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

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.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: 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; 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, 2002). 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 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, 2002) expresses how the magnitude of a decision variable for an economic agent depends on the magnitudes of the decision variables set by other economic agents. This provides the theoretical basis for a so-called spatial lag model, or, mixed regressive, spatial autoregressive model (Anselin, 1988):

\[ y = \rho Wy + X\beta + \varepsilon, \]  

(1)

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

Brueckner (2002) 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 \(-i\) subscripts 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:

\[ U(y_i, y_{-i}; x_i^*), \]

(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:

\[ y_i = R(y_{-i}, x_i^*). \]

(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:

\[ y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon, \]

(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 Wy 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).

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 (2002). 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 \( \varepsilon \) 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, 2002; Anselin and Moreno,
2002). These do not seem to fit the strategic interaction framework.

Brueckner (2002) 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 individual agent’s objective function:

\[ U(y_i, s_i, x_i^t) \]

(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 \( (x_i^t, \text{for example, the type of crop grown on the field}) \), as well as on the decisions taken by the other agents (how much water they use):

\[ s_i = H(y_i, y_{-i}; x_i^t). \]

(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 (2002) 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, Chilès and Delfiner, 1999, Chapter 8). A similar problem is encountered in the interpretation of a spatial lag model as the expression of a spatial diffusion process.7 While diffusion 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:

\[ P_i = \sum_j f(d_{ij})z_j. \]

(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

---

5 Familiar examples are Case et al. (1993), Besley and Case (1995), Murdoch et al. (1997), and Bivand and Szymanski (1997, 2000).
6 Examples are Brueckner (1998) and Saavedra (2000), among others.
7 See, for example, the discussion in Baller et al. (2001) and Messner and Anselin (2002).
8 In point pattern analysis, this identification problem is referred to as the problem of true versus 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).
9 For 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.
are the negative exponential, \( f(d_{ij}) = e^{-\gamma d_{ij}} \), and the inverse distance function, \( f(d_{ij}) = \frac{1}{d_{ij}^{\gamma}} \).

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

\[
[W\mathbf{x}]_i = \sum_{j \neq i} w_{ij}x_j, \tag{8}
\]

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. 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

\[
y = \mathbf{X}\beta + W\mathbf{X}\gamma + \varepsilon, \tag{9}
\]

where zero restrictions can be imposed on specific elements of the parameter vectors \( \beta \) and \( \gamma \). 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 \( \gamma \).

### 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 \( 1 \times k \) row vector of explanatory variables \( x'_i \) and an error \( \varepsilon_i \):

\[
y^*_i = x'_i\beta + \varepsilon_i. \tag{10}
\]

A matching spatial lag model would then be

\[
y^* = \rho W y^* + X\beta + \varepsilon, \tag{11}
\]

or, equivalently:

\[
y^*_i = \rho \sum_{j \neq i} w_{ij}y^*_j + x'_i\beta + \varepsilon_i, \tag{12}
\]

where \( y^* \) is the full \( n \times 1 \) 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 \( Wy^* \) that is present in the actors’ objective functions, such as Eq. (2), but not the observed \( Wy \). 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:

\[ y^* = (I - \rho W)^{-1} X^T \beta + (I - \rho W)^{-1} \epsilon, \quad (13) \]

or

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

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

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:

\[ y^*_i = \sum_j a_{ij} x_j^\epsilon + u_i, \quad (15) \]

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

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

\[ \sum_j a_{ij} x_j^\epsilon + u_i > 0. \quad (16) \]

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

\[ \Pr[y_i = 1] = \Pr[u_i < g_i(X, W, \beta, \rho)], \quad (17) \]

where \( g_i = \sum_j a_{ij} x_j^\epsilon \) depends on \( X, W, \beta \) and \( \rho \). 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 \( \epsilon_i \), i.e. for a spatial probit model, the random vector \( u \) will be multivariate normal with a covariance matrix \( \text{Cov}[u] = [(I - \rho W)^{(1/2)}]^{-1} \). 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 \( \epsilon_i \) does not follow a normal distribution, the transformed multivariate random variable \((I - \rho W)^{-1} \epsilon\) 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

\[ \Pr[y_i = 1] = \Pr[\epsilon_i < \rho \sum_j w_{ij} y^*_j + x'_i \beta]. \quad (18) \]

This expression is not operational, since the \( y^*_i \) that enter in the inequality condition are not observed, and are themselves determined by \( y^*_j \).

This simultaneous model contrasts with a conditional approach, in which either \( y_j \) or \( \Pr[y_j = 1] \) 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 expectation, maximization (EM) approach (McMillen, 1992), general method of moments (GMM) estimation (Pinkse and Slade, 1998), and simulation estimators, such as recursive importance sampling (Vijverberg, 1997; Beron

\(^{11}\) For 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.
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 characteristic 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 (β), while taking into account the spatial structure incorporated in the error covariance matrix. Formally, the main interest is in the familiar regression model:

\[ y = X\beta + \varepsilon, \]  

where the error covariance matrix, Cov[εε′], or, equivalently, E[εε′], specifies spatial covariance when the off-diagonal elements are non-zero, E[ε_iε_j] ≠ 0 (for i ≠ 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 ρ = 1 is outside the proper parameter space. For general, not row-standardized weights the parameter space is typically constrained to values much smaller than 1, so that (unscaled) first differencing is similarly not allowed.

More formally, a spatial first difference can be expressed as:

\[ y - W y = (X - WX)\beta + u, \]

or,

\[ (I - W)y = (I - W)X\beta + u. \]

Since the row elements of W sum to 1, the matrix (I - W) is singular. Instead of using a “pure” first difference, one can use a spatial first difference, which is given by:

\[ y - W y = (X - WX)\beta + u, \]

or,

\[ (I - W)y = (I - W)X\beta + u. \]

This 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. (1997a,b), Best et al. (1999), Lawson (2001), MacNab and Dean (2002), among others.

\[ 14 \text{ For a review of the econometric issues, see Lancaster (2000).} \]

\[ 15 \text{ For some early examples, see Martin (1974) and Getis (1995).} \]
difference, a spatial autoregressive parameter must be included, as in

\[(I - \rho W)y = (I - \rho W)X\beta + u. \tag{22}\]

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 pre-multiplying both sides of the equation by \((I - \rho W)^{-1}\) yields

\[y = X\beta + (I - \rho W)^{-1}u, \tag{23}\]

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

\[\varepsilon = (I - \rho W)^{-1}u, \tag{24}\]

\[(I - \rho W)\varepsilon = u, \tag{25}\]

\[\varepsilon = \rho W\varepsilon + u. \tag{26}\]

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 (12), 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

\[(I - \rho W)y = X\beta + u. \tag{27}\]

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 \(\beta\). However, as in the spatial error model, it is not possible to estimate the parameter \(\rho\) 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 observation of the locations of points, 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. 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).

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).


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

\(^{18}\) Note 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).
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 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.

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

### Table 1

<table>
<thead>
<tr>
<th>Implications of data models</th>
<th>Object</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIS</td>
<td>Vector</td>
<td>Raster</td>
</tr>
<tr>
<td>Spatial data</td>
<td>Points, lines, polygons</td>
<td>Surfaces</td>
</tr>
<tr>
<td>Location</td>
<td>Discrete</td>
<td>Continuous</td>
</tr>
<tr>
<td>Observations</td>
<td>Process realization</td>
<td>Sample</td>
</tr>
<tr>
<td>Spatial arrangement</td>
<td>Spatial weights</td>
<td>Distance function</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Lattice</td>
<td>Geostatistics</td>
</tr>
<tr>
<td>Prediction</td>
<td>Extrapolation</td>
<td>Interpolation</td>
</tr>
<tr>
<td>Models</td>
<td>Lag and error</td>
<td>Error</td>
</tr>
<tr>
<td>Asymptotics</td>
<td>Expanding domain</td>
<td>Infill</td>
</tr>
</tbody>
</table>

#### 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 \( \sigma_i^2 \) as well as \( n(n - 1)/2 \) off-diagonal terms \( \sigma_{ij} (= \sigma_{ji}) \) 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 do not 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 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 multi-level modeling, where one error component is associated 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 modeling 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 an \( n \times n \) 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 Fig. 1, but this can be easily generalized to any network structure. For example, in Fig. 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 Figs. 1 and 2 yield the same \( 6 \times 6 \) 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 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

---

20 See Anselin and Moreno (2002) for a review of some technical issues associated with specifying and testing for spatial error components.

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

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

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 files.

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 US 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 modeling 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.

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

---

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

25 One 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).
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 \( n_g \to \infty \). 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 \( w_{ij} = 1/d_{ij}^p \), 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, \( b_{ij} \neq b_{ji} \)). In practice, these parameters are often set at a priori, for example yielding a “gravity” like model with \( w_{ij} = 1/d_{ij}^2 \).

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 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 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

\[ w_{ij} = \sqrt{N_i N_j / d_{ij}}, \]

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, \( \delta \) 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.

\[ w_{ij} = 1/d_{ij}^2 \]

Other 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, \( \delta \) 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.
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 Pôtscher 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 → ∞). These approaches match the two different data models considered in the paper. 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


28 In 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. (1996), Best et al. (1999), Berger et al. (2001), and Damian et al. (2001), among others.

29 However, 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.
5.1. Ecological fallacy

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. 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 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). 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.

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 $\bar{x}_k = \frac{\sum_i x_{ik}}{n_k}$ (with $n_k$ as the group size) is the group average for that characteristic (e.g. county average income). In the literature, $\beta$ is referred to as the individual effect and $\gamma$ as the contextual effect. The corresponding macro-regression relates the group averages to each other, or

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

with the group averages as $\bar{y}_k = \frac{\sum_i y_{ik}}{n_k}$. 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

30 Note 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.

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

32 In 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).

33 For a related discussion of identification issues in economic models of interaction, see Manski (1993).
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 $\beta$ and $\gamma$ 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 ($\beta = 0$).\footnote{See Greenland (2002) and also Stoker (1993) for a more elaborate discussion.}

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:

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

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 $g = 1, \ldots, G$, would be:

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

with $w_{kg}$ as the elements of a group-level spatial weights matrix of dimension $G \times G$ 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 accomplished by means of an $n \times G$ aggregation matrix $H$, with elements $h_{ig} = 1$ for $i \in g$ and zero otherwise, such that

$$ W^G = H'W^n H,$$

with $W^n$ as the individual-level weights. This yields a $G \times G$ matrix with elements $w^n_{kg}$ obtained as the sums of all the weights $w^n_{ij}$ 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 $w^n_{ij} \neq 0$ 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,

$$ y_{ik} = \sum_{j=1}^{n} a_{ij}(\alpha + x_{jh}\beta + \bar{x}_{jh}\gamma) + \epsilon,$$
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 \( \tilde{y}_k \) 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

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.

Acknowledgements

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.

References


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


Conley, T.G., 1999. GMM estimation with cross-sectional dependence. J. Econ. 92, 1–45.


