence $RS_{\lambda_1\lambda_2}$ (see (68)) and the joint test $RS_{\lambda\rho}$ were satisfactory, although these type of tests provide less insightful guidance for an effective specification search. For joint and higher-order alternatives, these tests are optimal, and in practice they should be used along with the unadjusted and adjusted one-directional tests.

H. Operational Implementation and Illustration

As is the case for the estimation of spatial regression models, specification tests for spatial dependence are notably absent from econometric software, with the exception of SpaceStat (Anselin 1992b, 1995). Moreover, as pointed out, these tests cannot be obtained in the usual NR^2 format, which lends itself to straightforward implementation by means of auxiliary or augmented regressions. The closest to this situation is the Kelejian-Robinson test (54), provided one has an easy way to select the pairs of neighboring data points from the data. Typically, specification tests for spatial dependence must be implemented explicitly either by writing special-purpose software or by taking advantage of macros in econometric and statistical software. As in maximum likelihood estimation, the size of the weights matrix may be a constraint when the number of observations is large. This is particularly the case for Moran's I, where several operations are involved in the computation of the expected value and variance (46) and (47). Examples of the implementation of this test for small data sets in standard econometric software are given in Anselin and Hudak (1992) and Anselin et al. (1993b), for Shazam, Rats and Limdep, among others.

Given their importance for applied work, we now briefly describe implementation strategies for the RS tests for spatial error and spatial lag autocorrelation, RS_λ (59) and RS_ρ (71). First, note that the squared expression in the numerator equals N times a regression coefficient of an auxiliary regression of respectively We on e (in (59)) and Wy on e (in (71)). Once the lags are constructed, these coefficients can be obtained using standard software. The denominator in the expressions is slightly more complex. The trace elements $T = \mathrm{tr}(WW) + \mathrm{tr}(W'W)$ can easily be seen to equal, respectively, $\sum_i \sum_j w_{ij}.w_{ji}$ and $\sum_i \sum_j (w_{ij})^2$. When the spatial weights matrix consists of simple row-standardized contiguity weights, each element w_{ij} for a given i equals $1/k_i$, where k_i is the number of neighbors for observation i. Hence, $\sum_i \sum_j (w_{ij})^2 = \sum_i (1/k_i)$, which can easily be computed. The other trace term is $\sum_i \sum_j w_{ij}.w_{ji} = \sum_i (1/k_i)[\sum_j \delta_{ij}/k_j]$, where δ_{ij} is a binary variable indicating whether or not $w_{ij} \neq 0$. This requires only slighly more work to compute, similar to the sorting needed to establish the neighbor pairs in the Kelejian-Robinson test. Most importantly, the trace operations can be carried out without having to store a full matrix in memory, taking advantage of the sparse nature of spatial weights (for technical details, see Anselin 1995). Of course, for symmetric weights, the two traces are equal. In practice, this may occur when all observations are considered to have an equal number of neighbors, as in Pace and Barry (1996). The other term in the denominator of (71) is the residual sum of squares in a regression with WXb (i.e., the spatial lags for the predicted values from the OLS regression) on X, which can be obtained in a straightforward way.

To illustrate the various specification tests, we list the results of Moran's I, KR, and the RS and LR tests for the spatial model of crime in Table 2 (using a slightly different notation, most of these results are reported in Table 2, p. 87 of Anselin et al. 1996). All results are part of the standard SpaceStat regression diagnostic output. They reflect a situation that is often encountered in practice: strong significance of Moran's I and KR, as well as of both one-directional RS and LR tests. Clearly, spatial dependence is a problem, although without further investigation it is not obvious which form of spatial dependence is the proper alternative. Convincing evidence is provided by the robust tests RS^*_λ and RS^*_ρ. While the former is not at all significant,

Table 2 Specification Tests against Spatial Dependence[a]

Estimates	Test (equation number)	Value	p-value
OLS	Moran's I (z-value) (43)	2.95	0.003
OLS	Kelejian-Robinson (54)	11.55	0.009
OLS	$RS_{\lambda\rho}$ (78)	9.44	0.009
OLS	RS_λ (59)	5.72	0.02
OLS	RS^*_λ (82)	0.08	0.78
OLS	RS_ρ (71)	9.36	0.002
OLS	RS^*_ρ (84)	3.72	0.05
Lag-ML	LR_ρ (72)	9.97	0.002
Lag-ML	$RS_{\lambda\mid\rho}$ (79)	0.32	0.57
Err-ML	LR_λ (66)	7.99	0.005
Err-ML	$RS_{\rho\mid\lambda}$ (80)	1.76	0.18

[a] *Source*: From Anselin (1988a, Chap. 12; 1992a, Chap. 26; 1995) and Anselin et al. (1996).

the latter is significant at ρ slightly higher than 0.05. In other words, the impression of spatial error autocorrelation that may be given by an uncritical interpretation of Moran's I is spurious, since no evidence of such autocorrelation remains after robustifying for spatial lag dependence. Instead, a spatial lag model is the suggested alternative, consistent with the estimation results in Table 1.

V. CONCLUSIONS

In our review of methods to deal with spatial dependence in regression analysis, we have emphasized the distinguishing characteristics of spatial econometrics relative to time-series analysis. We highlighted the concept of spatial weights and the associated spatial lag operator which allow for the formal specification of neighbors in space, a much more general concept than its counterpart in time. In the estimation of spatial regression models, the maximum likelihood approach was shown to be prevalent and requiring nonlinear optimization of the likelihood function. The simplifying results from serial correlation in time series do not hold and estimation necessitates the explicit manipulation of matrices of dimension equal to the number of observations. Diagnostics for spatial effects in regression models may be based on the powerful score principle, but they do not boil down to simple significance tests of the coefficients in an auxiliary regression, as they do for time series.

The differences between the time domain and space are both puzzling and challenging, in terms of theory as well as from an applied perspective. They are the subject of active research efforts to develop diagnostics for multiple sources of misspecification, to discriminate between heterogeneity and spatial dependence, and to estimate models for complex forms of interaction in realistic data settings. Extensions to the space-time domain and to models for limited dependent variables are particularly challenging. We hope that our review of the fundamental concepts and basic methods will stimulate others to both apply these techniques as well as to pursue solutions for the remaining research questions.

ACKNOWLEDGMENTS

We would like to thank Aman Ullah and an anonymous referee for helpful suggestions, and Robert Rozovsky for very able research assistance. We also would like to thank Naoko Miki for her help in preparing the manuscript. However, we retain the responsibility for any remaining errors. The first author acknowledges ongoing support for the development of spatial econometric methods by the U.S. National Science Foundation, notably through grants SES 87-21875, SES 89-21385, and SBR 94-10612 as well as grant SES 88-10917 to the National Center for Geographic Information and Analysis (NCGIA). The second author acknowledges financial support by the Bureau of Economic and Business Research of the University of Illinois.

REFERENCES

Albert, P. and L. M. McShane (1995), A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data, *Biometrics*, 51, 627–638.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, MA.

Ancot, J-P., J. Paelinck, and J. Prins (1986), Some New Estimators in Spatial Econometrics, *Economics Letters*, 21, 245–249.

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

Anselin, L. (1982), A Note on Small Sample Properties of Estimators in a First-Order Spatial Autoregressive Model, *Environment and Planning A*, 14, 1023–1030.

Anselin, L. (1988a), *Spatial Econometrics: Methods and Models*, Kluwer, Dordrecht.

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), Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity, *Geographical Analysis*, 20, 1–17.

Anselin, L. (1990a), Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis, *Journal of Regional Science*, 30, 185–207.

Anselin, L. (1990b), Some Robust Approaches to Testing and Estimation in Spatial Econometrics, *Regional Science and Urban Economics*, 20, 141–163.

Anselin, L. (ed.) (1992a), *Space and Applied Econometrics*. Special Issue, *Regional Science and Urban Economics*, 22.

Anselin, L. (1992b), *SpaceStat: A Program for the Analysis of Spatial Data*, National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Anselin, L. (1995), *SpaceStat Version 1.80 User's Guide*, Regional Research Institute, West Virginia University, Morgantown, WV.

Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon (1996), Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics*, 26, 77–104.

Anselin, L. and A. Can (1996), Spatial Effects in Models of Mortgage Origination, Paper presented at the Mid Year AREUEA Conference, Washington, DC, May 28–29.

Anselin, L., R. Dodson, and S. Hudak (1993a), Linking GIS and Spatial Data Analysis in Practice, *Geographical Systems*, 1, 3–23.

Anselin, L., R. Dodson, and S. Hudak (1993b), *Spatial Data Analysis and GIS: Interfacing GIS and Econometric Software*, Technical Report 93-7, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Anselin, L. and R. Florax (eds.) (1995a), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin.

Anselin, L. and R. Florax (1995b), Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results, in L. Anselin and R. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 21–74.

Anselin, L. and S. Hudak (1992), Spatial Econometrics in Practice, a Review of Software Options, *Regional Science and Urban Economics*, 22, 509–536.

Anselin, L. and S. Rey (1991), Properties of Tests for Spatial Dependence in Linear Regression Models, *Geographic Analysis*, 23, 112–131.

Anselin, L. and S. Rey (eds.) (1997), *Spatial Econometrics*. Special Issue, *International Regional Science Review*, 20.

Anselin, L. and O. Smirnov (1996), Efficient Algorithms for Constructing Proper Higher Order Spatial Lag Operators, *Journal of Regional Science*, 36, 67–89.

Arora, S. and M. Brown (1977), Alternative Approaches to Spatial Autocorrelation: An Improvement over Current Practice, *International Regional Science Review*, 2, 67–78.

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.

Bartels, C. and R. Ketellapper (eds.) (1979), *Exploratory and Explanatory Analysis of Spatial Data*, Martinus Nijhoff, Boston.

Benirschka, M. and J. K. Binkley (1994), Land Price Volatility in a Geographically Dispersed Market, *American Journal of Agricultural Economics*, 76, 185–195.

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

Bera, A. K. and C. R. McKenzie (1986), Alternative Forms and Properties of the Score Test, *Journal of Applied Statistics*, 13, 13–25.

Bera, A. K. and C. R. McKenzie (1987), Additivity and Separability of the Lagrange Multiplier, Likelihood Ratio and Wald Tests, *Journal of Quantitative Economics*, 3, 53–63.

Beron, K. J., J. C. Murdoch, and W. P. M. Vijverberg (1996), Why Cooperate? An Interdependent Probit Model of Network Correlations, Working Paper, School of Social Sciences, University of Texas at Dallas, Richardson, TX.

Bera, A. K. and A. Ullah (1991), Rao's Score Test in Econometrics, *Journal of Quantitative Economics*, 7, 189–220.

Bera, A. K. and M. J. Yoon (1993), Specification Testing with Misspecified Alternatives, *Econometric Theory*, 9, 649–658.

Besag, J. (1974), Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society*, B, 36, 192–225.

Besley, T. and A. Case (1995), Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition, *American Economic Review*, 85, 25–45.

Bivand, R. (1992), Systat Compatible Software for Modeling Spatial Dependence among Observations, *Computers and Geosciences*, 18, 951–963.

Blommestein, H. (1983), Specification and Estimation of Spatial Econometric Models: A Discussion of Alternative Strategies for Spatial Economic Modelling, *Regional Science and Urban Economics*, 13, 250–271.

Blommestein, H. (1985), Elimination of Circular Routes in Spatial Dynamic Regression Equations, *Regional Science and Urban Economics*, 15, 121–130.

Blommestein, H. J. and N. A. Koper (1992), Recursive Algorithms for the Elimination of Redundant Paths in Spatial Lag Operators, *Journal of Regional Science*, 32, 91–111.

Bolduc, D., M. G. Dagenais, and M. J. Gaudry (1989), Spatially Autocorrelated Errors in Origin-Destination Models: A New Specification Applied to Urban Travel Demand in Winnipeg, *Transportation Research*, B 23, 361–372.

Bolduc, D., R. Laferrière, and G. Santarossa (1992), Spatial Autoregressive Error Components in Travel Flow Models, *Regional Science and Urban Economics*, 22, 371–385.

Bolduc, D., R. Laferrière, and G. Santarossa (1995), Spatial Autoregressive Error Components in Travel Flow Models, an Application to Aggregate Mode Choice, in L. Anselin and R. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 96–108.

Bowden, R. J. and D. A. Turkington (1984), *Instrumental Variables*, Cambridge University Press, Cambridge.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994), *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ.

Brandsma, A. S. and R. H. Ketellapper (1979a), A Biparametric Approach to Spatial Autocorrelation, *Environment and Planning A*, 11, 51–58.

Brandsma, A. S. and R. H. Ketellapper (1979b), Further Evidence on Alternative Procedures for Testing of Spatial Autocorrelation among Regression Disturbances, in C. Bartels and R. Ketellapper (eds.), *Exploratory and Explanatory Analysis in Spatial Data*, Martin Nijhoff, Boston, 111–136.

Brett, C. and C. A. P. Pinkse (1997), Those Taxes Are All over the Map! A Test for Spatial Independence of Municipal Tax Rates in British Columbia, *International Regional Science Review*, 20, 131–151.

Breusch, T. (1980), useful Invariance Results for Generalized Regression Models, *Journal of Econometrics*, 13, 327–340.

Brook, D. (1964), On the Distinction between the Conditional Probability and Joint Probability Approaches in the Specification of Nearest Neighbor Systems, *Biometrika*, 51, 481–483.

Brueckner, J. K. (1996), Testing for Strategic Interaction among Local Governments: The Case of Growth Controls, Discussion Paper, Department of Economics, University of Illinois, Champaign.

Burridge, P. (1980), On the Cliff-Ord Test for Spatial Autocorrelation, *Journal of the Royal Statistical Society B*, 42, 107–108.

Can, A. (1992), Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics*, 22, 453–474.

Can, A. (1996), Weight Matrices and Spatial Autocorrelation Statistics Using a Topological Vector Data Model, *International Journal of Geographical Information Systems*, 10, 1009–1017.

Can, A. and I. F. Megbolugbe (1997), Spatial Dependence and House Price Index Construction, *Journal of Real Estate Finance and Economics*, 14, 203–222.

Case, A. (1987), On the Use of Spatial Autoregressive Models in Demand Analysis, Discussion Paper 135, Research Program in Development Studies, Woodrow Wilson School, Princeton University.

Case, A. (1991), Spatial Patterns in Household Demand, *Econometrica*, 59, 953–965.

Case, A. (1992). Neighborhood Influence and Technological Change, *Regional Science and Urban Economics*, 22, 491–508.

Case, A. C., H. S. Rosen, and J. R. Hines (1993), Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States, *Journal of Public Economics*, 52, 285–307.

Cliff, A. and J. K. Ord (1972), Testing for Spatial Autocorrelation among Regression Residuals, *Geographic Analysis*, 4, 267–284.

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

Cliff, A. and J. K. Ord (1981), *Spatial Processes: Models and Applications*, Pion, London.

Cressie, N. (1991), Geostatistical Analysis of Spatial Data, in National Research Council, *Spatial Statistics and Digital Image Analysis*, National Academy Press, Washington, DC, 87–108.

Cressie, N (1993), *Statistics for Spatial Data*, Wiley, New York.

Dasgupta, S. and M. D. Perlman (1974), Power of the Noncentral F-test: Effect of Additional Variate on Hotelling's T^2-test, *Journal of the American Statistical Association*, 69, 174–180.

Davidson, R. and J. G. MacKinnon (1987), Implicit Alternatives and Local Power of Test Statistics, *Econometrica*, 55, 1305–1329.

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Deutsch, C. V. and A. G. Journel (1992), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, Oxford.

Doreian, P. (1980), Linear Models with Spatially Distributed Data, Spatial Disturbances or Spatial Effects, *Sociological Methods and Research*, 9, 29–60.

Doreian, P., K. Teuter, and C-H. Wang (1984), Network Autocorrelation Models, *Sociological Methods and Research*, 13, 155–200.

Dow, M. M., M. L. Burton, and D. R. White (1982), Network Autocorrelation: A Simulation Study of a Foundational Problem in Regression and Survey Study Research, *Social Networks*, 4, 169–200.

Dubin, R. (1988), Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms, *Review of Economics and Statistics*, 70, 466–474.

Dubin, R. (1992), Spatial Autocorrelation and Neighborhood Quality, *Regional Science and Urban Economics*, 22, 433–452.

Durbin, J. and G. S. Watson (1950), Testing for Serial Correlation in Least Squares Regression I, *Biometrika*, 37, 409–428.

Durbin, J. and G. S. Watson (1951), Testing for Serial Correlation in Least Squares Regression II, *Biometrika*, 38, 159–179.

Florax, R. and S. Rey (1995), The Impacts of Misspecified Spatial Interaction in Linear Regression Models, in L. Anselin and R. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 111–135.

Fomby, T. B., R. C. Hill, and S. R. Johnson (1984), *Advanced Econometric Methods*, Springer-Verlag, New York.

Getis, A. (1995), Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality, and Government Expenditures, in L. Anselin and R. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 172–185.

Godfrey, L. (1981), On the Invariance of the Lagrange Multiplier Test with Respect to Certain Changes in the Alternative Hypothesis, *Econometrica*, 49, 1443–1455.

Godfrey, L. (1988), *Misspecification Tests in Econometrics*, Cambridge University Press, New York.

Greene, W. H. (1993), *Econometric Analysis*, 2nd ed., Macmillan, New York.

Griffith, D. A. (1987), *Spatial Autocorrelation, A Primer*, Association of American Geographers, Washington, DC.

Griffith, D. A. (1993), *Spatial Regression Analysis on the PC: Spatial Statistics Using SAS*, Association of American Geographers, Washington, DC.

Haining, R. (1984), Testing a Spatial Interacting-Markets Hypothesis, *The Review of Economics and Statistics*, 66, 576–583.

Haining, R. (1988), Estimating Spatial Means with an Application to Remotely Sensed Data, *Communications in Statistics: Theory and Methods*, 17, 573–597.

Haining, R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.

Heijmans, R. and J. Magnus (1986a), On the First-Order Efficiency and Asymptotic Normality of Maximum Likelihood Estimators Obtained from Dependent Observations, *Statistica Neerlandica*, 40, 169–188.

Heijmans, R. and J. Magnus (1986b), Asymptotic Normality of Maximum Likelihood Estimators Obtained from Normally Distributed but Dependent Observations, *Econometric Theory*, 2, 374–412.

Hepple, L. W. (1995a), Bayesian Techniques in Spatial and Network Econometrics. 1: Model Comparison and Posterior Odds, *Environment and Planning A*, 27, 447–469.

Hepple, L. W. (1995b), Bayesian Techniques in Spatial and Network Econometrics. 2: Computational Methods and Algorithms, *Environment and Planning A*, 27, 615–644.

Holtz-Eakin, D. (1994), Public-Sector Capital and the Productivity Puzzle, *Review of Economics and Statistics*, 76, 12–21.

Hooper, P. and G. Hewings (1981), Some Properties of Space-Time Processes, *Geographical Analysis*, 13, 203–223.

Hordijk, L. (1974), Spatial Correlation in the Disturbances of a Linear Interregional Model, *Regional and Urban Economics*, 4, 117–140.

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

Hordijk, L. and J. Paelinck (1976), Some Principles and Results in Spatial Econometrics, *Recherches Economiques de Louvain*, 42, 175–197.

Huang, J. S. (1984). The Autoregressive Moving Average Model for Spatial Analysis, *Australian Journal of Statistics*, 26, 169–178.

Imhof, J. P. (1961), Computing the Distribution of Quadratic Forms in Normal Variables, *Biometrika*, 48, 419–426.

Isaaks, E. H. and R. M. Srivastava (1989), *An Introduction to Applied Geostatistics*, Oxford University Press, Oxford.

Johnson, N. L., and S. Kotz (1969), *Distributions in Statistics: Discrete Distributions*, Houghton Mifflin, Boston.

Johnston, J. (1984), *Econometric Models*, McGraw-Hill, New York.

Judge, G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T-C. Lee (1982), *Introduction to the Theory and Practice of Econometrics*, Wiley, New York.

Judge, G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985), *The Theory and Practice of Econometrics*, 2nd ed., Wiley, New York.

Kelejian, H. and D. Robinson (1992), Spatial Autocorrelation: A New Computationally Simple Test with an Application to Per Capita Country Police Expenditures, *Regional Science and Urban Economics*, 22, 317–331.

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), Spatial Correlation: A Suggested Alternative to the Autoregressive Model, in L. Anselin and R. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 75–95.

King, M. L. (1981), A Small Sample Property of the Cliff-Ord Test for Spatial Correlation, *Journal of the Royal Statistical Society B*, 43, 263–264.

King, M. L. (1987), Testing for Autocorrelation in Linear Regression Models: A Survey, in M. L. King and D. E. A. Giles (eds.), *Specification Analysis in the Linear Model*, Routledge and Kegan Paul, London, 19–73.

King, M. L. and M. A. Evans (1986), Testing for Block Effects in Regression Models Based on Survey Data, *Journal of the American Statistical Association*, 81, 677–679.

King, M. L. and G. H. Hillier (1985), Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model, *Journal of the Royal Statistical Society B*, 47, 98–102.

Koerts, J. and A. P. I. Abrahamse (1968), On the Power of the BLUS Procedure, *Journal of the American Statistical Association*, 63, 1227–1236.

Krugman, P. (1991), Increasing Returns and Economic Geography, *Journal of Political Economy*, 99, 483–499.

Land, K. and G. Deane (1992), On the Large-Sample Estimation of Regression Models with Spatial- or Network-Effects Terms: A Two Stage Least Squares Approach, in P. Marsden (ed.), *Sociological Methodology*, Jossey Bass, San Francisco, 221–248.

Leenders, R. T. (1995), *Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure and Their Interdependence*, Thesis Publishers, Amsterdam.

Legendre, P. (1993), Spatial Autocorrelation: Trouble or New Paradigm, *Ecology*, 74, 1659–1673.

LeSage, J. P. (1993), Spatial Modeling of Agricultural Markets, *American Journal of Agricultural Economics*, 75, 1211–1216.

LeSage, J. P. (1997), Bayesian Estimation of Spatial Autoregressive Models, *International Regional Science Review*, 20, 113–129.

Liang, K. Y. and S. L. Zeger (1986), Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73, 13–22.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Magnus, J. (1978), Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix, *Journal of Econometrics*, 7, 281–312; Corrigenda, *Journal of Econometrics*, 10, 261.

Mardia, K. V. and R. J. Marshall (1984), Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression, *Biometrika*, 71, 135–146.

Mardia, K. V. and A. J. Watkins (1989), On Multimodality of the Likelihood for the Spatial Linear Model, *Biometrika*, 76, 289–295.

MathSoft (1996), *S+Spatialstats User's Manual, Version 1.0*, MathSoft, Inc., Seattle.

McMillen, D. P. (1992), Probit with Spatial Autocorrelation, *Journal of Regional Science*, 32 335–348.

Moran, P. A. P. (1948), The Interpretation of Statistical Maps, *Biometrika*, 35, 255–260.

Moran, P. A. P. (1950a), Notes on Continuous Stochastic Phenomena, *Biometrika*, 37, 17–23.

Moran, P. A. P. (1950b), A Test for the Serial Independence of Residuals, *Biometrika*, 37, 178–181.

Murdoch, J. C., M. Rahmatian, and M. A. Thayer (1993), A Spatially Autoregressive Median Voter Model of Recreation Expenditures, *Public Finance Quarterly*, 21, 334–350.

Murdoch, J. C., T. Sandler, and K. Sargent (1996), A Tale of Two Collectives: Sulfur versus Nitrogen Oxides Emission Reduction in Europe, Working Paper, Department of Economics, Iowa State University, Ames, IA.

National Research Council (1991), *Spatial Statistics and Digital Image Analysis*, National Academy Press, Washington, DC.

Ord, J. K. (1975), Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, 70, 120–126.

Pace, R. K. and R. Barry (1996), Sparse Spatial Autoregressions, *Statistics and Probability Letters*, 2158, 1–7.

Pace, R. K. and R. Barry (1997), Quick Computation of Spatial Autoregressive Estimators, *Geographical Analysis*, 29 (forthcoming).

Paelinck, J. (1982), Operational Spatial Analysis, *Papers, Regional Science Association*, 50, 1–7.

Paelinck, J. and L. Klaassen (1979), Spatial Econometrics, Saxon House, Farnborough.

Pinkse, J. and M. E. Slade (1995), Contracting in Space, an Application of Spatial Statistics to Discrete-Choice Models, Working Paper, Department of Economics, University of British Columbia, Vancouver, BC.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics. A Comparative Approach*, The MIT Press, Cambridge, MA.

Rao, C. R. (1947), Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation, *Proceedings of the Cambridge Philosophical Society*, 44, 50–57.

Ripley, B. D. (1981), *Spatial Statistics*, Wiley, New York.

Ripley, B. D. (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

Romer, P. M. (1986), Increasing Returns and Long-Run Growth, *Journal of Political Economy*, 94, 1002–1037.

Saikkonen, P. (1989), Asymptotic Relative Efficiency of the Classical Test Statistics Under Misspecification, *Journal of Econometrics*, 42, 351–369.

Silvey, S. D. (1959), The Lagrange Multiplier Test, *Annals of Mathematical Statistics*, 30, 389–407.

Tiefelsdorf, M. and B. Boots (1995), The Exact Distribution of Moran's *I*, *Environment and Planning A*, 27, 985–999.

Tobler, W. (1979), Cellular Geography, in S. Gale and G. Olsson (eds.), *Philosophy in Geography*, Reidel, Dordrecht, 379–386.

Upton, G. J. and B. Fingleton (1985), *Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data*, Wiley, New York.

Upton, G. J. and B. Fingleton (1989), *Spatial Data Analysis by Example. Volume 2: Categorical and Directional Data*, Wiley, New York.

Warnes, J. J. and B. D. Ripley (1987), Problems with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes, *Biometrika*, 74, 640–642.

White, H. (1980), A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817–838.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, Orlando.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

White, H. and I. Domowitz (1984), Nonlinear Regression with Dependent Observations, *Econometrica*, 52, 143–161.

Whittle, P. (1954), On Stationary Processes in the Plane, *Biometrika*, 41, 434–449.

Zeger, S. L. and K. Y. Liang (1986), Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, 42, 121–130.

Zeger, S. L., K. Y. Liang, and P. S. Albert (1988), Models for Longitudinal Data: A Generalized Estimating Equations Approach, *Biometrics*, 44, 1049–1060.