

Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models*

LUC ANSELIN

Regional Economics Applications Laboratory (REAL)

and

Department of Agricultural and Consumer Economics

University of Illinois, Urbana-Champaign

Urbana, IL 61801

USA

anselin@uiuc.edu

July 13, 2002

Abstract

This paper reviews a number of conceptual issues pertaining to the implementation of an explicit “spatial” perspective in applied econometrics. It provides an overview of the motivation for including spatial effects in regression models, both from a theory-driven as well as from a data-driven perspective. Considerable attention is paid to the inferential framework necessary to carry out estimation and testing and the different assumptions, constraints and implications embedded in the various specifications available in the literature. The review combines insights from the traditional spatial econometrics literature as well as from geostatistics, biostatistics and medical image analysis.

Key Words: spatial econometrics, spatial autocorrelation, spatial processes, model specification.

1 Introduction

Recent years have seen a virtual explosion in the application of spatial models in a range of fields in the social sciences in general, and in applied economics in particular (recent reviews are given in, e.g., Anselin and Bera 1998, Goodchild et al. 2000,

*This research was supported in part by NSF Grant BCS-9978058 to the Center for Spatially Integrated Social Science (CSISS) and by NSF/EPA Grant SES-0084213. Parts of the paper were presented earlier at the Pre-Conference Workshop on Spatial Analysis, Annual Meetings of the American Agricultural Economics Association, Chicago, IL, the ICPSR Summer Workshop on Spatial Regression, Ann Arbor, MI, the Joseph Havlicek Jr. Memorial Lecture in Applied Econometrics, The Ohio State University, Columbus, OH, and short courses at the Wharton School, University of Pennsylvania, Philadelphia, PA, and the University at Albany, Albany, NY. Comments, questions and suggestions from the participants at these presentations contributed to refining the arguments outlined in the paper and are greatly appreciated.

Anselin 2001b,c). Over time, the methodology of *spatial econometrics* (Paelinck and Klaassen 1979, Anselin 1988) has matured and evolved from an aspect of spatial statistics with primary application in regional science and analytical geography (Ord 1975, Cliff and Ord 1981), to an increasingly visible thread in formal econometric theory (for example, Conley 1999, Kelejian and Prucha 1999, 2001, Baltagi and Li 2001, Lee 2002). Similarly, when dealing with aggregate cross-sectional data in empirical work, testing for spatial autocorrelation and estimating models that formally incorporate spatial effects is no longer exceptional. Arguably, spatial regression techniques are beginning to become part of the toolbox of applied econometrics. In agricultural and resource economics, illustrations of this perspective can be found in Benirschka and Binkley (1994), Bockstael (1996), Weiss (1996), Nelson and Hellerstein (1997), Bell and Bockstael (2000), Florax et al. (2001), Hurley et al. (2001), Anselin et al. (2002), Irwin and Bockstael (2002), Kim et al. (2002) and Roe et al. (2002), among others.

While undoubtedly considerable progress has been made, most applications of spatial econometrics are rather limited in the way in which *spatial interaction* is incorporated in the model specifications. The typical approach is to distinguish between so-called *spatial lag* and *spatial error* models (see Anselin 1988). The former incorporate a spatially lagged dependent variable (Wy) on the right hand side of the regression model. Spatial error autocorrelation is either modeled directly, following the general principles of geostatistics, or by utilizing a spatial autoregressive process for the error term (for a recent review of these models, see Anselin and Bera 1998, Anselin 2001b). What is not always well understood in this process is that different spatial models induce sometimes radically different spatial correlation patterns, which do not necessarily match the underlying theoretical interaction model.

In this paper, I review some issues in the specification and interpretation of spatial regression models. The objective is to pull together results from a variety of disciplines in which different modeling strategies have been pursued, including spatial econometrics, biostatistics, medical image analysis and geostatistics. The review is aimed at a general audience of applied econometricians, without assuming familiarity with spatial econometrics. Hence, the approach taken is primarily pedagogic and mostly a reformulation and elaboration of ideas that have been outlined in some form in earlier review papers (specifically, Anselin and Bera 1998, Anselin 2001b,c, 2003). The emphasis throughout is less on technical aspects than on the underlying concepts and intuition. The objective is to highlight *unusual* results and suggest additional ways in which spatial models may be introduced in applied econometrics. A more technical treatment is pursued in a companion paper.

The focus of the paper is exclusively on spatial correlation in linear regression models, leaving the discussion of spatial *heterogeneity* aside.¹ Also, space-time models and spatial panel specifications are not considered explicitly, although most of the structures considered can be implemented in a space-time context without modification.²

The remainder of the paper consists of four sections and a conclusion. First, theoretical motivations for the inclusion of spatial dependence in a regression model are

¹For a general review of approaches to model spatial heterogeneity in a regression context, see, among others, Anselin (1988, 1990), Casetti (1997), Fotheringham et al. (1998) and LeSage (2002).

²See for example Elhorst (2001) for a recent review of some of the issues, and Anselin (1988, 2001b), for the specification of spatial panel data models.

considered. This is followed by a similar focus on data-driven motivations. Next, a number of conceptual issues are reviewed that pertain to the foundations for statistical inference in spatial regression models, including data models, the construction of spatial weights and asymptotics. Finally, some comments are formulated on the issue of ecological regression, i.e., the application of spatial regression models to aggregate units, such as counties or states.

2 Theory-Driven Specifications

The inclusion of spatial effects in applied econometric models is typically motivated either on theoretical grounds, following from the formal specification of spatial interaction in an economic model, or on practical grounds, due to peculiarities of the data used in an empirical analysis. I consider the theoretical perspective first.

In their description of the definition of *spatial econometrics*, Paelinck and Klaassen (1979) stressed the importance of spatial interdependence, the asymmetry of spatial relations, and the relevance of factors located in “other spaces.” This early formulation of the importance of spatial interaction was mostly based on pragmatic grounds. However, more recently, these concerns are also reflected in theoretical economic models of *interacting agents* and *social interaction*. Such models deal with questions of how the interaction between economic agents can lead to emergent collective behavior and aggregate patterns, and they assign a central role to location, space and spatial interaction. The substantive concepts receive different labels in various subfields, such as social norms, neighborhood effects, peer group effects, social capital, strategic interaction, copy-catting, yardstick competition and race to the bottom, to name a few. However, an important commonality is the need to formally specify the range and strength of the relations between the interacting agents, which in empirical practice translates into the need to specify a structure for spatial correlation.

Examples of such theoretical frameworks in economics are models of complex behavior built on principles from statistical mechanics, such as interacting particle systems and random field models (Brock and Durlauf 1995, Akerlof 1997, Durlauf 1997); macroeconomic models with mean field interaction (Aoki 1996); models for neighborhood spillover effects (Durlauf 1994, Borjas 1995, Glaeser et al. 1996); and models of increasing returns, path dependence and imperfect competition underlying the *new economic geography* (Fujita et al. 1999).

Rather than providing a detailed review of how specific spatial econometric models follow from these theoretical considerations, I will first focus on two particularly interesting forms that have seen considerable application in practice: spatial reaction functions and potential variables. Next, I present some remarks on spatial latent variable models.

2.1 Spatial Reaction Function

A spatial reaction function (Brueckner 2003) expresses how the magnitude of a decision variable for an economic agent depends on the magnitudes of the decision vari-

ables set by other economic agents.³ This provides the theoretical basis for a so-called *spatial lag* model, or, mixed regressive, spatial autoregressive model (Anselin 1988):

$$y = \rho W y + X \beta + \varepsilon, \quad (1)$$

where, as usual, y is an n by 1 vector of observations on the dependent (decision) variable, W is an n by n spatial weights matrix that formalizes the network structure (nodes and links) of the social network of the n agents,⁴ ρ is the spatial autoregressive parameter, X is an n by k matrix of observations on the exogenous variables, with an associated k by 1 regression coefficient vector β , and ε is a vector of random error terms.

Brueckner (2003) develops two theoretical frameworks for strategic interaction that yield a reaction function as the equilibrium solution. One is referred to as a *spillover* model, in which an agent i chooses the level of a decision variable, y_i , but the values of the y chosen by other agents (say, y_{-i} , where the $-i$ subscript refers to all agents other than i) affect its objective function as well. For example, this would be relevant in a situation where a farmer would determine the amount of farmland devoted to a crop by taking into account the amounts allocated by the other farmers in the system. Consequently, the objective function for each agent is:

$$U(y_i, y_{-i}; x_i'), \quad (2)$$

with x_i' as a row vector of (exogenous) characteristics of i . The solution to the usual objective maximization problem yields the reaction function as

$$y_i = R(y_{-i}, x_i'). \quad (3)$$

The spatial lag model (1) is an implementation of the reaction function obtained by specifying a linear functional form for R and by restricting the set of interacting agents to the “neighbor” structure expressed in the spatial weights W . Even though this imposes a large number of zero constraints in W in the structural form (1), the corresponding reduced form reveals a *global* range of spillovers:

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

in which the “Leontief inverse” $(I - \rho W)^{-1}$ links the decision variable y_i to all the x_i in the system through a so-called *spatial multiplier*. In addition, (4) illustrates how the dependent variable y_i at i is determined by the error terms at all locations in the system, and not just the error at i . This *simultaneity* makes the spatially lagged $W y$ variable endogenous, which necessitates specialized estimation techniques, such as maximum likelihood estimation or instrumental variables approaches (see, e.g., Ord 1975, Anselin 1988, Kelejian and Robinson 1993, Kelejian and Prucha 1998).

³Brueckner (2003) develops his argument in the context of public economics and refers to the “agents” as “jurisdictions,” which are the decision-making agents in this setting. Here, I will use the more generic term agents. See also below, Section 5 for a discussion of the interpretation of reaction functions when the units are aggregates of agents, which is an important distinction.

⁴A more formal treatment of the spatial weights is postponed until Section 4.2.

The particular form of the spatial multiplier in (4) is only one example out of a taxonomy of models for spatial spillovers, as presented in Anselin (2003). Different ranges for the spatial spillovers can be incorporated by applying the spatial lag operator (pre-multiplication by the spatial weights matrix W) to the y , X or ε terms in a regression specification. However, it is important to note that models that include Wy all induce a *global* form of spillovers.

Local forms of spillover are obtained from spatial lags for the explanatory variables (WX , see section 2.2) and particular error covariances, such as those induced by a spatial moving average model and a spatial error components model (see Anselin 2003, Anselin and Moreno 2003). These do not seem to fit the strategic interaction framework.

Brueckner (2003) illustrates how a number of empirical applications of strategic interaction models are special cases of his spillover model, with applications to state expenditures, pollution abatement and other forms of yardstick competition.⁵

A second theoretical framework is referred to as the *resource flow* model. Here, the agent's decision variable is not directly affected by the levels chosen by other agents, but only indirectly. The indirect effect follows from the presence of the value of a "resource" in the the individual agent's objective function,

$$U(y_i, s_i, x'_i), \quad (5)$$

where s_i is the amount of the resource available to agent i . For example, this could pertain to a farmer's decision of how much irrigation to apply to a field, where the resource s_i would be the amount of water available for this purpose. The interaction between agents follows from the way in which the resource is distributed among them, which depends both on the characteristics of each agent (x'_i , for example, the type of crop grown on the field), as well as on the decisions taken by the other agents (how much water they use):

$$s_i = H(y_i, y_{-i}; x'_i). \quad (6)$$

After substituting (6) into (5), the interaction variables y_{-i} become part of the objective function, and the resulting equilibrium solution takes the same form as the reaction function (3) for the spillover model. Brueckner (2003) illustrates how a number of tax competition and other strategic interaction models fall in the resource flow category and thus also suggest a spatial lag specification.⁶

It is important to note that the spillover and the resource flow models both lead to the same spatial lag econometric specification. Put differently, the spatial econometric model as such is not sufficient to identify the economic mechanism that leads to the presence (and empirical evidence) of spatial interaction. This is an example of the *inverse problem*, which is pervasive in spatial data analysis (see, for example, Ch. 8 in Chilès and Delfiner 1999). A similar problem is encountered in the interpretation of a spatial lag model as the expression of a spatial diffusion process.⁷ While diffusion

⁵Familiar examples are Case et al. (1993), Besley and Case (1995), Murdoch et al. (1997), and Bivand and Szymanski (1997, 2000).

⁶Examples are Brueckner (1998) and Saavedra (2000), among others.

⁷See, for example, the discussion in Baller et al. (2001) and Messner and Anselin (2003).

processes will lead to equilibrium outcomes that are compatible with a spatial lag specification, other processes may yield the same outcome as well. In other words, these different processes are *observationally equivalent*.

The essence of the problem is that a single cross-sectional data set contains insufficient information to identify the precise nature of the underlying mechanism. This is only one example of the kinds of identification problems encountered in spatial econometric models, as shown by Manski (1993) and Kelejian and Prucha (1997), among others.⁸

2.2 Potential Variables

A *potential* variable formally expresses the importance of “other spaces” in a regression specification. The theoretical motivation for this goes back to the early treatment of spatial interaction in the regional science literature by Isard (1960). There, the potential for interaction between an origin i and all destinations j was formulated as a sum of “mass” terms in the destination, suitably downscaled by a distance decay function. Specifically, with z_j as a measure of mass (e.g., income, population size) and $f(d_{ij})$ as a distance decay function, the potential at i becomes:

$$P_i = \sum_j f(d_{ij})z_j. \quad (7)$$

Note that the destinations need not be the same as the origins for this concept to work, although they typically are.⁹ Commonly used distance decay functions are the negative exponential, $f(d_{ij}) = e^{-\gamma d_{ij}}$, and the inverse distance function, $f(d_{ij}) = d_{ij}^{-\gamma}$.¹⁰

A concept related to notion of a potential is the *spatial cross-regressive* term or spatially lagged explanatory variable (WX), discussed in Florax and Folmer (1992). Such a variable consists of a weighted sum of values at other locations, or, for each observation i ,

$$[WX]_i = \sum_{j \neq i} w_{ij}x_j, \quad (8)$$

where the importance of each link i - j is expressed in the weights. The weights are perfectly general, and can include the distance decay specifications given above. Typically, they are based on geographic contiguity of the units of observation. The non-zero elements in the i -th row of W determine the range of interaction that affects location i , or, the range of spatial spillover.

In contrast to the spatially lagged dependent variable that follows from the spatial reaction function, the spatial cross-regressive term does not imply a multiplier effect.

⁸In point pattern analysis, this identification problem is referred to as the problem of *true vs. apparent* contagion. In a nutshell, the information in a cross-section is not sufficient to distinguish between clustering as resulting from a contagious process, or clustering as a result of spatial/structural heterogeneity (for details, see, among others, Upton and Fingleton 1985, Cressie 1993).

⁹For example, in Anselin et al. (1997) a variable is included that incorporates the effects of counties surrounding an MSA, as a “ring” variable, whereas those counties are not part of the data set for the dependent variable.

¹⁰The parameter γ is either estimated jointly with the other parameters of the model, which turns it into a non-linear specification, or set a priori. Common choices in the inverse distance model are the integers 1 and 2, the latter following from the Newtonian gravity model. See Isard (1960), for an extensive discussion.

Whether its range is global (including all other observations) or local (limited to a few “neighbors”) depends on the specification of the spatial weights, that is, on the extent to which zero restrictions (of the form $w_{ij} = 0$) have been imposed.

In a regression specification, the same variables may be included in non-lagged and spatially lagged form, as in

$$y = X\beta + WX\gamma + \varepsilon, \quad (9)$$

where zero restrictions can be imposed on specific elements of the parameter vectors β and γ . In addition, when both the original values (x_k) and the spatial lag (Wx_k) for the same variable are included, tests for the importance of distance decay can be performed (distance decay implies $\gamma_k < \beta_k$). Finally, in contrast to the spatial lag model, the spatial cross-regressive specification does not require specialized estimation methods and ordinary least squares remains unbiased for γ .

2.3 Spatial Latent Variable Models

So far, the theoretical models considered were formulated for a continuous dependent variable y . In applied econometrics, a more relevant specification often pertains to discrete dependent variables, where only a limited number of values are observed, such as the presence or absence of an action, or whether one out of a small number of alternative decisions has been taken. The standard approach to modeling such phenomena is to develop a specification for an unobserved underlying *latent* dependent variable for each agent, say y_i^* . The link between the latent variable and the observed discrete phenomenon is obtained by specifying a threshold, say c , such that y_i is observed whenever $y_i^* > c$. A typical application of this approach is in spatial land use models, where only the outcome is observed (one out of a set of discrete land use decisions), but the decision process is related to a latent variable, such as profitability, in a random utility framework.

A *spatial* latent variable model is a specification for this process where spatial correlation is introduced between the decision variables and/or in the error structure of the model. As a familiar point of departure, consider a latent linear regression model, where the unobserved dependent variable y_i^* is related to a 1 by k row vector of explanatory variables x_i' and an error term ε_i ,

$$y_i^* = x_i'\beta + \varepsilon_i \quad (10)$$

A matching spatial lag model would then be

$$y^* = \rho W y^* + X\beta + \varepsilon, \quad (11)$$

or, equivalently,

$$y_i^* = \rho \sum_{j \neq i} w_{ij} y_j^* + x_i'\beta + \varepsilon_i, \quad (12)$$

where y^* is the full n by 1 vector of the latent dependent variables. Note that this is compatible with a spatial reaction function for the latent variables, but *not* necessarily for the observed discrete outcomes. In other words, it is the latent $W y^*$ that is present in the actors' objective functions, such as equation (2), but *not* the observed $W y$. For

example, this would imply that it is the unobserved profitability of the neighbors parcels that enters in the utility function of a spatial land use model, but not the observed actual land uses.

The matching reduced form of the latent spatial lag process is as in (4), but again pertaining to the vector of latent variables

$$y^* = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} \varepsilon, \quad (13)$$

or,

$$y^* = (I - \rho W)^{-1} X \beta + u, \quad (14)$$

where $u = (I - \rho W)^{-1} \varepsilon$.

For simplicity, let the threshold $c = 0$ and let y_i be binary, taking on the value of 1 whenever $y_i^* > 0$. From (13), it follows that

$$y_i^* = \sum_j a_{ij} x_j' \beta + u_i, \quad (15)$$

where a_{ij} is the element i, j of the Leontief inverse matrix $(I - \rho W)^{-1}$, and $u_i = \sum_j a_{ij} \varepsilon_j$. The summation over j implies that y_i^* is determined not only by x_i' , but also by all the other x_j' in the system, and not only by ε_i , but by the other error terms ε_j in the system as well.

The discrete variable y_i is observed whenever $y_i^* > 0$ in (15), or

$$\sum_j a_{ij} x_j' \beta + u_i > 0. \quad (16)$$

For a symmetric random variable u_i , this yields the familiar condition

$$\text{Prob}[y_i = 1] = \text{Prob}[u_i < g_i(X, W, \beta, \rho)], \quad (17)$$

where $g_i = \sum_j a_{ij} x_j' \beta$ depends on X , W , β and ρ . Note that in (17), u_i is not i.i.d. as in the usual (non-spatial) model, but is a random variable whose marginal distribution is determined (in part) by the covariance matrix of the multivariate random vector u .

In the case of a standard normal ε_i , i.e., for a spatial probit model, the random vector u will be multivariate normal with a covariance matrix $\text{Cov}[u] = [(I - \rho W)'(I - \rho W)]^{-1}$. An important consequence of this complex covariance structure is that the marginal u_i will be *heteroskedastic*. This makes standard probit estimation inconsistent. In addition, due to the high degree of covariance, it is necessary to integrate out the $n - 1$ other random variables in order to obtain the marginal distribution for each individual u_i . Note that when ε_i does not follow a normal distribution, the transformed multivariate random variable $(I - \rho W)^{-1} \varepsilon$ is not necessarily well defined. For example, this is the case for a logit specification and for models of counts (Poisson models), where the resulting multivariate specification is intractable.

The essence of the problem is again the *simultaneity* of the y_i^* , which precludes an easy solution such as

$$\text{Prob}[y_i = 1] = \text{Prob} \left[\varepsilon_i < \rho \sum_{j \neq i} w_{ij} y_j^* + x_i' \beta \right]. \quad (18)$$

This expression is not operational, since the y_j^* that enter in the inequality condition are not observed, and are themselves determined by y_i^* .

This *simultaneous* model contrasts with a *conditional* approach, in which either y_j or $\text{Prob}[y_j = 1]$ are (assumed to be) observable. An implication of the conditional approach is that the spatial pattern of the y_j cannot be explained by the model. In other words, in order to be operational, a conditional model requires that the values of the y_j for the neighbors are obtained separately from the spatial model (12). More importantly, the two approaches are *not* equivalent, a point sometimes lost in the interpretation of results obtained in empirical practice.

A distinct advantage of the conditional approach is that standard estimation techniques can be applied, as long as the spatial lag term can be inserted as an observable on the right hand side of the condition (18). In practice, this can be implemented by a judicious spatial resampling (cluster sampling of non-contiguous clusters), although this typically involves a substantial loss of information.¹¹ In contrast, the complexity of the multivariate interactions in (17) invalidates standard probit or tobit techniques and requires specialized estimation and tests. This is still very much an active area of research. A number of suggestions have been formulated, such as the use of the EM (expectation, maximization) approach (McMillen 1992), general method of moments (GMM) estimation (Pinkse and Slade 1998), and simulation estimators, such as recursive importance sampling (Vijverberg 1997, Beron and Vijverberg 2002) and the Gibbs sampler (LeSage 2000).¹²

3 Data-Driven Specifications

In practice, the motivation for applying a spatial econometric model is typically not driven by formal theoretical concerns, but instead is a result of data “problems.” For example, the scale and location of the process under study does not necessarily match the available data, such as when agricultural land markets are studied with data at the county level. This mismatch will tend to result in model error structures that show a systematic spatial pattern. Also, explanatory variables are often “constructed” by spatial interpolation to make their scale compatible with that of the dependent variable. Again, this spatial prediction will tend to result in prediction “errors” that show systematic spatial variation. This problem is commonly encountered in models where economic outcomes are related to environmental or resource variables, such as air or water quality. Spatially aggregate measures of the latter are computed by interpolating measures obtained for a small set of monitoring stations, whose locations do not coincide with those of the economic agents. Another often encountered situation is when data on important variables are missing, and those variables show spatial structure, as is often the case in studies of tropical land use and deforestation. A common charac-

¹¹For example, if observations typically have four neighbors, the effective sample size in a non-contiguous subsample would be one fourth of the original sample size. This loss of degrees of freedom will result in a lower precision of the estimates, limiting this approach in practice to very large data sets.

¹²A review of estimation issues in spatial latent variable models is contained in Fleming (2002). Testing for spatial autocorrelation in probit and tobit models is considered in Kelejian and Prucha (2001) and Pinkse (2002).

teristic of these data problems is that the error term in a regression model will tend to be spatially correlated.¹³

In contrast to theory-driven models, which can be referred to as dealing with *substantive* spatial correlation, the correlation in the error models is referred to as a *nuisance*. From a technical viewpoint, the parameters used in the specification of the structure of the spatial correlation can therefore often be considered to be nuisance parameters, which facilitates estimation in some instances.¹⁴ The main objective of the econometric exercise is to obtain unbiased/consistent and efficient estimates for the regression parameters in the model (the β), while taking into account the spatial structure incorporated in the error covariance matrix. Formally, the main interest is in the familiar regression model,

$$y = X\beta + \varepsilon, \quad (19)$$

where the error covariance matrix, $Cov[\varepsilon\varepsilon']$, or, equivalently, $E[\varepsilon\varepsilon']$, specifies spatial covariance when the off-diagonal elements are non-zero, $E[\varepsilon_i\varepsilon_j] \neq 0$ (for $i \neq j$), in accordance with a given “spatial ordering” (Kelejian and Robinson 1992). Specific forms for the covariance structure are either specified directly (in so-called direct representation models) or follow from a spatial stochastic process model (such as a spatial autoregressive or spatial moving average model).

3.1 Spatial Filtering

An interesting perspective in the context of spatial correlation as a nuisance is the so-called *spatial filtering* approach. Similar to first differencing for time series, one can consider a form of spatial differencing.¹⁵ However, unlike the time series case, for row-standardized spatial weights, the first differencing leads to singularity, since $\rho = 1$ is outside the proper parameter space. For general, not row-standardized weights the parameter space is typically constrained to values much smaller than one, so that (unscaled) first differencing is similarly not allowed.

More formally, a spatial first difference can be expressed as

$$y - Wy = (X - WX)\beta + u, \quad (20)$$

or,

$$(I - W)y = (I - W)X\beta + u. \quad (21)$$

Since the row elements of W sum to one, the matrix $(I - W)$ is singular. Instead of using a “pure” first difference, a spatial autoregressive parameter must be included, as in

$$(I - \rho W)y = (I - \rho W)X\beta + u. \quad (22)$$

¹³This pertains not only to the error terms in classical linear regression models, but extends to generalized linear models (such as Poisson regression models) and generalized additive models (e.g., models for rates) as well. Examples can be found in Gotway and Stroup (1997), Waller et al. (1997b), Best et al. (1999), Lawson (2001), MacNab and Dean (2002), among others.

¹⁴For a review of the econometric issues, see Lancaster (2000).

¹⁵For some early examples, see Martin (1974) and Getis (1995).

The operation where a vector (or matrix) of data is pre-multiplied with the matrix expression $(I - \rho W)$ is called a *spatial filter*. Further premultiplying both sides of the equation by $(I - \rho W)^{-1}$ yields

$$y = X\beta + (I - \rho W)^{-1}u, \quad (23)$$

which is equivalent to a model with a spatially autoregressive error term:

$$\varepsilon = (I - \rho W)^{-1}u \quad (24)$$

$$(I - \rho W)\varepsilon = u \quad (25)$$

$$\varepsilon = \rho W\varepsilon + u. \quad (26)$$

In other words, specifying a spatial autoregressive process for the error term is equivalent to carrying out a standard regression on spatially filtered variables. However, unlike the time series case, the spatial autoregressive parameter cannot be obtained from a straightforward auxiliary regression, but estimation must be carried out jointly with that of the other model parameters. As a result, the spatial filter is mostly a convenient interpretation, but not a solution to the estimation problem.

Similar to the approach taken in (22), the spatial lag model (1) can be expressed as a spatial filter as well. However, the filter only pertains to the left hand side of the equation, as in

$$(I - \rho W)y = X\beta + u. \quad (27)$$

This can be interpreted as a way to clean the dependent variable y of the effects of spatial correlation, while maintaining the “correct” (i.e., consistent and efficient) estimates for β . However, as in the spatial error model, it is not possible to estimate the parameter ρ separately from the other parameters of the model, so that there is no gain in the estimation. Moreover, model (27) relates deviations from a spatial mean in y to levels for X , which may not be appropriate in many contexts. It should only be considered as a last resort, when there is no substantive basis for a lag model, but strong empirical evidence in its favor, such as indicated by the results of model diagnostics. Typically, however, other problems, such as scale mismatch, a poor selection of the weights and more serious misspecifications are likely to be the culprit and should be considered before resorting to an interpretation of the lag model as a spatial filter.

4 Inferential Framework

In this section, I review some issues that are seldom made explicit in applied work, but that are fundamental for the statistical inference in spatial econometrics. Three aspects in particular are considered, the data model underlying the statistical analysis, the choice of spatial weights and distance decay functions, and the asymptotic approach toward inference.

4.1 Data Model

In his classic text, Cressie (1993) outlines a taxonomy for spatial statistical analysis, distinguishing between point pattern analysis, geostatistical models and so-called lattice or regional models. In point pattern analysis, the main interest focuses on the

location of the observations as such, and whether this suggests clustering or other non-random patterns. Since point pattern analysis is seldom used in economic analysis, it will not be further considered here.¹⁶ Instead, the focus is on the distinction between the geostatistical and lattice approaches and the contexts in which they are appropriate in applied econometrics.

The fundamental difference between the geostatistical and lattice approaches can be related to the notion of a *data model* from the computer data base and geographic information science literatures. A data model is an abstraction of reality in a form amenable for analysis by a computer. In dealing with spatial data, the basic distinction is between *objects* and *fields* (Goodchild 1992).¹⁷

Objects are discrete entities and are typically represented in a geographic information systems (GIS) as points, lines and polygons (in a so-called vector GIS). In economic analysis, these objects correspond to economic agents or “jurisdictions,” with discrete locations in space, such as addresses, census tracts and counties. In contrast, fields pertain to continuous spatial distributions, represented as surfaces (in a so-called raster GIS). In economic analysis, one can envisage fields as price or risk surfaces, for example in the study of land values, crop yields or air quality.

Sometimes it is not immediately obvious whether an object or field approach is more appropriate. For example, land values could be studied as characteristics of discrete spatial objects (parcels) or could be viewed as samples from a continuous land value surface. Similarly, location-specific yield measures in a precision agriculture application could be conceived of as samples from a continuous yield surface, or, alternatively, be associated with a regular lattice overlaid on the field. The implications of the choice of framework for statistical inference are far-reaching.

Of the two, the object view and associated lattice data perspective seem to be the more natural for the study of discrete economic agents, and is the one typically associated with spatial econometrics. However, unlike “standard” econometric analysis, the observations in a spatial analysis of objects are no longer a representative *sample* from a population of objects. Instead, they consist of a single data point on the complete *spatial pattern* among them.

For example, a cross-sectional data set on economic variables for U.S. states is not a sample from a population of imaginary states, but its spatial pattern (e.g., as shown by a map for state incomes) is a single observation from all the possible stochastic patterns that an underlying mechanism may generate. In order to carry out statistical inference, a notion of a superpopulation or spatial random process is required (e.g., a Markov random field, MRF). This assumes the existence (conceptually) of a stochastic process that may generate many possible spatial patterns, of which the observed data is one. The objective of the analysis is then to characterize the spatial process by means of the observed spatial pattern. Both a spatial lag and a spatial error specification can be accommodated in this framework.

A number of unusual features of this approach are worth pointing out. Since the complete spatial pattern is the observation in lattice models, missing values are hard

¹⁶Apart from Cressie (1993), early comprehensive reviews of point pattern statistics are provided in Ripley (1981), Diggle (1983) and Upton and Fingleton (1985). A “modern” approach, based on Bayesian hierarchical modeling is outlined in Wolpert and Ickstadt (1998), Ickstadt and Wolpert (1999) and Best et al. (2000).

¹⁷See also Egenhofer et al. (1999) for a recent overview.

to deal with.¹⁸ In other words, a fully filled out space must be observed, without any “holes.” For example, in the analysis of land values or crop yields for parcels in a region, this would require that *all* the parcels in the region are observed.

Also, typically, the spatial units (such as U.S. states) are contiguous and exhaust the space, so that a notion of *interpolation* is impractical. For example, it would be hard to imagine predicting for a “state” in between Kansas and Colorado. Instead, spatial prediction applies to *extrapolation*, or the application of a model estimated from the observed spatial pattern to another set of spatial units, outside the observed set, or for a different time period.

In a lattice approach, observations can be viewed as nodes on a network, with links between them indicating the connectedness between nodes. This representation is very general, and easily extends beyond a pure geographic setting to economic and social networks (e.g., Friedkin 1998). It requires the formal specification of the network structure, which is implemented by means of spatial weights (see section 4.2).

A slightly different case occurs where a sample of discrete units is observed, each with a relevant set of “neighbors,” as in some cluster sample designs. For example, this may occur for a data set of land use by parcel, where each observation is matched to its nearest neighbors. While similar to the lattice setup, an important distinction is that the values for the neighbors are assumed known, and the pattern for the neighbors themselves is not explained. More precisely, the *conditional* distribution of land uses, conditional upon that observed for the neighbors is being modelled, not the joint distribution of all the land uses in the system. This important distinction was also encountered in the discussion of spatial latent variable models.

When the data model is a field, a geostatistical perspective is appropriate, since it views the observations as sample points from a continuous surface. The objective of a geostatistical analysis is to infer the spatial distribution for the surface from information provided by the pairwise association between the sample points, expressed as a function of the distance that separates them (for extensive reviews, see, among others, Cressie 1993, Goovaerts 1997, Chilès and Delfiner 1999, Stein 1999).

The geostatistical perspective is a natural framework when dealing with an incomplete set of spatial observations, where the objective is to predict values for unobserved locations. This focus on spatial *interpolation* (kriging), is a distinctive characteristic of geostatistics, and contrasts with the emphasis on estimation and inference in the lattice perspective. For applied econometric work, it is important to note that in a geostatistical approach to spatial regression models, the main interest therefore lies in optimal prediction (rather than estimation), exploiting the spatial patterns in the error term. In a geostatistical approach, there is no direct counterpart to a spatial lag model or spatial reaction function.

An additional aspect of the geostatistical perspective is that the choice of the number and location of sample points becomes part of the analysis, in contrast to the lattice perspective, where the locations are given and fixed. The optimal design of spatial sampling networks is a topic that is receiving increasing attention, particularly in the

¹⁸Note that this pertains to a classical statistical analysis. In a Bayesian viewpoint, both data and parameters are considered to be random, so that missing values can be incorporated in the same way as unobserved parameter values. A review of issues pertaining to “data augmentation,” although without treating spatial aspects of the issue can be found in Tanner (1996).

Table 1: Implications of Data Models

	Object	Field
GIS	vector	raster
Spatial Data	points, lines, polygons	surfaces
Location	discrete	continuous
Observations	process realization	sample
Spatial Arrangement	spatial weights	distance function
Statistical Analysis	lattice	geostatistics
Prediction	extrapolation	interpolation
Models	lag and error	error
Asymptotics	expanding domain	infill

area of environmental monitoring.¹⁹

In applied work, it is important to select the proper data model for the analysis. Primarily, this boils down to making a distinction between a design consisting of discrete objects and a design that is conceptualized as a sample from a continuous spatial surface. In many applications in applied econometrics, the latter is artificial when it comes to modeling economic agents. However, when the set of spatial observations is incomplete (i.e., with missing values or holes in the layout), the geostatistical/field approach is the only one that remains internally consistent. Also, a hybrid form is possible, when some of the explanatory variables are “interpolated” from a geostatistical model, but the model itself pertains to discrete agents, leading to a spatial errors in variables specification (see Anselin 2001c).

The contrast between the two data models is summarized in Table 1 .

4.2 Spatial Weights

A fundamental problem in the analysis of spatial correlation in a pure cross-sectional setting is the lack of identification of the parameters of the complete covariance matrix. The covariance matrix contains n potentially different variance terms σ_i^2 as well as $n(n-1)/2$ off-diagonal terms σ_{ij} ($= \sigma_{ji}$, since the covariance matrix is symmetric). Clearly, a single cross-section of n observations contains insufficient information to allow for the estimation of the individual variance-covariance terms. Asymptotics don't help, since the problem gets worse as the sample size grows (an *incidental parameter* problem). In sum, it will be necessary to impose a structure on the variance-covariance and to express it as a function of a small number of estimable parameters. In spatial regression analysis, this is approached from two main perspectives, matching the data models outlined in section 4.1. In a geostatistics-inspired approach, the covariance is specified *directly* as a function of the distance between pairs of observations. Different specifications for this distance decay function have been employed, but most are some variant of a negative exponential model. Examples and a discussion of estimation and

¹⁹Some recent reviews of the salient issues can be found in Arbia and Lafratta (1997), Müller (1998), and Wikle and Royle (1999), among others.

Table 2: Spatial Weights Matrix

0	1	0	1	1	0	0.00	0.33	0.00	0.33	0.33	0.00
1	0	0	1	1	0	0.33	0.00	0.00	0.33	0.33	0.00
0	0	0	0	1	1	0.00	0.00	0.00	0.00	0.50	0.50
1	1	0	0	1	0	0.33	0.33	0.00	0.00	0.33	0.00
1	1	1	1	0	0	0.25	0.25	0.25	0.25	0.00	0.00
0	0	1	0	0	0	0.00	0.00	1.00	0.00	0.00	0.00

identification issues can be found in Cook and Pocock (1983), Mardia and Marshall (1984), Dubin (1988, 1992), Dubin et al. (1999), and Anselin (2001a), among others.

In contrast, using an object view and corresponding lattice model, the covariance structure follows *indirectly* from the specification of the spatial weights matrix that underlies a spatial process model (Markov random field). Different specifications for the weights and their consequences are reviewed in the remainder of this section.

Before proceeding with this, it is worthwhile to briefly consider a third way of imposing structure, referred to in the literature as *spatial error components*. The full covariance structure follows from decomposing the error term into *components* and imposing a model for the variance and covariance of these terms. This approach is prevalent in hierarchical and multilevel modelling, where one error component is association with a model for (excess) heterogeneity and the other with a model for spatial variation, following the suggestion of Besag et al. (1991). Examples of the incorporation of spatial random effects in hierarchical Bayesian models in biostatistics are reviewed in, among others, Waller et al. (1997a,b) and Best et al. (1999). Applications using a multilevel modelling framework are illustrated in Langford et al. (1999a,b), Leyland et al. (2000) and Leyland (2001). Error components were introduced in spatial econometric specifications by Kelejian and Robinson (1995).²⁰

Formally, the spatial weights matrix is a $n \times n$ positive matrix (W) which specifies “neighborhood sets” for each observation. In each row i , a non-zero element w_{ij} defines j as being a neighbor of i . By convention, an observation is not a neighbor to itself, so that the diagonal elements are zero ($w_{ii} = 0$). Note that this definition is much broader than the term neighbor suggests. In most applications in applied econometrics, the neighbors are *contiguous* spatial units, as in Figure 1, but this can be easily generalized to any network structure. For example, in Figure 2, the six observations are nodes on a network and the existence of a neighbor relation matches the links between the nodes.²¹ The layouts in both Figure 1 and Figure 2 yield the same 6 by 6 spatial weights matrix, illustrated on the left hand side of Table 2 (from Anselin and Smirnov 1996). This is usually referred to as a *binary* contiguity matrix, since the weights are set to one for neighbors, and zero for others. For ease of interpretation and to make the parameter estimates between different models more comparable, the spatial weights matrix is typically row-standardized, as shown on the right hand side of Table 2. Each element in the standardized matrix, $w_{ij}^s = w_{ij} / \sum_j w_{ij}$, is between 0 and 1, which suggests that

²⁰See Anselin and Moreno (2003) for a review of some technical issues associated with specifying and testing for spatial error components.

²¹Figures 1 and 2 were first used as illustrations in Anselin and Smirnov (1996).

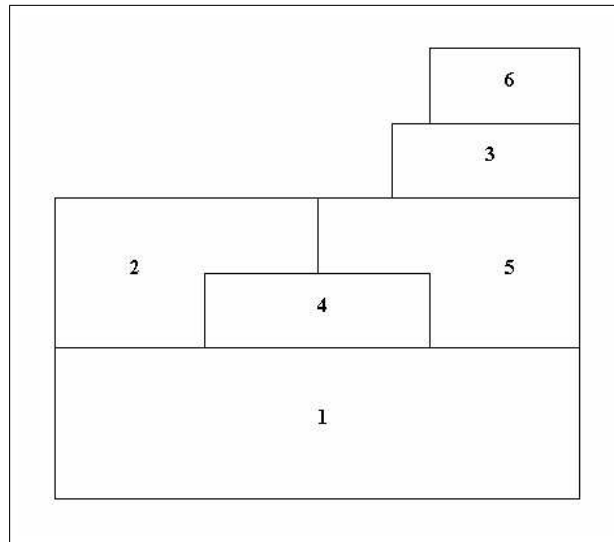


Figure 1: Example Spatial Layout.

a spatial lag operation (pre-multiplying a vector of observations by W) corresponds to an averaging of the neighboring values.²²

The specification of the weights matrix is a matter of some arbitrariness and is often cited as a major weakness of the lattice approach. A range of suggestions have been offered in the literature, based on contiguity, distance, as well as more general metrics.²³

A number of issues related to the specification of spatial weights require careful consideration in practice. First, even when the weights are based on simple contiguity, different weights structures may result for the same spatial layout. In the classic example of a regular square grid layout, the options are referred to as the *rook* case (only common boundaries), the *bishop* case (only common vertices) and the *queen* case (both boundaries and vertices). Depending on the criterion chosen, a location will have either four (rook, bishop) or eight (queen) neighbors on average (apart from edge effects). This implies quite different covariance structures for the associated random processes. Even in irregular spatial layouts, a decision must be made as to whether units that only share a common vertex should be considered to be neighbors (queen) or not (rook).

In practice, the construction of the neighbor structure of irregular spatial units is based on the digital boundary files in a GIS. Imprecision in the storage of the polygons and vertices can cause problems in this respect, yielding “islands” or other unexpected connectedness structures when deriving the spatial arrangement from these boundary

²²Note that the resulting matrix is no longer symmetric, since $\sum_j w_{ij} \neq \sum_i w_{ji}$, which needs to be accounted for in the computations of maximum likelihood estimates.

²³For more extensive reviews, see Cliff and Ord (1973, 1981), Upton and Fingleton (1985), Anselin (1988) and Anselin and Bera (1998).

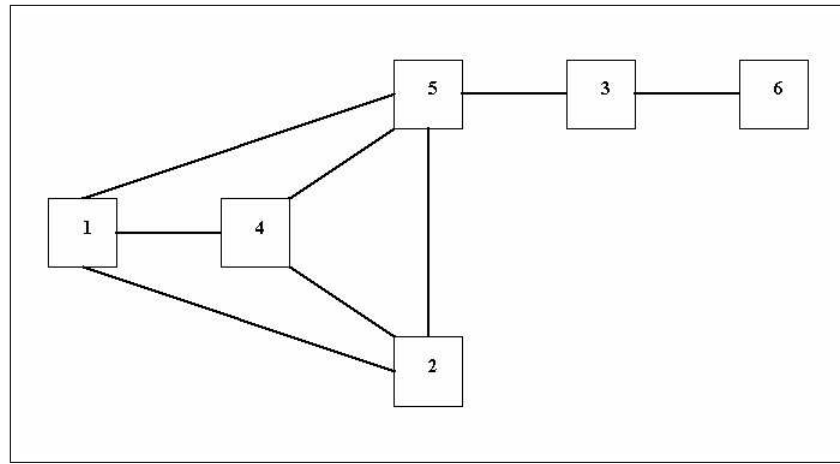


Figure 2: Spatial Network Structure.

files.²⁴

A second type of problem occurs when the spatial weights are based on a distance criterion, such that two units i and j are defined as neighbors when the distance between them (or, for areal units, the distance between their centroids) is less than a given critical value. When there is a high degree of heterogeneity in the spatial distribution of points or in the areas of regions, there may be no satisfactory critical distance. In those instances, a “small” distance will tend to yield a lot of islands (or, unconnected observations). Also, a distance chosen to ensure that each unit has at least one neighbor may result in an unacceptably large number of neighbors for the smaller units.

In empirical applications, this problem is encountered when building distance-based spatial weights for U.S. counties (Western counties have much larger areas than Eastern counties) or urban census tracts (core census tracts are much smaller than suburban census tracts). Similarly, when modelling land use or land values based on parcel data, problems will occur when the area of the parcels is highly variable. A common solution to this problem is to constrain the neighbor structure to the k -nearest neighbors, thereby precluding islands and forcing each unit to have the same number of neighbors. Whether or not this is appropriate in any given situation remains an empirical matter.²⁵

A third issue may arise when the weights are based on “economic” distance (Case et al. 1993) or another general metric, such as derived from a social network structure (Doreian 1980). Care must be taken to ensure that the resulting weights are meaningful, finite and non-negative. In addition, the “zero-distance problem” must be accounted for. The latter occurs when a distance measure, such as $d_{ij} = |z_i - z_j|$, becomes zero, due to rounding problems or because two observations show identical socio-economic

²⁴In GIS terminology, the polygons should be “clean” before the topology can be “built”.

²⁵One consequence of the choice of k -nearest neighbor weights is that the weights matrix becomes asymmetric. This is qualitatively different from a row-standardized (asymmetric) matrix derived from a symmetric contiguity matrix (for technical details, see Smirnov and Anselin 2001).

profiles. As a result, inverse distance weights such as $w_{ij} = 1/d_{ij}$ are undefined.

It is also important to maintain the weights matrix as *exogenous*. When the same variables are used to compute a general distance metric as are included in the model, the weights are unlikely to remain exogenous. Consequently, the resulting model specification becomes highly non-linear with endogeneity that must be instrumented out. Typically, this is not the result one has in mind when designing a weights matrix.

A slightly different type of “economic” weights follows when a block structure is imposed, as in Case (1991, 1992). This is a form of hierarchical spatial model, where all units that share a common higher order level are considered to be neighbors. For example, this would make all counties in the same state neighbors, yielding a block-diagonal spatial weights matrix. This structure precludes neighbors between the higher order levels and between lower order units that may be contiguous across higher order levels (such as neighboring counties in adjoining states). When implemented in row-standardized form, this type of weight also has a peculiar side effect. Since each weight in effect becomes $w_{ij} = 1/n_g$, where n_g is the number of units in the higher order level (counties in a state), the effect of each individual neighbor will disappear as $n_g \rightarrow \infty$. This happens in the limit, since the asymptotics operate on the cross-sectional dimension. Consequently, in the limit, the weights matrix becomes effectively zero, eliminating the effect of the spatial correlation.

Interestingly, this “economic weights” specification is the one employed in a recent paper by Lee (2002), where it is argued that OLS is consistent for the spatial lag parameter (the standard result is that it is not). Given the peculiar structure for the weights, this turns out to be a very special result, and does not pertain to the type of spatial “correlation” typically implemented in empirical spatial econometric work. Also, it would seem that the result is of limited practical use, since any type of meaningful correlation structure should *not* disappear in the limit.

The various weights specifications considered so far all share the property that their elements are fixed. It is straightforward to extend this notion and to incorporate parameters in the weights matrix, for example when the weights reflect the notion of a potential (see section 2.2), as in $w_{ij} = 1/d_{ij}^\beta$, or $w_{ij} = e^{-\beta d_{ij}}$ (see Anselin 1988, Chapter 3). The generalized Cliff-Ord weights are another well-known example, with $w_{ij} = b_{ij}^\alpha/d_{ij}^\beta$, where b_{ij} is the share of the common border between units i and j in the perimeter of i (and, typically, $b_{ij} \neq b_{ji}$), and α and β are parameters (Cliff and Ord 1973, 1981). In practice, these parameters are often set a priori, for example yielding a “gravity” like model with $w_{ij} = 1/d_{ij}^2$.²⁶

However, when the parameters of the weights elements are jointly estimated with the parameters in the model, the resulting specification is highly non-linear (for examples, see Anselin 1988, Bolduc et al. 1992, 1995). Moreover, when a scaling factor, such as a spatial autoregressive coefficient, is included together with parameterized weights, both sets of parameters are not necessarily identified. It is also important to

²⁶Other examples found in the literature are $w_{ij} = \sqrt{N_i N_j}/d_{ij}$, where N stands for the population (or mass) in an area (Ferrándiz et al. 1995), or $w_{ij} = N_i N_j/d_{ij}$ with a cutoff distance beyond which all weights are set to zero (Ferrándiz et al. 1999), or also $w_{ij} = e^{-d_{ij}/\delta}$ (Best et al. 1999). In the latter example, δ was set to a value of 33 to make the magnitude of the weight equal to 0.01 for two units that were the median inter-unit distance apart.

note that a weights matrix parameterized as a distance function is *not* equivalent to the direct representation model for the *covariance*. The structure of the corresponding covariance depends not only on the weights, but also on the choice of the spatial process. However, irrespective of the latter, there is no one-to-one match between the weights and the covariance.

It is also important to keep in mind that weights that are a function of distance depend on the scale of the distance metric. Ignoring this *scale-dependence* may lead to unexpected results (such as a zero weights matrix) when the coordinate units from which the distances are computed are non-standard.

There is very little formal guidance in the choice of the “correct” spatial weights in any given application. When the focus is on a model for substantive spatial dependence, care should be taken to match the spatial interaction patterns suggested by the theoretical framework (for example, a spatial reaction function implying a specific range of interaction). In other situations, the specification is much more ad hoc and sensitivity analysis of the results is very important. In practice, model validation techniques, such as a comparison of goodness-of-fit, or cross-validation, may provide ways to eliminate bad choices. Fortunately, empirical investigations can increasingly exploit both time and space dimensions (spatial panel data analysis), which opens up a number of opportunities to relax the structure of the weights matrix and employ non-parametric or semi-parametric methods to estimate a generic covariance structure, avoiding some of the strong priors required in the cross-sectional setting.²⁷

4.3 Asymptotics

Classical (as opposed to Bayesian) statistical inference in models that incorporate spatial correlation is based on asymptotic properties.²⁸ These properties only hold under a fairly restrictive set of assumptions, or, *regularity conditions*, which impose constraints on the degree of heterogeneity and range of dependence of the spatial stochastic process that is considered to generate the data. Unlike what is often assumed, these regularity conditions and the associated laws of large numbers (to prove consistency) and central limit theorems (to ascertain asymptotic normality) are quite special and *not* straightforward generalizations of the time series case of dynamic heterogeneous processes (treated, for example, in Pötscher and Prucha 1997). A few of the distinguishing characteristics of asymptotics in space are worth considering.

In a spatial setting, there are two fundamentally different ways to grow the “sample” to the limit (i.e., to obtain $n \rightarrow \infty$). These approaches match the two different data models considered in the paper.

²⁷See also Frees (1995), Driscoll and Kraay (1998) and Chen and Conley (2001) for a lattice perspective, and Sampson and Guttorp (1992), Guttorp and Sampson (1994) and Damian et al. (2001) for a geostatistical approach.

²⁸In a Bayesian approach to spatial regression analysis, both the data and the model parameters are considered to be random variables. Inference is based on an analysis of the posterior distribution of the model parameters, which is constructed by combining a prior distribution with a likelihood using Bayes’ theorem. In spatial regression analysis, this requires the specification of priors for the structure and the parameters of the spatial covariance matrix (in addition to the other model parameters). Recent overviews of this approach are provided in Clayton and Kaldor (1987), Handcock and Stein (1993), Ecker and Gelfand (1997), LeSage (1997), Wikle et al. (1998), Best et al. (1999), Berger et al. (2001), and Damian et al. (2001), among others.

In the field perspective and associated geostatistical approach to spatial modeling, the asymptotics are based on a fixed region, from which an increasingly denser sample of points is taken, or *infill* asymptotics (Cressie 1993). Intuitively, the denser and denser samples provide more and more information on the spatial distribution of the underlying surface. In contrast, in the object view and associated lattice model approach, there is no surface, and the asymptotics are obtained by adding more and more discrete objects to the sample, or, by *expanding* the *domain*.

The two paradigms are not equivalent. In fact, properties that hold under one framework do not necessarily hold under the other (Lahiri 1996). To illustrate this point, consider a setting where the spatial weights matrix defines neighbors as those points within a given fixed distance band. In an expanding domain framework, there is no problem, since adding new objects only affects the neighbor structure for those observations at the margin.²⁹ In other words, the number of neighbors and the implied “range” of spatial correlation is not (substantially) affected by growing the sample. In contrast, with infill asymptotics, the sample would become increasingly denser, resulting in more and more points meeting the critical distance criterion for each observation. Therefore, the number of neighbors will increase with the sample size, effectively removing the spatial correlation (for row-standardized weights) as the sample grows. This runs counter to the regularity conditions required for expanding domain asymptotics.

In a nutshell, the regularity conditions required for the expanding domain asymptotics in lattice models boil down to limits on the heterogeneity of the process (variance and higher order moments) and constraints on the range of spatial dependence. The latter can be thought of as a formal expression of Tobler’s “first law of geography,” according to which everything depends on everything else, but “close” things more so (Tobler 1979). Similar regularity conditions are required in the geostatistical approach, to ensure that the covariance structure that follows from the specified distance decay function is positive definite (for technical details, see Cressie 1993). Formally, the conditions pertain to summability and differentiability of the elements of the covariance matrix. An extensive and unifying technical treatment of these issues was recently provided by Kelejian and Prucha (1998, 1999, 2001) (for antecedents, see also Magnus 1978, Mardia and Marshall 1984, Mandy and Martins-Filho 1994).

In applied spatial econometrics based on the lattice approach, sufficient conditions are typically satisfied by limiting the number of neighbors in the weights and ensuring this does not grow with the sample size. Conditions on the heterogeneity are complicated by the fact that many spatial processes induce heteroskedasticity (or, non-stationarity), which must be properly accounted for.³⁰ In practice, weights based on contiguity or similar principles (distance bands) will satisfy these regularity conditions. Matters are less straightforward when complex weights are introduced (such as parameterized distance functions) for which it is not always possible to establish

²⁹However, as discussed in Anselin and Kelejian (1997), the fact that the weights change at the boundary is a non-standard situation. As Kelejian and Prucha (1998, 1999) have pointed out, this requires the use of *triangular arrays* as well as specialized central limit theorems in order to establish the asymptotic properties.

³⁰Note that in some treatments stationarity is a crucial assumption, which rules out spatial processes that induce heteroskedasticity. See, for example, the central limit theorems based on Bolthausen (1982) used as the basis for the GMM estimator in Conley (1999). Also, note that in a geostatistical approach, there is no such induced heteroskedasticity.

that the regularity conditions are satisfied. Evidence of the violation of these assumptions is sometimes provided by “weird” results, such as negative variance estimates and explosive spatial interaction functions.

5 Ecological Regression

In empirical practice, the estimation of models such as a spatial reaction function, specified in the form of a spatial lag model (1), is often carried out for aggregate spatial units of observation, such as counties or census tracts. In the statistical literature, this is referred to as *ecological regression*, and often criticized as yielding invalid inference, the so-called *ecological fallacy* problem. More precisely, the ecological fallacy pertains to cross-level inference or cross-level bias. This is what happens when parameters and other characteristics of a distribution are estimated at an aggregate level, but behavioral and socio-economic relations are inferred for another, disaggregate level.

5.1 Ecological Fallacy

An enormous literature has been devoted to the problem of ecological fallacy in sociology, political science and economics, going back to the classics of Gehlke and Biehl (1934), Robinson (1950) and Goodman (1953).³¹ In economics, this issue is closely related to the aggregation problem, or the extent to which micro-relationships can be inferred from macro estimates (for recent review of the relevant issues, see Stoker 1993). In general, unless extremely rigid (and unrealistic) homogeneity constraints are imposed, it is impossible to transfer findings from the macro level to a micro interpretation.³²

In practice, this is easily overlooked, but even in very simple situations, and with a high degree of homogeneity, the ecological approach creates problems of interpretation. Consider a regression at the individual level, where individuals are stratified by group, and both individual-level variates as well as group-wise aggregates are included in the model (the example is adapted from Greenland 2002, p. 390):

$$y_{ik} = \alpha + x_{ik}\beta + \bar{x}_k\gamma + \epsilon_{ik}, \quad (28)$$

where x_{ik} is a characteristic of individual i in group k (e.g., income for household i in county k) and $\bar{x}_k = \sum_i x_{ik}/n_k$ (with n_k as the group size) is the group average for that characteristic (e.g., county average income). In the literature, β is referred to as the individual effect and γ as the contextual effect.³³ The corresponding macro regression relates the group averages to each other, or

$$\bar{y}_k = \alpha + \bar{x}_k(\beta + \gamma) + \bar{\epsilon}_k, \quad (29)$$

³¹For recent overviews of this extensive literature, see also Achen and Shively (1995) and King (1997).

³²In a very stylized setting, often used in voting rights analysis, King (1997) has suggested a “solution” to this problem, which he refers to as “ecological inference” or *ei*. The essence of this approach is to treat the unobserved individual parameters in a random coefficient framework and to simulate their posterior distribution, using additional information from individual-level constraints. For a discussion of the role of spatial effects in this model, see Anselin and Cho (2002).

³³For a related discussion of identification issues in economic models of interaction, see Manski (1993).

with the group averages as $\bar{y}_k = \sum_i y_{ik}/n_k$. When the groups do not contain an equal number of members, the error term in (29) will become heteroskedastic. In other words, at the aggregate level, heteroskedasticity should be expected, and an i.i.d. assumption for the errors is incompatible with the aggregation rule.

In addition, the coefficient of the average \bar{x}_k in the aggregate model no longer allows for the separate identification of the individual and contextual effects, but confounds the two. More precisely, even when there is no within-group heterogeneity (all the groups have the same β and γ coefficients), the estimate from the aggregate model only corresponds with an individual-level coefficient when there is no contextual effect ($\gamma = 0$). Similarly, it only corresponds to a “pure” contextual effect when there is no individual effect ($\beta = 0$).³⁴

5.2 Spatial Aggregation

In spatial analysis, an additional twist is added to the ecological regression problem, in that it is not only the level of aggregation that matters, but also how “elemental units” are combined spatially. This is referred to as the zoning problem or the *modifiable areal unit problem* (MAUP), first illustrated by the “million or so correlation coefficients” in Openshaw and Taylor (1979). Considerable attention has been paid to the interrelation between MAUP and spatial correlation (Arbia 1989). More recently, statisticians have approached this problem as a special case of the *change of support problem* (COSP) (see Cressie 1996, Gotway and Young 2002).

Extending the example in (28) with a spatial autoregressive term, some of the complexities of spatial ecological regression become apparent. At the individual level, a spatial lag specification would be

$$y_{ik} = \rho \sum_{j=1}^n w_{ij} y_{jh} + \alpha + x_{ik} \beta + \bar{x}_k \gamma + \varepsilon_{ik}, \quad (30)$$

where it is important to note that the non-zero weights in w_{ij} are not limited to neighbors that belong to the same group. The specification in (30) is typically what one has in mind when implementing a spatial reaction function for economic agents i .

By comparison, a spatial lag specification at the aggregate level, for the groups g , with $g = 1, \dots, G$, would be

$$\bar{y}_k = \lambda \sum_{g=1}^G w_{kg} \bar{y}_g + \alpha + \bar{x}_k (\beta + \gamma) + \bar{\varepsilon}_k, \quad (31)$$

with w_{kg} as the elements of a group-level spatial weights matrix of dimension G by G that reflects the neighbor structure for the aggregate spatial units.

While the regressive part in (31) is a straightforward extension of the non-spatial case, the autoregressive part is *not* an aggregate of the spatial autoregressive terms in (30). Several factors preclude a simple aggregation. Consider the case of an aggregated spatial weights matrix based on simple contiguity that would be constructed from collapsing the rows and columns for the elements in each group. Formally, this is

³⁴See Greenland (2002) and also Stoker (1993) for a more elaborate discussion.

accomplished by means of an n by G aggregation matrix H , with elements $h_{ig} = 1$ for $i \in g$ and zero otherwise, such that

$$W^G = H'W^nH, \quad (32)$$

with W^n as the individual-level weights. This yields a G by G matrix with elements w_{kg}^G obtained as the sums of all the weights w_{ij}^n for which $i \in k$ and $j \in g$. If the original weights W^n contained non-zero elements for agents in the same group, the aggregate weights from (32) should have non-zero diagonal elements, $w_{kk} \neq 0$. This is typically ruled out (see section 4.2). Therefore, zero diagonals in the weights for an aggregate-level model are inconsistent with a spatial aggregation of the individual spatial weights.

The aggregated between-group weights w_{kg} (with $k \neq g$) are the sum of the weights for individual agents that were cross-group neighbors in each group (the number of weights $w_{ij}^n \neq 0$ for which $i \in k$ and $j \in g$). In a spatial aggregation over irregular units, the share of such neighbors in each group is unlikely to be constant, yielding unequal weights w_{kg} for any given k . In contrast, in a typical group-level application, contiguity weights would be set equal for all elements in the same row, again violating the proper spatial aggregation.

More importantly, the aggregate over groups of the individual-level spatial lag terms is not equal to the spatial lag of the aggregate values. This can be traced back to the aggregation over the reduced form (4). At the individual level, the y_{ik} are a weighted average of the x_j in the system, with the weights for each i corresponding to the row elements of the inverse matrix $A = (I - \rho W)^{-1}$. Formally, and ignoring the structure of the error term, at the individual level,

$$y_{ik} = \sum_{j=1}^n a_{ij}(\alpha + x_{jh}\beta + \bar{x}_{jh}\gamma) + e, \quad (33)$$

where j is the index of the individual observation and h is a generic group indicator. The summation in (33) is over all j and includes elements in group k as well as in other groups in the system. The weights in a_{ij} are unequal. The aggregation of A by means of the matrix H used above will yield a group level spatial multiplier $A^G = H'AH$. For the same reasons as outlined for the spatial weights, the spatial multiplier matrix for the groups ($[I - \lambda W^G]^{-1}$) will be inconsistent with a spatial aggregation of the individual-level weights. As a result, the group average of the individual y_{ik} obtained from the reduced form will not equal the \bar{y}_k from the group-level reduced form.

Given these problems, one might be tempted to dismiss spatial lag models estimated for aggregate units. Clearly, a naive interpretation of the parameters of such a model is misguided when they are considered as proxies for individual-level parameters that reflect substantive spatial dependence, e.g., as implied by a spatial reaction function. However, there remain many contexts where the interest is in the aggregate object considered in its own merit, and not as an aggregation of lower level units. For example, in many applications of policy evaluation, the focus would be on state or county-level economic indicators as such, without requiring an explicit link to the micro units. In such situations, there is no problem with the use of ecological spatial lag models as specification of substantive economic relations.³⁵

³⁵See also Schwartz (1994), for related arguments in the public health arena.

6 Conclusions

As a spatial perspective is becoming increasingly common in applied econometric work, it is important to keep in mind the formal framework within which proper estimation and inference can be carried out. Different model specifications imply different spatial correlation structures that may not always be compatible with the economic theory behind the interaction model. The data are often problematic, and prevent the use of an optimal spatial scale in many empirical situations. Choices must be made about the data model and appropriate statistical paradigm. The specification of spatial weights is often ambiguous and the conceptual interaction model does not always match the formal simultaneous or conditional specification. Moreover, the danger of ecological fallacy lurks everywhere.

With more user friendly software available, these important choices may be hidden from the analyst, or a particular perspective forced on the unsuspecting practitioner. The goal of this paper was to focus attention on a number of conceptual issues that must be resolved in order to obtain a sound design for a spatial econometric analysis. While the discussion was mostly informal, the main ideas remain valid and hopefully will guide applied econometricians in future spatial work. A more technical treatment is offered in a companion paper.

References

- Achen, C. H. and Shively, W. P. (1995). *Cross-Level Inference*. University of Chicago Press, Chicago, IL.
- Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, 65:1005–1027.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Anselin, L. (1990). Spatial dependence and spatial structure instability in applied regression analysis. *Journal of Regional Science*, 30:185–207.
- Anselin, L. (2001a). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, 97:113–139.
- Anselin, L. (2001b). Spatial econometrics. In Baltagi, B., editor, *A Companion to Theoretical Econometrics*, pages 310–330. Blackwell, Oxford.
- Anselin, L. (2001c). Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics*, 83(3):705–710.
- Anselin, L. (2003). Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review*, 26. forthcoming.
- Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and Giles, D. E., editors, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York.

- Anselin, L., Bongiovanni, R., and Lowenberg-DeBoer, J. (2002). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. Technical Report 02-T-2, Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign, IL.
- Anselin, L. and Cho, W. T. (2002). Spatial effects and ecological inference. *Political Analysis*, 10(3):276–297.
- Anselin, L. and Kelejian, H. H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review*, 20:153–182.
- Anselin, L. and Moreno, R. (2003). Properties of tests for spatial error components. *Regional Science and Urban Economics*. forthcoming.
- Anselin, L. and Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, 36:67–89.
- Anselin, L., Varga, A., and Acs, Z. (1997). Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics*, 42:422–448.
- Aoki, M. (1996). *New Approaches to Macroeconomic Modelling*. Cambridge University Press, Cambridge.
- Arbia, G. (1989). *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Arbia, G. and Lafratta, G. (1997). Evaluating and updating the sample design in repeated environmental surveys: Monitoring air quality in Padua. *Journal of Agricultural, Biological and Environmental Statistics*, 2:451–466.
- Baller, R., Anselin, L., Messner, S., Deane, G., and Hawkins, D. (2001). Structural covariates of U.S. county homicide rates: Incorporating spatial effects. *Criminology*, 39(3):561–590.
- Baltagi, B. and Li, D. (2001). Double length artificial regressions for testing spatial dependence. *Econometric Reviews*, 20(1):31–40.
- Bell, K. P. and Bockstael, N. E. (2000). Applying the generalized moments estimation approach to spatial problems involving micro-level data. *The Review of Economics and Statistics*, 82:72–82.
- Benirschka, M. and Binkley, J. K. (1994). Land price volatility in a geographically dispersed market. *American Journal of Agricultural Economics*, 76:185–195.
- Berger, J. O., Oliveira, V. D., and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96:1361–1374.

- Beron, K. J. and Vijverberg, W. P. (2002). Probit in a spatial context: a Monte Carlo approach. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg. forthcoming.
- Besag, J., York, J., and A.Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59.
- Besley, T. and Case, A. (1995). Incumbent behavior: Vote-seeking, tax-setting and yardstick competition. *American Economic Review*, 85:25–45.
- Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999). Bayesian models for spatially correlated disease and exposure data. In Bernardo, J., Berger, J., Dawid, A., and Smith, F., editors, *Bayesian Statistics 6*, pages 131–156. Oxford University Press, New York, NY.
- Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association*, 95:1076–1088.
- Bivand, R. and Szymanski, S. (1997). Spatial dependence through local yardstick competition: Theory and testing. *Economics Letters*, 55:257–265.
- Bivand, R. and Szymanski, S. (2000). Modeling the spatial impact of the introduction of compulsory competitive tendering. *Regional Science and Urban Economics*, 30:203–219.
- Bockstael, N. E. (1996). Modeling economics and ecology: The importance of a spatial perspective. *American Journal of Agricultural Economics*, 78:1168–1180.
- Bolduc, D., Laferrière, R., and Santarossa, G. (1992). Spatial autoregressive error components in travel flow models. *Regional Science and Urban Economics*, 22:381–385.
- Bolduc, D., Laferrière, R., and Santarossa, G. (1995). Spatial autoregressive error components in travel flow models, an application to aggregate mode choice. In Anselin, L. and Florax, R. J., editors, *New Directions in Spatial Econometrics*, pages 96–108. Springer-Verlag, Berlin.
- Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10:1047–1050.
- Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, 85:365–390.
- Brock, W. A. and Durlauf, S. N. (1995). Discrete choice with social interactions I: Theory. Working Paper W5291, National Bureau of Economic Research, Cambridge, MA.
- Brueckner, J. K. (1998). Testing for strategic interaction among local governments: The case of growth controls. *Journal of Urban Economics*, 44:438–467.

- Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, 26. forthcoming.
- Case, A., Rosen, H. S., and Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52:285–307.
- Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, 59:953–965.
- Case, A. C. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics*, 22:491–508.
- Casetti, E. (1997). The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review*, 20:9–33.
- Chen, X. and Conley, T. G. (2001). A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 105:59–83.
- Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics, Modeling Spatial Uncertainty*. John Wiley & Sons, New York, NY.
- Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.
- Cliff, A. and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion, London.
- Cliff, A. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. Pion, London.
- Conley, T. G. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics*, 92:1–45.
- Cook, D. and Pocock, S. (1983). Multiple regression in geographic mortality studies, with allowance for spatially correlated errors. *Biometrics*, 39:361–371.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.
- Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3:159–180.
- Damian, D., Sampson, P. D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12:161–178.
- Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Doreian, P. (1980). Linear models with spatially distributed data, spatial disturbances or spatial effects. *Sociological Methods and Research*, 9:29–60.
- Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80:549–560.

- Dubin, R. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated errors. *Review of Economics and Statistics*, 70:466–474.
- Dubin, R. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22:433–452.
- Dubin, R., Pace, R. K., and Thibodeau, T. G. (1999). Spatial autoregression techniques for real estate data. *Journal of Real Estate Literature*, 7:79–95.
- Durlauf, Stephen, N. (1994). Spillovers, stratification and inequality. *European Economic Review*, 38:836–845.
- Durlauf, Stephen, N. (1997). Statistical mechanics approaches to socioeconomic behavior. In Arthur, B. W., Durlauf, S. N., and Lane, D. A., editors, *The Economy as an Evolving Complex System II*, pages 81–104. Addison-Wesley, Reading, MA.
- Ecker, M. and Gelfand, A. (1997). Bayesian variogram modeling for an isotropic spatial process. *Journal of Agricultural, Biological and Environmental Statistics*, 2:347–369.
- Egenhofer, M. J., Glasgow, J., Günther, O., Herring, J. R., and Peuquet, D. J. (1999). Progress in computational methods for representing geographical concepts. *International Journal of Geographical Information Science*, 13(8):775–796.
- Elhorst, J. P. (2001). Dynamic models in space and time. *Geographical Analysis*, 33:119–140.
- Ferrándiz, J., López, A., Llopis, A., Morales, M., and Tejerizo, M. (1995). Spatial interaction between neighbouring counties: Cancer mortality data in Valencia (Spain). *Biometrics*, 51:665–678.
- Ferrándiz, J., López, A., and Sanmartín, P. (1999). Spatial regression models in epidemiological studies. In Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., and Bertollini, R., editors, *Disease Mapping and Risk Assessment for Public Health*, pages 203–215. John Wiley, New York.
- Fleming, M. (2002). Techniques for estimating spatially dependent discrete choice models. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg. forthcoming.
- Florax, R. and Folmer, H. (1992). Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators. *Regional Science and Urban Economics*, 22:405–432.
- Florax, R. J., Voortman, R., and Brouwer, J. (2001). Spatial dimensions of precision agriculture: Estimating millet yield functions for Sahelian coversands. Technical report, Department of Spatial Economics, Free University, Amsterdam, The Netherlands.

- Fotheringham, A. S., Brundson, C., and Charlton, M. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30:1905–1927.
- Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics*, 69:393–414.
- Friedkin, N. E. (1998). *A Structural Theory of Social Influence*. Cambridge University Press, New York, NY.
- Fujita, M., Krugman, P., and Venables, A. (1999). *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge, MA.
- Gehlke, C. and Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29:169–170.
- Getis, A. (1995). Spatial filtering in a regression framework: Examples using data on urban crime, regional inequality, and government expenditures. In Anselin, L. and Florax, R. J., editors, *New Directions in Spatial Econometrics*, pages 172–185. Springer-Verlag, Berlin.
- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. (1996). Crime and social interactions. *Quarterly Journal of Economics*, 111:507–548.
- Goodchild, M. F. (1992). Geographical data modeling. *Computers and Geosciences*, 18:401–408.
- Goodchild, M. F., Anselin, L., Appelbaum, R., and Harthorn, B. (2000). Toward spatially integrated social science. *International Regional Science Review*, 23(2):139–159.
- Goodman, L. A. (1953). Ecological regression and behavior of individuals. *American Sociological Review*, 18:663–664.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, NY.
- Gotway, C. A. and Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological and Environmental Statistics*, 2(2):157–178.
- Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97:632–648.
- Greenland, S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine*, 21:389–395.
- Guttorp, P. and Sampson, P. (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. In Patil, G. and Rao, C., editors, *Handbook of Statistics XII: Environmental Statistics*, pages 663–690. North-Holland/Elsevier, New York, NY.

- Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35:403–410.
- Hurley, T., Kilian, B., Malzer, G., and Dikici, H. (2001). The value of information for variable rate nitrogen applications: A comparison of soil test, topographical and remote sensing information. Technical report, Department of Applied Economics, University of Minnesota.
- Ickstadt, K. and Wolpert, R. L. (1999). Spatial regression for marked point processes. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 6*, pages 323–341. Oxford University Press, New York, NY.
- Irwin, E. G. and Bockstael, N. E. (2002). Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*, 2:31–54.
- Isard, W. (1960). *Methods of Regional Analysis*. MIT Press, Cambridge, MA.
- Kelejian, H. H. and Prucha, I. (1997). Estimation of spatial regression models with autoregressive errors by two stage least squares procedures: A serious problem. *International Regional Science Review*, 20:103–111.
- Kelejian, H. H. and Prucha, I. (1998). A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.
- Kelejian, H. H. and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.
- Kelejian, H. H. and Prucha, I. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2):219–257.
- Kelejian, H. H. and Robinson, D. P. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita country police expenditures. *Regional Science and Urban Economics*, 22:317–333.
- Kelejian, H. H. and Robinson, D. P. (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 Robinson, D. P. (1995). Spatial correlation: A suggested alternative to the autoregressive model. In Anselin, L. and Florax, R. J., editors, *New Directions in Spatial Econometrics*, pages 75–95. Springer-Verlag, Berlin.
- Kim, C.-W., Phipps, T. T., and Anselin, L. (2002). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management*. forthcoming.
- King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press, Princeton, NJ.

- Lahiri, S. (1996). On the inconsistency of estimators under infill asymptotics for spatial data. *Sankhya A*, 58:403–417.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413.
- Langford, I. H., Leyland, A. H., Rasbash, J., and Goldstein, H. (1999a). Multilevel modelling of the geographical distributions of diseases. *Applied Statistics*, 48:253–268.
- Langford, I. H., Leyland, A. H., Rasbash, J., Goldstein, H., Day, R. J., and McDonald, A.-L. (1999b). Multilevel modelling of area-based health data. In Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., and Bertollini, R., editors, *Disease Mapping and Risk Assessment for Public Health*, pages 217–228. John Wiley & Sons, Chichester.
- Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology*. John Wiley & Sons, New York, NY.
- Lee, L.-F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, 18(2):252–277.
- LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20:113–129.
- LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32:19–35.
- LeSage, J. P. (2002). A family of geographically weighted regression models. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg. forthcoming.
- Leyland, A. H. (2001). Spatial analysis. In Leyland, A. H. and Goldstein, H., editors, *Multilevel Modelling of Health Statistics*, pages 143–157. John Wiley & Sons, Chichester.
- Leyland, A. H., Langford, I. H., Rabash, J., and Goldstein, H. (2000). Multivariate spatial models for event data. *Statistics in Medicine*, 19:2469–2478.
- MacNab, Y. C. and Dean, C. B. (2002). Spatial-temporal modelling of rates for the construction of disease maps. *Statistics in Medicine*, 21:347–358.
- 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.
- Mandy, D. and Martins-Filho, C. (1994). A unified approach to asymptotic equivalence of Aitken and feasible Aitken instrumental variables estimators. *International Economic Review*, 35:957–979.

- Manski, C. F. (1993). Identification of endogenous social effects: The reflexion problem. *Review of Economic Studies*, 60:531–542.
- Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146.
- Martin, R. (1974). On spatial dependence, bias and the use of first spatial differencing in regression analysis. *Area*, 6:185–194.
- McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32:335–348.
- Messner, S. F. and Anselin, L. (2003). Spatial analyses of homicide with areal data. In Goodchild, M. and Janelle, D., editors, *Spatially Integrated Social Science*. Oxford University Press, New York, NY. forthcoming.
- Müller, W. G. (1998). *Collecting Spatial Data, Optimum Design of Experiments for Random Fields*. Physica-Verlag, Heidelberg.
- Murdoch, J. C., Sandler, T., and Sargent, K. (1997). A tale of two collectives: Sulfur versus nitrogen oxides emission reduction in Europe. *Economica*, 64:281–301.
- Nelson, G. C. and Hellerstein, D. (1997). Do roads cause deforestation? using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*, 79:80–88.
- Openshaw, S. and Taylor, P. (1979). A million or so correlation coefficients. In Wrigley, N., editor, *Statistical Methods in the Spatial Sciences*, pages 127–144. Pion, London.
- Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126.
- Paelinck, J. and Klaassen, L. (1979). *Spatial Econometrics*. Saxon House, Farnborough.
- Pinkse, J. (2002). Moran-flavored tests with nuisance parameter. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg. Forthcoming.
- Pinkse, J. and Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85:125–154.
- Pötscher, B. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models*. Springer-Verlag, Berlin.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.
- Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15:351–357.

- Roe, B., Irwin, E. G., and Sharp, J. S. (2002). Pigs in space: Modeling the spatial structure of hog production in traditional and nontraditional production regions. *American Journal of Agricultural Economics*, 84(2):259–278.
- Saavedra, L. A. (2000). A model of welfare competition with evidence from AFDC. *Journal of Urban Economics*, 47:248–279.
- Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87:108–119.
- Schwartz, S. (1994). The fallacy of ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health*, 84:819–823.
- Smirnov, O. and Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics and Data Analysis*, 35:301–319.
- Stein, M. L. (1999). *Interpolation of Spatial Data, Some Theory for Kriging*. Springer-Verlag, New York.
- Stoker, T. M. (1993). Empirical approaches to the problem of aggregation over individuals. *Journal of Economic Literature*, 33:1827–1874.
- Tanner, M. A. (1996). *Tools for Statistical Inference*. Springer-Verlag, New York, NY.
- Tobler, W. (1979). Cellular geography. In Gale, S. and Olsson, G., editors, *Philosophy in Geography*, pages 379–386. Reidel, Dordrecht.
- Upton, G. J. and Fingleton, B. (1985). *Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data*. Wiley, New York.
- Vijverberg, W. P. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics*, 76:281–307.
- Waller, L., Carlin, B., and Xia, H. (1997a). Structuring correlation within hierarchical spatio-temporal models for disease rates. In Grégoire, T., Brillinger, D., Russek-Cohen, P., Warren, W., and Wolfinger, R., editors, *Modeling Longitudinal and Spatially Correlated Data*, pages 309–319. Springer-Verlag, New York.
- Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997b). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92:607–617.
- Weiss, M. D. (1996). Precision farming and spatial econometric analysis: Research challenges and opportunities. *American Journal of Agricultural Economics*, 78:1275–1280.
- Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.

- Wikle, C. K. and Royle, J. A. (1999). Space-time dynamic design of environmental monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(4):489–507.
- Wolpert, R. L. and Ickstadt, K. (1998). Poisson/Gamma random field models for spatial statistics. *Biometrika*, 85:251–267.