A Companion to Theoretical Econometrics
Edited by Badi H. Baltagi

CHAPTER FOURTEEN

Spatial Econometrics

*Luc Anselin**

1 Introduction

Spatial econometrics is a subfield of econometrics that deals with spatial interaction (*spatial autocorrelation*) and spatial structure (*spatial heterogeneity*) in regression models for cross-sectional and panel data (Paelinck and Klaassen, 1979; Anselin, 1988a). Such a focus on location and spatial interaction has recently gained a more central place not only in applied but also in theoretical econometrics. In the past, models that explicitly incorporated "space" (or geography) were primarily found in specialized fields such as regional science, urban, and real estate economics and economic geography (e.g. recent reviews in Anselin, 1992a; Anselin and Florax, 1995a; Anselin and Rey, 1997; Pace *et al.*, 1998). However, more recently, spatial econometric methods have increasingly been applied in a wide range of empirical investigations in more traditional fields of economics as well, including, among others, studies in demand analysis, international economics, labor economics, public economics and local public finance, and agricultural and environmental economics.[1]

This new attention to specifying, estimating, and testing for the presence of spatial interaction in the mainstream of applied and theoretical econometrics can be attributed to two major factors. One is a growing interest within theoretical economics in *models* that move towards an explicit accounting for the interaction of an economic agent with other heterogeneous agents in the system. These new theoretical frameworks of "interacting agents" model strategic interaction, social norms, neighborhood effects, copy-catting, and other peer group effects, and raise interesting questions about how the individual interactions can lead to emergent collective behavior and aggregate patterns. Models used to estimate such phenomena require the specification of how the magnitude of a variable of interest (say crime) at a given location (say a census tract) is determined by the values of the same variable at other locations in the system (such as neighboring census tracts). If such a dependence exists, it is referred to as spatial autocorrelation. A second driver behind the increased interest in spatial econometric techniques is the need to handle spatial *data*. This has been stimulated by the explosive diffusion of geographic information systems (GIS) and the associated availability of geocoded data (i.e. data sets that contain the location of the observational units). There is a growing recognition that standard econometric techniques often fail in the presence of spatial autocorrelation, which is commonplace in geographic (cross-sectional) data sets.[2]

Historically, spatial econometrics originated as an identifiable field in Europe in the early 1970s because of the need to deal with sub-country data in regional econometric models (e.g. Paelinck and Klaassen, 1979). In general terms, spatial econometrics can be characterized as the set of techniques to deal with methodological concerns that follow from the explicit consideration of *spatial effects*, specifically spatial autocorrelation and spatial heterogeneity. This yields four broad areas of interest: (i) the formal *specification* of spatial effects in econometric models; (ii) the *estimation* of models that incorporate spatial effects; (iii) specification *tests* and diagnostics for the presence of spatial effects; and (iv) spatial *prediction* (interpolation). In this brief review chapter, I will focus on the first three concerns, since they fall within the central preoccupation of econometric methodology.

The remainder of the chapter is organized as follows. In Section 2, I outline some foundations and definitions. In Section 3, the specification of spatial regression models is treated, including the incorporation of spatial dependence in panel data models and models with qualitative variables. Section 4 focuses on estimation and Section 5 on specification testing. In Section 6, some practical implementation and software issues are addressed. Concluding remarks are formulated in Section 7.

2 Foundations

2.1 Spatial autocorrelation

In a regression context, spatial effects pertain to two categories of specifications. One deals with spatial dependence, or its weaker expression, *spatial autocorrelation*, and the other with *spatial heterogeneity*.[3] The latter is simply structural instability, either in the form of non-constant error variances in a regression model (heteroskedasticity) or in the form of variable regression coeffcients. Most of the methodological issues related to spatial heterogeneity can be tackled by means of the standard econometric toolbox.[4] Therefore, given the space constraints for this chapter, the main focus of attention in the remainder will be on spatial dependence.

The formal framework used for the statistical analysis of spatial autocorrelation is a so-called spatial stochastic process (also often referred to as a spatial random field), or a collection of random variables y, indexed by location i,

$$\{y_i, i \in D\}, \tag{14.1}$$

where the index set D is either a continuous surface or a finite set of discrete locations. (See Cressie (1993), for technical details.) Since each random variable is "tagged" by a location, spatial autocorrelation can be formally expressed by the moment condition,

$$\mathrm{cov}[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] \neq 0, \quad \text{for } i \neq j \tag{14.2}$$

where i, j refer to individual observations (locations) and y_i (y_j) is the value of a random variable of interest at that location. This covariance becomes meaningful from a spatial perspective when the particular configuration of nonzero i, j pairs has an interpretation in terms of spatial structure, spatial interaction or the spatial arrangement of the observations. For example, this would be the case when one is interested in modeling the extent to which technological innovations in a county spill over into neighboring counties.

The spatial covariance can be modeled in three basic ways. First, one can specify a particular functional form for a spatial stochastic process generating the random variable in (14.1), from which the covariance structure would follow. Second, one can model the covariance structure directly, typically as a function of a small number of parameters (with any given covariance structure corresponding to a class of spatial stochastic processes). Third, one can leave the covariance unspecified and estimate it nonparametrically.[5] I will review each of these approaches in turn.

Spatial stochastic process models

The most often used approach to formally express spatial autocorrelation is through the specification of a functional form for the spatial stochastic process (14.1) that relates the value of a random variable at a given location to its value at other locations. The covariance structure then follows from the nature of the process. In parallel to time series analysis, spatial stochastic processes are categorized as spatial autoregressive (SAR) and spatial moving average (SMA) processes, although there are several important differences between the cross-sectional and time series contexts.[6]

For example, for an $N \times 1$ vector of random variables, y, observed across space, and an $N \times 1$ vector of iid random errors ε, a simultaneous spatial autoregressive (SAR) process is defined as

$$(\mathrm{y} - \mu\mathrm{i}) = \rho\mathrm{W}(\mathrm{y} - \mu\mathrm{i}) + \varepsilon, \quad \text{or} \quad (\mathrm{y} - \mu\mathrm{i}) = (\mathrm{I} - \rho\mathrm{W})^{-1}\varepsilon, \tag{14.3}$$

where μ is the (constant) mean of y_i, i is an $N \times 1$ vector of ones, and ρ is the spatial autoregressive parameter.

Before considering the structure of this process more closely, note the presence of the $N \times N$ matrix W, which is referred to as a *spatial weights matrix*. For each location in the system, it specifies which of the other locations in the system affect the value at that location. This is necessary, since in contrast to the unambiguous notion of a "shift" along the time axis (such as y_{t-1} in an autoregressive model), there is no corresponding concept in the spatial domain, especially when observations are located irregularly in space.[7] Instead of the notion of shift, a *spatial lag operator* is used, which is a weighted average of random variables at "neighboring" locations.[8]

The spatial weights crucially depend on the definition of a neighborhood set for each observation. This is obtained by selecting for each location i (as the row) the neighbors as the columns corresponding to nonzero elements w_{ij} in a fixed (nonstochastic) and positive $N \times N$ spatial weights matrix W.[9] A spatial lag for y at i then follows as

$$[\mathrm{Wy}]_i = \sum_{j=1,\ldots,N} w_{ij} \cdot y_j, \tag{14.4}$$

or, in matrix form, as

$$\mathrm{Wy}. \tag{14.5}$$

Since for each i the matrix elements w_{ij} are only nonzero for those $j \in S_i$ (where S_i is the neighborhood set), only the matching y_j are included in the lag. For ease of interpretation, the elements of the spatial weights matrix are typically row-standardized, such that for each i, $\Sigma_j w_{ij} = 1$. Consequently, the spatial lag may be interpreted as a weighted average (with the w_{ij} being the weights) of the neighbors, or as a spatial smoother.

It is important to note that the elements of the weights matrix are nonstochastic and exogenous to the model. Typically, they are based on the geographic arrangement of the observations, or contiguity. Weights are nonzero when two locations share a common boundary, or are within a given distance of each other. However, this notion is perfectly general and alternative specifications of the spatial weights (such as economic distance) can be considered as well (Anselin, 1980, ch. 8; Case, Rosen, and Hines, 1993; Pinkse and Slade, 1998).

The constraints imposed by the weights structure (the zeros in each row), together with the specific form of the spatial process (autoregressive or moving average) determine the variance–covariance matrix for y as a function of two parameters, the variance σ^2 and the spatial coefficient, ρ. For the SAR structure in (14.3), this yields (since $E[\mathrm{y} - \mu \mathrm{i}] = 0$)

$$\mathrm{cov}[(\mathrm{y} - \mu\mathrm{i}), (\mathrm{y} - \mu\mathrm{i})] = E[(\mathrm{y} - \mu\mathrm{i})(\mathrm{y} - \mu\mathrm{i})'] = \sigma^2[(\mathrm{I} - \rho\mathrm{W})'(\mathrm{I} - \rho\mathrm{W})]^{-1}. \tag{14.6}$$

This is a full matrix, which implies that shocks at any location affect all other locations, through a so-called *spatial multiplier* effect (or, global interaction).[10]

A major distinction between processes in space compared to the time domain is that even with iid error terms ε_i, the diagonal elements in (14.6) are not constant.[11] Furthermore, the heteroskedasticity depends on the neighborhood structure embedded in the spatial weights matrix W. Consequently, the process in y is not covariance-stationary. Stationarity is only obtained in very rare cases, for example on regular lattice structures when each observation has an identical weights structure, but this is of limited practical use. This lack of stationarity has important implications for the types of central limit theorems (CLTs) and laws of large numbers (LLNs) that need to be invoked to obtain asymptotic properties for estimators and specification test, a point that has not always been recognized in the literature.

Direct representation

A second commonly used approach to the formal specification of spatial autocorrelation is to express the elements of the variance–covariance matrix in a parsimonious fashion as a "direct" function of a small number of parameters and one or more exogenous variables. Typically, this involves an inverse function of some distance metric, for example,

$$\text{cov}[\varepsilon_i, \varepsilon_j] = \sigma^2 f(d_{ij}, \varphi), \tag{14.7}$$

where ε_i and ε_j are regression disturbance terms, σ^2 is the error variance, d_{ij} is the distance separating observations (locations) i and j, and f is a distance decay function such that $\frac{\partial f}{\partial d} < 0$ and $|f(d_{ij}, \varphi)| \leq 1$, with $\varphi \in \Phi$ as a $p \times 1$ vector of parameters on an open subset Φ of R^p. This form is closely related to the variogram model used in geostatistics, although with stricter assumptions regarding stationarity and isotropy. Using (14.7) for individual elements, the full error covariance matrix follows as

$$E[\varepsilon\varepsilon'] = \sigma^2\Omega(d_{ij}, \varphi), \tag{14.8}$$

where, because of the scaling factor σ^2, the matrix $\Omega(d_{ij}, \varphi)$ must be a positive definite spatial correlation matrix, with $\omega_{ii} = 1$ and $|\omega_{ij}| \leq 1$, $\forall\ i, j$.[12] Note that, in contrast to the variance for the spatial autoregressive model, the direct representation model does not induce heteroskedasticity.

In spatial econometrics, models of this type have been used primarily in the analysis of urban housing markets, e.g. in Dubin (1988, 1992), and Basu and Thibodeau (1998). While this specification has a certain intuition, in the sense that it incorporates an explicit notion of spatial clustering as a function of the distance separating two observations (i.e. positive spatial correlation), it is also fraught with a number of estimation and identification problems (Anselin, 2000a).

Nonparametric approaches

A nonparametric approach to estimating the spatial covariance matrix does not require an explicit spatial process or functional form for the distance decay. This is common in the case of panel data, when the time dimension is (considerably) greater than the cross-sectional dimension ($T >> N$) and the "spatial" covariance is estimated from the sample covariance for the residuals of each set of location pairs (e.g. in applications of Zellner's SUR estimator; see Chapter 5 by Fiebig in this volume).

Applications of this principle to spatial autocorrelation are variants of the well known Newey–West (1987) heteroskedasticity and autocorrelation consistent covariance matrix and have been used in the context of generalized methods of moments (GMM) estimators of spatial regression models (see Section 4.3). Conley (1996) suggested a covariance estimator based on a sequence of weighted averages of sample autocovariances computed for subsets of observation pairs that fall within a given distance band (or spatial window). Although not presented as such, this has a striking similarity to the nonparametric estimation of a semivariogram in geostatistics (see, e.g. Cressie, 1993, pp. 69–70), but the assumptions of stationarity and isotropy required in the GMM approach are stricter than those needed in variogram estimation. In a panel data setting, Driscoll and Kraay (1998) use a similar idea, but avoid having to estimate the spatial covariances by distance bands. This is accomplished by using only the cross-sectional averages (for each time period) of the moment conditions, and by relying on asymptotics in the time dimension to yield an estimator for the spatial covariance structure.

2.2 Aymptotics in spatial stochastic processes

As in time series analysis, the properties of estimators and tests for spatial series are derived from the asymptotics for stochastic processes. However, these properties are not simply extensions to two dimensions of the time series results. A number of complicating factors are present and to date some formal results for the spatial dependence case are still lacking. While an extensive treatment of this topic is beyond the scope of the current chapter, three general comments are in order. First, the intuition behind the asymptotics is fairly straightforward in that regularity conditions are needed to limit the extent of spatial dependence (memory) and heterogeneity of the spatial series in order to obtain the proper (uniform) laws of large numbers and central limit theorems to establish consistency and asymptotic normality. In this context, it is important to keep in mind that both SAR and SMA processes yield heteroskedastic variances, so that the application of results for dependent *stationary* series are not applicable.[13] In addition to the usual moment conditions that are similar in spirit to those for heterogeneous dependent processes in time (e.g. Pötscher and Prucha, 1997), specific spatial conditions will translate into constraints on the spatial weights and on the parameter space for the spatial coefficients (for some specific examples, see, e.g. Anselin and Kelejian, 1997; Kelejian and Prucha, 1999b; Pinkse and Slade, 1998; Pinkse, 2000). In practice, these conditions are likely to be satisfied by most spatial weights that are based on simple contiguity, but this is not necessarily the case for general weights, such as those based on economic distance.

A second distinguishing characteristic of asymptotics in space is that the limit may be approached in two different ways, referred to as *increasing domain* asymptotics and *infill* asymptotics.[14] The former consists of a sampling structure where new "observations" are added at the edges (boundary points), similar to the underlying asymptotics in time series analysis. Infill asymptotics are appropriate when the spatial domain is bounded, and new observations are added in between existing ones, generating a increasingly denser surface. Many results for increasing domain asymptotics are not directly applicable to infill asymptotics (Lahiri, 1996). In most applications of spatial econometrics, the implied structure is that of an increasing domain.

Finally, for spatial processes that contain spatial weights, the asymptotics require the use of CLT and LLN for triangular arrays (Davidson, 1994, chs. 19, 24). This is caused by the fact that for the boundary elements the "sample" weights matrix changes as new data points are added (i.e. the new data points change the connectedness structure for existing data points).[15] Again, this is an additional degree of complexity, which is not found in time series models.

3 Spatial Regression Models

3.1 Spatial lag and spatial error models

In the standard linear regression model, spatial dependence can be incorporated in two distinct ways: as an additional regressor in the form of a spatially lagged dependent variable (Wy), or in the error structure ($E[\varepsilon_i \varepsilon_j] \neq 0$). The former is referred to as a *spatial lag* model and is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction. This is interpreted as substantive spatial dependence in the sense of being directly related to a *spatial model* (e.g. a model that incorporates spatial interaction, yardstick competition, etc.). Spatial dependence in the regression disturbance term, or a *spatial error* model is referred to as nuisance dependence. This is appropriate when the concern is with correcting for the potentially biasing influence of the spatial autocorrelation, due to the use of *spatial data* (irrespective of whether the model of interest is spatial or not).

Formally, a spatial lag model, or a mixed regressive, spatial autoregressive model is expressed as

$$y = \rho W y + X\beta + \varepsilon, \tag{14.9}$$

where ρ is a spatial autoregressive coefficient, ε is a vector of error terms, and the other notation is as before.[16] Unlike what holds for the time series counterpart of this model, the spatial lag term Wy is correlated with the disturbances, even when the latter are iid. This can be seen from the reduced form of (14.9),

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon, \tag{14.10}$$

in which each inverse can be expanded into an infinite series, including both the explanatory variables and the error terms at all locations (the spatial multiplier). Consequently, the spatial lag term must be treated as an endogenous variable and proper estimation methods must account for this endogeneity (OLS will be biased and inconsistent due to the simultaneity bias).

A spatial error model is a special case of a regression with a non-spherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence. Consequently, OLS remains unbiased, but it is no longer efficient and the classical estimators for standard errors will be biased. The spatial structure can be specified in a number of different ways, and (except for the non-parametric approaches) results in a error variance–covariance matrix of the form

$$E[\varepsilon\varepsilon'] = \Omega(\theta), \tag{14.11}$$

where θ is a vector of parameters, such as the coefficients in an SAR error process.[17]

3.2 Spatial dependence in panel data models

When observations are available across space as well as over time, the additional dimension allows the estimation of the full covariance of one type of association, using the other dimension to provide the asymptotics (e.g. in SUR models with $N << T$). However, as in the pure cross-sectional case, there is insufficient information in the NT observations to estimate the complete $(NT)^2$ covariance matrix $\text{cov}[y_{it}, y_{js}] \neq 0$, (with $i \neq j$ and $t \neq s$) without imposing some structure. For small N and large T, the asymptotics in the time domain can be exploited to obtain a nonparametric estimate of cross-sectional dependence, while time dependence must be parameterized. Similarly, for large N and small T, the asymptotics in the spatial domain can be exploited to yield a nonparametric estimate of serial (time) dependence, while spatial dependence must be parameterized. As in the pure cross-sectional case, the latter requires the use of a spatial weights matrix. In each of these situations, asymptotics are only needed in one of the dimensions while the other can be treated as fixed.

When both spatial as well as serial dependence are parameterized, a range of specifications can be considered, allowing different combinations of the two. For ease of exposition, assume that the observations are stacked by time period, i.e. they can be considered as T time slices of N cross-sectional units. Restricting attention to "lag" dependence, and with $f(z)$ as a generic designation for the regressors (which may be lagged in time and/or space), four types of models can be distinguished.

1. *pure space-recursive*, in which the dependence pertains to neighboring locations in a different period, or,

$$y_{it} = \gamma[\mathrm{W}\mathrm{y}_{t-1}]_i + f(z) + \varepsilon_{it}, \tag{14.12}$$

   where, using the same notational convention as before, $[\mathrm{W}\mathrm{y}_{t-1}]_i$ is the ith element of the spatial lag vector applied to the observations on the dependent variable in the previous time period (using an $N \times N$ spatial weights matrix for the cross-sectional units).
2. *time–space recursive*, in which the dependence relates to the same location as well as the neighboring locations in another period, or,

$$y_{it} = \lambda y_{it-1} + \gamma[\mathrm{W}\mathrm{y}_{t-1}]_i + f(z) + \varepsilon_{it} \tag{14.13}$$

3. *time–space simultaneous*, with both a time-wise and a spatially lagged dependent variable, or,

$$y_{it} = \lambda y_{it-1} + \rho[\mathrm{W}\mathrm{y}_{t}]_i + f(z) + \varepsilon_{it} \tag{14.14}$$

   where $[\mathrm{W}\mathrm{y}_t]_i$ is the ith element of the spatial lag vector in the same time period.

4. *time–space dynamic*, with all forms of dependence, or,

$$y_{it} = \lambda y_{it-1} + \rho[Wy_t]_i + \gamma[Wy_{t-1}]_i + f(z) + \varepsilon_{it}. \quad (14.15)$$

In order to estimate the parameters of the time–space simultaneous model, asymptotics are needed in the cross-sectional dimension, while for the time–space dynamic model, asymptotics are needed in both dimensions. For the other models, the type of asymptotics required are determined by the dependence structure in the error terms. For example, the pure space-recursive model with iid errors satisfies the assumptions of the classical linear model and can be estimated by means of OLS.

Spatial lag and spatial error dependence can be introduced into the cross-sectional dimension of traditional panel data models in a straightforward way. For example, in a spatial SUR model, both autoregressive as well as regression parameters are allowed to vary by time period, in combination with a nonparametric serial covariance. The spatial lag formulation of such a model would be (in the same notation as before):

$$y_{it} = \rho_t[Wy_t]_i + x'_{it}\beta_t + \varepsilon_{it} \quad (14.16)$$

with $\mathrm{var}[\varepsilon_{it}] = \sigma_t^2$ and $E[\varepsilon_{it}\varepsilon_{is}] = \sigma_{ts}$.[18]

An important issue to consider when incorporating spatial dependence in panel data models is the extent to which fixed effects may be allowed. Since the estimation of the spatial process models requires asymptotics in the cross-sectional domain ($N \to \infty$), fixed effects (i.e. a dummy variable for each location) would suffer from the incidental parameter problem and no consistent estimator exists. Hence, fixed cross-sectional effects are incompatible with spatial processes and instead a random effects specification must be considered.

3.3 Spatial dependence in models for qualitative data

Empirical analysis of interacting agents requires models that incorporate spatial dependence for discrete dependent variables, such as counts or binary outcomes (Brock and Durlauf, 1995). This turns out to be quite complex and continues to be an active area of research. While an extensive discussion of the technical aspects associated with spatial discrete choice models is beyond the scope of the current chapter, the salient issues may be illustrated with a spatial version of the probit model, which has recently received considerable attention.[19]

The point of departure is the familiar expression for a linear model in a latent (unobserved) dependent variable y_i^*

$$y_i^* = x_i'\beta + \varepsilon_i, \quad (14.17)$$

where ε_i is a random variable for which a given distribution is assumed (e.g. the normal for the probit model). The realization of y_i^* is observed in the form of discrete events, $y_i = 1$ for $y_i^* \geq 0$, and $y_i = 0$ for $y_i^* < 0$. The discrete events are related to the underlying probability model through the error term, for example, $y_i^* \geq 0$ implies $-x_i'\beta < \varepsilon_i$, and, therefore,

$$E[y_i] = P[y_i = 1] = \Phi[x_i'\beta], \tag{14.18}$$

where Φ is the cumulative distribution function for the standard normal.

Spatial autocorrelation can be introduced into this model in the form of a spatial autoregressive process for the error term ε_i in (14.17), or

$$\varepsilon_i = \lambda \sum_j w_{ij}\varepsilon_j + u_i, \tag{14.19}$$

where λ is an autoregressive parameter, the w_{ij} are the elements in the ith row of a spatial weights matrix, and u_i may be assumed to be iid standard normal. As a consequence of the spatial multiplier in the autoregressive specification, the random error at each location now becomes a function of the random errors at all other locations as well. Its distribution is multivariate normal with $N \times N$ variance–covariance matrix

$$E[\varepsilon\varepsilon'] = [(I - \lambda W)'(I - \lambda W)]^{-1}. \tag{14.20}$$

As pointed out above, besides being nondiagonal, (14.20) is also heteroskedastic. Consequently, the usual inequality conditions that are at the basis of (14.18) no longer hold, since each location has a different variance. Moreover, $P[-x_i'\beta < \varepsilon_i]$ can no longer be derived from the univariate standard normal distribution, but rather must be expressed explicitly as the marginal distribution of a N-dimensional multivariate normal vector, whose variance–covariance matrix contains off-diagonal elements that are a function of the autoregressive parameter λ. This is non-standard and typically not analytically tractable, which greatly complicates estimation and specification testing. Similar issues are faced in the spatial lag model for a latent variable.[20]

4 ESTIMATION

4.1 Maximum likelihood estimation

Maximum likelihood (ML) estimation of spatial lag and spatial error regression models was first outlined by Ord (1975).[21] The point of departure is an assumption of normality for the error terms. The joint likelihood then follows from the multivariate normal distribution for y. Unlike what holds for the classic regression model, the joint loglikelihood for a spatial regression does not equal the sum of the loglikelihoods associated with the individual observations. This is due to the two-directional nature of the spatial dependence, which results in a Jacobian term that is the determinant of a full $N \times N$ matrix, e.g. $|I - \rho W|$.

For the SAR error model, the loglikelihood is based on the multivariate normal case, for example, as used in the general treatment of Magnus (1978). Since $\varepsilon \sim MVN(0, \Sigma)$, it follows that, with $\varepsilon = y - X\beta$ and $\Sigma = \sigma^2[(I - \lambda W)'(I - \lambda W)]^{-1}$,

$$\ln L = -(N/2)\ln(2\pi) - (N/2)\ln\sigma^2 + \ln|I - \lambda W| \\ -(1/2\sigma^2)(y - X\beta)'(I - \lambda W)'(I - \lambda W)(y - X\beta). \quad (14.21)$$

Closer inspection of the last term in (14.21) reveals that, conditional upon λ (the spatial autoregressive parameter), a maximization of the loglikelihood is equivalent to the minimization of the sum of squared residuals in a regression of a spatially filtered dependent variable $y^* = y - \lambda Wy$ on a set of spatially filtered explanatory variables $X^* = X - \lambda WX$. The first order conditions for $\hat{\beta}_{ML}$ indeed yield the familiar generalized least squares estimator:

$$\hat{\beta}_{ML} = [(X - \lambda WX)'(X - \lambda WX)]^{-1}(X - \lambda WX)'(y - \lambda Wy) \quad (14.22)$$

and, similarly, the ML estimator for σ^2 follows as:

$$\hat{\sigma}^2_{ML} = (e - \lambda We)'(e - \lambda We)/N \quad (14.23)$$

with $e = y - X\hat{\beta}_{ML}$. However, unlike the time series case, a consistent estimator for λ cannot be obtained from the OLS residuals and therefore the standard two-step FGLS approach does not apply.[22] Instead, the estimator for λ must be obtained from an explicit maximization of a concentrated likelihood function (for details, see Anselin, 1988a, ch. 6, and Anselin and Bera, 1998).

The loglikelihood for the spatial lag model is obtained using the same general principles (see Anselin, 1988, ch. 6 for details) and takes the form

$$\ln L = -(N/2)\ln(2\pi) - (N/2)\ln\sigma^2 + \ln|I - \rho W| \\ -(1/2\sigma^2)(y - \rho Wy - X\beta)'(y - \rho Wy - X\beta). \quad (14.24)$$

The minimization of the last term in (14.24) corresponds to OLS, but since this ignores the log Jacobian $\ln|I - \rho W|$, OLS is not a consistent estimator in this model. As in the spatial error model, there is no satisfactory two-step procedure and estimators for the parameters must be obtained from an explicit maximization of the likelihood. This is greatly simplified since both $\hat{\beta}_{ML}$ and $\hat{\sigma}^2_{ML}$ can be obtained conditional upon ρ from the first order conditions:

$$\hat{\beta}_{ML} = (X'X)^{-1}X'(y - \rho Wy), \quad (14.25)$$

or, with $\hat{\beta}_0 = (X'X)^{-1}X'y$, $e_0 = y - X\hat{\beta}_0$, $\hat{\beta}_L = (X'X)^{-1}X'Wy$, $e_L = y - X\hat{\beta}_L$,

$$\hat{\beta}_{ML} = \hat{\beta}_0 - \rho\hat{\beta}_L \quad (14.26)$$

and

$$\hat{\sigma}^2_{ML} = (e_0 - \rho e_L)'(e_0 - \rho e_L)/N. \quad (14.27)$$

This yields a concentrated loglikelihood in a single parameter, which is straightforward to optimize by means of direct search techniques (see Anselin (1980, 1988a) for derivations and details).

Both spatial lag and spatial error models are special cases of a more general specification that may include forms of heteroskedasticity as well. This also provides the basis for ML estimation of spatial SUR models with spatial lag or spatial error terms (Anselin, 1980, ch. 10). Similarly, ML estimation of error components models with spatial lag or spatial error terms can be implemented as well. Spatial models with discrete dependent variables are typically not estimated by means of ML, given the prohibitive nature of evaluating multiple integrals to determine the relevant marginal distributions.[23]

Finally, it is important to note that models with spatial dependence do not fit the classical framework (e.g. as outlined in Rao, 1973) under which the optimal properties (consistency, asymptotic efficiency, asymptotic normality) of ML estimators are established. This implies that these properties do not necessarily hold and that careful consideration must be given to the explicit formulation of regularity conditions. In general terms, aside from the usual restrictions on the variance and higher moments of the model variables, these conditions boil down to constraints on the range of dependence embodied in the spatial weights matrix.[24] In addition, to avoid singularity or explosive processes, the parameter space for the coefficient in a spatial process model is restricted to an interval other than the familiar –1, +1. For example, for an SAR process, the parameter space is $1/\omega_{min} < \rho < 1/\omega_{max}$, where ω_{min} and ω_{max} are the smallest (on the real line) and largest eigenvalues of the spatial weights matrix W. For row-standardized weights, $\omega_{max} = 1$, but $\omega_{min} > -1$, such that the lower bound on the parameter space is less than –1 (Anselin, 1980). This must be taken into account in practical implementations of estimation routines.

4.2 Spatial two-stage least squares

The endogeneity of the spatially lagged dependent variable can also be addressed by means of an instrumental variables or two-stage least squares (2SLS) approach (Anselin, 1980, 1988a, 1990; Kelejian and Robinson, 1993; Kelejian and Prucha, 1998). As demonstrated in Kelejian and Robinson (1993), the choice of an instrument for Wy follows from the conditional expectation in the reduced form (14.10),

$$E[y \mid X] = (I - \rho W)^{-1} X\beta = X\beta + \rho W X\beta + \rho^2 W^2 X\beta + \dots . \qquad (14.28)$$

Apart from the exogenous variables X (which are always instruments), this includes their spatial lags as well, suggesting WX as a set of instruments.

Under a set of reasonable assumptions that are easily satisfied when the spatial weights are based on contiguity, the spatial two-stage least squares estimator achieves the consistency and asymptotic normality properties of the standard 2SLS (see, e.g. the theorems spelled out in Schmidt, 1976).[25] A straightforward extension is the application of 3SLS to the spatial SUR model with a spatial lag (Anselin, 1988a, ch. 10).

4.3 Method of moments estimators

Recently, a number of approaches have been outlined to estimate the coefficients in a spatial error model as an application of general principles underlying the method of moments. Kelejian and Prucha (1999a) develop a set of moment conditions that yield estimation equations for the parameter of an SAR error model. Specifically, assuming an iid error vector u, the following three conditions readily follow

$$E[u'u/N] = \sigma^2$$

$$E[u'W'Wu/N] = \sigma^2(1/N)\mathrm{tr}(W'W) \tag{14.29}$$

$$E[u'Wu/N] = 0$$

where tr is the matrix trace operator. Replacing u by e – λWe (with e as the vector of OLS residuals) in (14.29) yields a system of three equations in the parameters λ, λ^2, and σ^2. Kelejian and Prucha (1999a) suggest the use of nonlinear least squares to obtain a consistent generalized moment estimator for λ from this system, which can then be used to obtain consistent estimators for the β in an FGLS approach. Since the λ is considered as a nuisance parameter, its significance (as a test for spatial autocorrelation) cannot be assessed, but its role is to provide a consistent estimator for the regression coefficients.[26]

A different approach is taken in the application of Hansen's (1982) generalized method of moments estimator (GMM) to spatial error autocorrelation in Conley (1996). This estimator is the standard minimizer of a quadratic form in the sample moment conditions, where the covariance matrix is obtained in nonparametric form as an application of the ideas of Newey and West (1987). Specifically, the spatial covariances are estimated from weighted averages of sample covariances for pairs of observations that are within a given distance band from each other. Note that this approach requires covariance stationarity, which is only satisfied for a restricted set of spatial processes (e.g. it does not apply to SAR error models).

Pinkse and Slade (1998) use a set of moment conditions to estimate a probit model with SAR errors. However, they focus on the induced heteroskedasticity of the process and do not explicitly deal with the spatial covariance structure.[27]

The relative efficiency of the new methods of moments approaches relative to the more traditional maximum likelihood techniques remains an area of active investigation.

4.4 Other estimation methods

A number of other approaches have been suggested to deal with the estimation of spatial regression models. An early technique is the so-called coding method, originally examined in Besag and Moran (1975).[28] This approach consists of selecting a subsample from the data such that the relevant neighbors are removed (a non-contiguous subsample). This in effect eliminates the simultaneity bias in the spatial lag model, but at the cost of converting the model to a conditional one and with a considerable reduction of the sample size (down to 20 percent of the original sample for irregular lattice data). The advantage of this approach is that standard methods may be applied (e.g. for discrete choice models). However, it is not an efficient procedure and considerable arbitrariness is involved in the selection of the coding scheme.

Another increasingly common approach consists of the application of computational estimators to spatial models. A recent example is the recursive importance sampling (RIS) estimator (Vijverberg, 1997) applied to the spatial probit model in Beron and Vijverberg (2000).

A considerable literature also exists on Bayesian estimation of spatial models, but a detailed treatment of this is beyond the current scope.

5 SPECIFICATION TESTS

5.1 Moran's I

The most commonly used specification test for spatial autocorrelation is derived from a statistic developed by Moran (1948) as the two-dimensional analog of a test for univariate time series correlation (see also Cliff and Ord, 1973). In matrix notation, Moran's I statistic is

$$I = [N/S_0)(\mathrm{e}'\mathrm{We}/\mathrm{e}'\mathrm{e}), \tag{14.30}$$

with e as a vector of OLS residuals and $S_0 = \Sigma_i \Sigma_j w_{ij}$, a standardization factor that corresponds to the sum of the weights for the nonzero cross-products. The statistic shows a striking similarity to the familiar Durbin–Watson test.[29]

Moran's I test has been shown to be locally best invariant (King, 1981) and consistently outperforms other tests in terms of power in simulation experiments (for a recent review, see Anselin and Florax, 1995b). Its application has been extended to residuals in 2SLS regression in Anselin and Kelejian (1997), and to generalized residuals in probit models in Pinkse (2000). General formal conditions and proofs for the asymptotic normality of Moran's I in a wide range of regression models are given in Pinkse (1998) and Kelejian and Prucha (1999b). The consideration of Moran's I in conjunction with spatial heteroskedasticity is covered in Kelejian and Robinson (1998, 2000).

5.2 ML based tests

When spatial regression models are estimated by maximum likelihood, inference on the spatial autoregressive coefficients may be based on a Wald or asymptotic t-test (from the asymptotic variance matrix) or on a likelihood ratio test (see Anselin, 1988a, ch. 6; Anselin and Bera, 1998). Both approaches require that the alternative model (i.e. the spatial model) be estimated. In contrast, a series of test statistics based on the Lagrange Multiplier (LM) or Rao Score (RS) principle only require estimation of the model under the null. The LM/RS tests also allow for the distinction between a spatial error and a spatial lag alternative.[30]

An LM/RS test against a spatial error alternative was originally suggested by Burridge (1980) and takes the form

$$LM_{err} = [e'We/(e'e/N)]^2/[tr(W^2 + W'W)]. \tag{14.31}$$

This statistic has an asymptotic $\chi^2(1)$ distribution and, apart from a scaling factor, corresponds to the square of Moran's *I*.[31] From several simulation experiments (Anselin and Rey, 1991; Anselin and Florax, 1995b) it follows that Moran's *I* has slightly better power than the LM_{err} test in small samples, but the performance of both tests becomes indistinguishable in medium and large size samples. The LM/RS test against a spatial lag alternative was outlined in Anselin (1988c) and takes the form

$$LM_{lag} = [e'Wy/(e'e/N)]^2/D, \tag{14.32}$$

where $D = [(WX\beta)'(I - X(X'X)^{-1}X')(WX\beta)/\sigma^2] + tr(W^2 + W'W)$. This statistic also has an asymptotic $\chi^2(1)$ distribution.

Since both tests have power against the other alternative, it is important to take account of possible lag dependence when testing for error dependence and vice versa. This can be implemented by means of a joint test (Anselin, 1988c) or by constructing tests that are robust to the presence of local misspecification of the other form (Anselin *et al.*, 1996).

The LM/RS principle can also be extended to more complex spatial alternatives, such as higher order processes, spatial error components and direct representation models (Anselin, 2000), to panel data settings (Anselin, 1988b), and to probit models (Pinkse, 1998, 2000; Pinkse and Slade, 1998). A common characteristic of the LM/RS tests against spatial alternatives is that they do not lend themselves readily to a formulation as an NR^2 expression based on an auxiliary regression. However, as recently shown in Baltagi and Li (2000a), it is possible to obtain tests for spatial lag and spatial error dependence in a linear regression model by means of Davidson and MacKinnon's (1988) double length artificial regression approach.

6 IMPLEMENTATION ISSUES

To date, spatial econometric methods are not found in the main commercial econometric and statistical software packages, although macro and scripting facilities may be used to implement some estimators (Anselin and Hudak, 1992). The only comprehensive software to handle both estimation and specification testing of spatial regression models is the special-purpose SpaceStat package (Anselin, 1992b, 1998). A narrower set of techniques, such as maximum likelihood estimation of spatial models is included in the Matlab routines of Pace and Barry (1998), and estimation of spatial error models is part of the S+Spatialstats add-on to S-Plus (MathSoft, 1996).[32]

In contrast to maximum likelihood estimation, method of moments and 2SLS can easily be implemented with standard software, provided that spatial lags can be computed. This requires the construction of a spatial weights matrix, which must often be derived from information in a geographic information system. Similarly, once a spatial lag can be computed, the LM/RS statistics are straightforward to implement.

The main practical problem is encountered in maximum likelihood estimation where the Jacobian determinant must be evaluated for every iteration in a nonlinear optimization procedure. The original solution to this problem was suggested by Ord (1975), who showed how the log Jacobian can be decomposed in terms that contain the eigenvalues of the weights matrix ω_i,

$$\ln|I - \rho W| = \sum_{i=1}^{n} \ln(1 - \rho\omega_i). \tag{14.33}$$

This is easy to implement in a standard optimization routine by treating the individual elements in the sum as observations on an auxiliary term in the log-likelihood (see Anselin and Hudak, 1992). However, the computation of the eigenvalues quickly becomes numerically unstable for matrices of more than 1,000 observations. In addition, for large data sets this approach is inefficient in that it does not exploit the high degree of sparsity of the spatial weights matrix. Recently suggested solutions to this problem fall into two categories. Approximate solutions avoid the computation of the Jacobian determinant, but instead approximate it by a polynomial function or by means of simulation methods (e.g. Barry and Pace, 1999). Exact solutions are based on Cholesky or LU decomposition methods that exploit the sparsity of the weights (Pace and Barry, 1997a, 1997b), or use a characteristic polynomial approach (Smirnov and Anselin, 2000). While much progress has been made, considerable work remains to be done to develop efficient algorithms and data structures to allow for the analysis of very large spatial data sets.

7 CONCLUDING REMARKS

This review chapter has been an attempt to present the salient issues pertaining to the methodology of spatial econometrics. It is by no means complete, but it is hoped that sufficient guidance is provided to pursue interesting research directions. Many challenging problems remain, both methodological in nature as well as in terms of applying the new techniques to meaningful empirical problems. Particularly in dealing with spatial effects in models other than the standard linear regression, much needs to be done to complete the spatial econometric toolbox. It is hoped that the review presented here will stimulate statisticians and econometricians to tackle these interesting and challenging problems.

Notes

* This paper benefited greatly from comments by Wim Vijverberg and two anonymous referees. A more comprehensive version of this paper is available as Anselin (1999).

1 A more extensive review is given in Anselin and Bera (1998) and Anselin (1999).

2 An extensive collection of recent applications of spatial econometric methods in economics can be found in Anselin and Florax (2000).

3 In this chapter, I will use the terms spatial dependence and spatial autocorrelation interchangeably. Obviously, the two are not identical, but typically, the weaker form is used, in the sense of a moment of a joint distribution. Only seldom is the focus on the complete joint density (a recent exception can be found in Brett and Pinkse (1997)).

4 See Anselin (1988a), for a more extensive discussion.

5 One would still need to establish the class of spatial stochastic processes that would allow for the consistent estimation of the covariance; see Frees (1995) for a discussion of the general principles.

6 See Anselin and Bera (1998) for an extensive and technical discussion.

7 On a square grid, one could envisage using North, South, East and West as spatial shifts, but in general, for irregular spatial units such as counties, this is impractical, since the number of neighbors for each county is not constant.

8 In Anselin (1988a), the term spatial lag is introduced to refer to this new variable, to emphasize the similarity to a distributed lag term rather than a spatial shift.

9 By convention, $w_{ii} = 0$, i.e. a location is never a neighbor of itself. This is arbitrary, but can be assumed without loss of generality. For a more extensive discussion of spatial weights, see Anselin (1988a, ch. 3), Cliff and Ord (1981), Upton and Fingleton (1985).

10 See Anselin and Bera (1998) for further details.

11 See McMillen (1992) for an illustration.

12 The specification of spatial covariance functions is not arbitrary, and a number of conditions must be satisfied in order for the model to be "valid" (details are given in Cressie (1993, pp. 61–3, 67–8 and 84–6)).

13 Specifically, this may limit the applicability of GMM estimators that are based on a central limit theorem for stationary mixing random fields such as the one by Bolthausen (1982), used by Conley (1996).

14 Cressie (1993, pp. 100–1).

15 See Kelejian and Prucha (1999a, 1999b).

16 For ease of exposition, the error term is assumed to be iid, although various forms of heteroskedasticity can be incorporated in a straightforward way (Anselin, 1988a, ch. 6).

17 Details and a review of alternative specifications are given in Anselin and Bera (1998).

18 For further details, see Anselin (1988a, 1988b). A recent application is Baltagi and Li (2000b).

19 Methodological issues associated with spatial probit models are considered in Case (1992), McMillen (1992), Pinkse and Slade (1998) and Beron and Vijverberg (2000).

20 For an extensive discussion, see Beron and Vijverberg (2000).

21 Other classic treatments of ML estimation in spatial models can be found in Whittle (1954), Besag (1974), and Mardia and Marshall (1984).

22 For a formal demonstration, see Anselin (1988a) and Kelejian and Prucha (1997).

23 For details, see, e.g. McMillen (1992), Pinkse and Slade (1998), Beron and Vijverberg (2000), and also, for general principles, Poirier and Ruud (1988).

24 For a careful consideration of these issues, see Kelejian and Prucha (1999a).

25 For technical details, see, e.g. Kelejian and Robinson (1993), Kelejian and Prucha (1998).

26 A recent application of this method is given in Bell and Bockstael (2000). An extension of this idea to the residuals of a spatial 2SLS estimation is provided in Kelejian and Prucha (1998).
27 See also Case (1992) and McMillen (1992) for a similar focus on heteroskedasticity in the spatial probit model.
28 See also the discussion in Haining (1990, pp. 131–3).
29 For example, for row-standardized weights, $S_0 = N$, and $I = e'We/e'e$. See Anselin and Bera (1998) for an extensive discussion.
30 Moran's I is not based on an explicit alternative and has power against both (see Anselin and Rey, 1991).
31 As shown in Anselin and Kelejian (1997) these tests are asymptotically equivalent.
32 Neither of these toolboxes include specification tests. Furthermore, S+Spatialstats has no routines to handle the spatial lag model.

References

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

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

Anselin, L. (1988b). A test for spatial autocorrelation in seemingly unrelated regressions. *Economics Letters* 28, 335–41.

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

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

Anselin, L. (1992a). Space and applied econometrics. Special Issue, *Regional Science and Urban Economics* 22.

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

Anselin, L. (1998). *SpaceStat Version 1.90*. http://www.spacestat.com.

Anselin, L. (1999). Spatial econometrics, An updated review. Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign.

Anselin, L. (2000). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference* (forthcoming).

Anselin, L., and A. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D.E.A. Giles (eds.) *Handbook of Applied Economic Statistics*, pp. 237–89. New York: Marcel Dekker.

Anselin, L., and R. Florax (1995a). Introduction. In L. Anselin and R. Florax (eds.) *New Directions in Spatial Econometrics*, pp. 3–18. Berlin: Springer-Verlag.

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*, pp. 21–74. Berlin: Springer-Verlag.

Anselin, L., and R. Florax (2000). *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.

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

Anselin, L., and H.H. Kelejian (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review* 20, 153–82.

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

Anselin, L., and S. Rey (1997). Introduction to the special issue on spatial econometrics. *International Regional Science Review* 20, 1–7.

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

Baltagi, B., and D. Li (2000a). Double length artificial regressions for testing spatial dependence. *Econometric Review*, forthcoming.

Baltagi, B., and D. Li (2000b). Prediction in the panel data model with spatial correlation. In L. Anselin and R. Florax (eds.) *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.

Barry, R.P., and R.K. Pace (1999). Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and its Applications* 289, 41–54.

Basu, S., and T.G. Thibodeau (1998). Analysis of spatial autocorrelation in housing prices. *Journal of Real Estate Finance and Economics* 17, 61–85.

Bell, K.P., and N.E. Bockstael (2000). Applying the generalized moments estimation approach to spatial problems involving micro-level data. *Review of Economics and Statistics* 82, 72–82.

Beron, K.J., and W.P.M. Vijverberg (2000). Probit in a spatial context: a Monte Carlo approach. In L. Anselin and R. Florax (eds.) *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B* 36, 192–225.

Besag, J., and P.A.P. Moran (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika* 62, 555–62.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *Annals of Probability* 10, 1047–50.

Brett, C., and J. Pinkse (1997). Those taxes all over the map! A test for spatial independence of municipal tax rates in British Columbia. *International Regional Science Review* 20, 131–51.

Brock, W.A., and S.N. Durlauf (1995). Discrete choice with social interactions I: Theory. NBER Working Paper No. W5291. Cambridge, MA: National Bureau of Economic Research.

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

Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics* 22, 491–508.

Case, A., 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 (1973). *Spatial Autocorrelation*. London: Pion.

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

Conley, T.G. (1996). *Econometric modelling of cross-sectional dependence*. Ph.D. dissertation. Department of Economics, University of Chicago, Chicago, IL.

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

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Davidson, R. and J.G. MacKinnon (1988). Double-length artificial regressions. *Oxford Bulletin of Economics and Statistics* 50, 203–17.

Driscoll, J.C., and A.C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80, 549–60.

Dubin, R. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics* 70, 466–74.

Dubin, R. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics* 22, 433–52.

Frees, E.W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69, 393–414.

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

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–54.

Kelejian, H., and I. Prucha (1997). Estimation of spatial regression models with autoregressive errors by two stage least squares procedures: a serious problem. *International Regional Science Review* 20, 103–11.

Kelejian, H., and I. Prucha (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.

Kelejian, H., and I. Prucha (1999a). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–33.

Kelejian, H.H., and I. Prucha (1999b). On the asymptotic distribution of the Moran I test statistic with applications. Working Paper, Department of Economics, University of Maryland, College Park, MD.

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 (1998). A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28, 389–417.

Kelejian, H.H., and D.P. Robinson (2000). The influence of spatially correlated heteroskedasticity on tests for spatial correlation. In L. Anselin and R. Florax (eds.) *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag (forthcoming).

King, M. (1981). A small sample property of the Cliff–Ord test for spatial correlation. *Journal of the Royal Statistical Society B* 43, 263–4.

Lahiri, S.N. (1996). On the inconsistency of estimators under infill asymptotics for spatial data. *Sankhya A* 58, 403–17.

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, 135–46.

MathSoft (1996). *S+SpatialStats User's Manual for Windows and Unix*. Seattle, WA: MathSoft, Inc.

McMillen, D.P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science* 32, 335–48.

Moran, P.A.P. (1948). The interpretation of statistical maps. *Biometrika* 35, 255–60.

Newey, W.K., and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–8.

Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120–6.

Pace, R.K., and R. Barry (1997a). Sparse spatial autoregressions. *Statistics and Probability Letters* 33, 291–7.

Pace, R.K., and R. Barry (1997b). Quick computation of spatial autoregressive estimators. *Geographical Analysis* 29, 232–46.

Pace, R.K., and R. Barry (1998). *Spatial Statistics Toolbox 1.0*. Real Estate Research Institute, Lousiana State University, Baton Rouge, LA.

Pace, R.K., R. Barry, C.F. Sirmans (1998). Spatial statistics and real estate. *Journal of Real Estate Finance and Economics* 17, 5–13.

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

Pinkse, J. (1998). Asymptotic properties of the Moran and related tests and a test for spatial correlation in probit models. Working Paper, Department of Economics, University of British Columbia, Vancouver, BC.

Pinkse, J. (2000). Moran-flavored tests with nuisance parameters. In L. Anselin and R. Florax (eds.) *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.

Pinkse, J., and M.E. Slade (1998). Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125–54.

Poirier, D.J., and P.A. Ruud (1988). Probit with dependent observations. *Review of Economic Studies* 55, 593–614.

Pötscher, B.M., and I.R. Prucha (1997). *Dynamic Nonlinear Econometric Models*. Berlin: Springer.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications* (2nd edn). New York: Wiley.

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

Smirnov, O., and L. Anselin (2000). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. Computational Statistics and Data Analysis.

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

Vijverberg, W. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* 76, 281–307.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika* 41, 434–49.