Chapter 29: Spatial Econometrics

Luc Anselin
University of Illinois, Urbana-Champaign
anselin@uiuc.edu

May 20, 2005

Contents

1 Introduction **2**

2 Theoretical Motivations **4**
2.1 Social and Spatial Interaction 5
2.2 Measurement Error and Common Shocks 7
2.3 Identification Issues 8

3 Specification of Spatial Error Dependence **9**
3.1 Spatial Stochastic Processes 9
3.1.1 Spatial Weights 10
3.1.2 Spatial Autoregressive Process – SAR 11
3.1.3 Conditional Autoregressive Process – CAR 12
3.1.4 Spatial Moving Average Process – SMA 13
3.2 Direct Representation 14
3.3 Nonparametric Distance Models 15
3.4 Error Component Models 16

4 Specification of Spatial Lag Dependence **17**
4.1 Mixed Regressive Spatial Autoregressive Model 17
4.1.1 Spatial Multipliers 18
4.2 Higher Order Models 19

5 Estimation **21**
5.1 Asymptotics in Space 22
5.2 Maximum Likelihood Estimation 23
5.2.1 Spatial Lag Model 23
5.2.2 Spatial Error Models 25
5.2.3 Computational Issues 26
5.3 Method of Moments/Instrumental Variables 28
5.3.1 Spatial Two Stage Least Squares 28
5.3.2 Spatial Correlation as a Nuisance Parameter 30

5.3.3 Spatial Lag with Spatial SAR Errors 30
5.3.4 Heteroskedastic and Autocorrelation Consistent Estimators 32

6 Specification Testing 33
6.1 Tests Against Spatial Autocorrelation 33
6.1.1 Moran's I Test 33
6.1.2 Kelejian-Robinson Test 35
6.2 Tests Based on Maximum Likelihood 36
6.2.1 Spatial Error Autocorrelation 37
6.2.2 Spatial Lag 39
6.2.3 Higher Order Models 39
6.2.4 Robust Tests 40
6.3 Tests Against Multiple Sources of Misspecification 41
6.4 Specification Search 42

7 Spatial Effects in Panel Data 43
7.1 Spatial Effects in Pooled Cross-Section and Time Series Models . 44
7.1.1 Model Specification 44
7.1.2 Estimation 45
7.1.3 Specification Testing 46
7.2 Spatial Effects in Error Component Models 47
7.3 Spatial Seemingly Unrelated Regression Models (SUR) 50

8 Spatial Effects in Models with Limited Dependent Variables 52
8.1 Spatial Probit 53
8.2 Estimation of Spatial Probit 55
8.3 Specification Tests in Spatial Probit 58

9 Future Directions 60

1 Introduction

Spatial econometrics consists of a subset of econometric methods that is concerned with spatial aspects present in cross-sectional and space-time observations. Variables related to location, distance and arrangement (topology) are treated explicitly in model specification, estimation, diagnostic checking and prediction. More specifically, spatial econometrics deals with two basic forms of spatial effects in regression models, categorized as *spatial dependence* and *spatial heterogeneity* (Anselin 1988b).

Spatial dependence is a special case of cross-sectional dependence, in the sense that the *structure* of the correlation or covariance between observations at different locations is derived from a specific ordering, determined by the relative position (distance, spatial arrangement) of the observations in geographic space (or, in general, in network space). While similar to correlation in the time domain, the distinct nature of spatial dependence requires a specialized set of

techniques, that are *not* a straightforward extension of time series methods to two dimensions (Anselin 1988b).

Spatial heterogeneity is a special case of observed or unobserved heterogeneity, a familiar problem in standard econometrics. In contrast to spatial dependence, tackling this issue does not require a separate set of methods. The only spatial aspect of the heterogeneity is the additional information that may be provided by spatial structure. For example, this may inform models for heteroskedasticity, spatially varying coefficients, random coefficients and spatial structural change (spatial regimes). This will not be further considered here.[1]

The designation of a separate field of "spatial econometrics" was originally suggested by the Belgian economist Jean Paelinck in the early 1970s, to refer to methodological aspects associated with incorporating dependence in cross-sectional multiregional econometric models. An early overview is contained in the 1979 *Spatial Econometrics* book by Paelinck and Klaassen (Paelinck and Klaassen 1979).[2] In it, they outlined five important characteristics of the field: (i) the role of spatial interdependence; (ii) the asymmetry in spatial relations; (iii) the importance of explanatory factors located in other spaces; (iv) differentiation between ex post and ex ante interaction; and (v) explicit modeling of space. In Anselin (1988b), these ideas were elaborated upon and situated more formally within econometric estimation and specification testing.

Initially, the development and application of spatial econometric methods was largely confined to the applied fields of economics (e.g., urban and regional economics, environmental economics, real estate economics), regional science and quantitative economic geography. Early reviews of the state of the art can be found in, among others, Cliff and Ord (1973, 1981), Hordijk (1974), Anselin (1980, 1988b, 1992), Upton and Fingleton (1985), Haining (1990), and Anselin and Florax (1995a). This early work was heavily influenced by the statistical literature, where the interest in spatial aspects dates back to the early 20th century, dealing with concerns about the presence of spatial autocorrelation in the design of agricultural experiments (see Cressie 1993, p. 7). In spatial statistics, fundamental results regarding the estimation of spatial models were obtained by Whittle (1954), Besag (1974), Ord (1975), and Ripley (1981), among others (for more in-depth overviews, see Cressie 1993, Waller and Gotway 2004).

Arguably, the distinction between spatial statistics and spatial econometrics is subtle. As suggested in Anselin (1988b), the main difference lies in the emphasis on the economic models (and the theory behind them) as the basis for econometric specification in spatial econometrics, whereas the data tend to be more central in spatial statistics. A great degree of cross fertilization has occurred, although distinct emphases remain, especially in terms of a greater prominence of a Bayesian (hierarchical) perspective in modern spatial statistics (see, e.g., Waller et al. 1997a,b, Wikle et al. 1998, Royle and Berliner 1999, Banerjee et al. 2004, and the references cited therein).

[1] For overviews of the salient issues, see, e.g., Anselin (1988b, 1990), Jones and Casetti (1992), Casetti (1997), Fotheringham et al. (2002), as well as in the statistical literature, Gelfand et al. (2003) and Gamerman et al. (2003).

[2] See also Hordijk and Paelinck (1976), Hordijk (1979).

The initially somewhat marginal status of spatial econometrics relative to other subfields of econometrics has changed dramatically. Spurred both by advances in theory (social and spatial interaction) and in technology (geographic information systems), interest in spatial modeling has grown considerably in economics, as evidenced by a large number of both applied and theoretical papers that have appeared in recent years (for extensive literature reviews, see Florax and van der Vlist 2003, Anselin et al. 2004b). Several special journal issues have been devoted to the topic, including Nelson (2002), Anselin (2003a), Florax and van der Vlist (2003), and LeSage et al. (2004), as well as a number of edited volumes, Anselin et al. (2004a), Getis et al. (2004), and LeSage and Pace (2004).

This chapter reviews the basic methodology as well as recent advances in the specification, estimation and testing of spatial econometric models. It borrows from, builds upon and extends earlier reviews presented in Anselin and Bera (1998) and Anselin (2001a,b, 2002). In the next three sections, the basic specification problem of spatial econometrics is outlined, first motivated in terms of theory, then with regard to models of spatial error dependence and spatial lag dependence. This is followed by an overview of estimation and specification testing, presented in the context of the linear regression model. Two specialized sections deal with spatial effects in, respectively, panel data settings and models with limited dependent variables. The chapter closes with some comments on promising future directions.

2 Theoretical Motivations

In general, spatial regression specifications fall into two broad categories, referred to in the literature as *spatial lag* and *spatial error* models (Anselin 1988b). Before considering estimation and specification testing, in this section, a few comments are formulated on the theoretical foundations for such specifications.

To focus the discussion, consider the regression model:

$$y_i = g(y_J, \theta) + x_i'\beta + \varepsilon_i, \tag{1}$$

where y_i is an observation on the dependent variable at location i, $i = 1, \ldots, n$, $g(y_J, \theta)$ is a function of the values of the dependent variable observed at *neighboring* locations $j \in J$, with $j \neq i$ and with the set J of neighbors to be specified, x_i is a $k \times 1$ vector of observations on explanatory (exogenous) variables, ε_i is an error term, and θ and β are vectors of parameters. The function g of observations on the dependent variable at locations other than i is referred to as a *spatially lagged dependent variable*, or spatial lag term. It is typically formulated as a weighted average of neighboring values, where the neighbors are specified through the use of a so-called spatial weights matrix.[3] Specifically, a mixed regressive, spatial autoregressive model (Anselin 1988b) takes the form

[3] A more extensive treatment of spatial weights is postponed until Section 3.1.1.

(in matrix notation):

$$y = \rho W y + X\beta + \varepsilon, \tag{2}$$

where y is a $n \times 1$ vector of observations on the dependent variable, W is a $n \times n$ spatial weights matrix, which specifies the neighbors used in the averaging (resulting in the spatial lag term, Wy), ρ is a spatial autoregressive parameter, X a $n \times k$ matrix of observations on the explanatory variables, β a matching vector or parameters, and ε a $n \times 1$ vector of error terms.

The inclusion of the terms $g(y_J, \theta)$ or Wy on the right hand side of the equation is motivated by theory as the equilibrium outcome of processes of social and spatial interaction, considered in Section 2.1.

The case of spatial error autocorrelation is a special form of a non-spherical error variance-covariance matrix. Specifically,

$$\mathrm{E}[\varepsilon_i \varepsilon_j] = \sigma_{ij} \neq 0, \tag{3}$$

for $j \neq i$, yielding a general variance-covariance matrix $\mathrm{E}[\varepsilon\varepsilon'] = \Sigma$. The non-zero off-diagonal elements correspond to a notion of *spatial* covariance when they follow a given spatial structure or "spatial ordering" (Kelejian and Robinson 1992). The spatial ordering specifies for which pairs of locations i, j (with $i \neq j$) the covariance will be non-zero. The ordering is based on concepts such as contiguity (neighbors in space) or on the use of a distance metric. The consideration of spatially correlated disturbance terms is motivated by measurement error or by the presence of common shocks, reviewed briefly in Section 2.2.

A short overview of identification issues is presented in Section 2.3

2.1 Social and Spatial Interaction

Theoretical economic models of *interacting agents* and *social interaction* move away from the traditional notion of atomistic agents and instead formalize how interaction among the agents can lead to collective behavior and aggregate patterns. These models have recently received considerable attention, as evidenced in the development of theoretical frameworks to explain social phenomena such as peer effects, neighborhood effects, spatial spillovers, network effects, and the like (for examples and reviews, see, among others, Glaeser et al. 1992, 1996, Akerlof 1997, Durlauf 1997, 2004, Brock and Durlauf 2001a,b, Manski 2000, Conley and Topa 2002). In his recent review, Manski (2000) suggests three specific channels of interaction from an economic perspective: constraint interactions, expectations interactions and preference interactions. This leads to specifications where the actions of other agents are included in the right hand side of a model for individual behavior. In the review of Brock and Durlauf (2001a), a further distinction is made between global interaction, specified as the average behavior of the population, and local interaction, specified as the average behavior of a reference group ("neighborhood"). The implications of these specifications for model identification and estimation are complex, and

treated at length in Manski (1993, 1995), and Brock and Durlauf (2001a,b), among others.

In the spatial econometric literature, the spatial lag model is commonly conceptualized as representing the empirical counterpart to the equilibrium solution of strategic interaction, or, a *spatial reaction function* (Brueckner 1998, 2003). The reaction function of Brueckner (2003) takes the form:

$$y_i = R(y_{-i}, x_i), \tag{4}$$

where y_i stands for the level of decision variable y for decision maker i, y_{-i} reflects a function of the decision variables chosen by the other decision makers, and x_i is a vector of exogenous characteristics of i.[4] A linear functional form for R readily yields a specification such as equation (1) or (2).

Brueckner (2003) demonstrates how the reaction function (4) can be obtained through two different behavioral mechanisms for strategic interaction. In one, termed *spillover*, the level of the decision variable chosen by the other agents (y_{-i}) enters directly into the utility (objective function) of each individual agent. This yields a utility function such as:

$$U(y_i, y_{-i}; x_i). \tag{5}$$

This model can be extended to apply to a representative agent for an aggregate spatial unit, such as a state, e.g., in the widely cited study of spillovers among state expenditures in Case et al. (1993). Other examples in the literature apply this framework as the motivation for the use of a spatial lag specification in studies of the demand for public goods, pollution abatement and other forms of yardstick competition (e.g., Besley and Case 1995, Murdoch et al. 1993, 1997, Bivand and Szymanski 1997).

A second behavioral mechanism is termed *resource flow* by Brueckner (2003). In this approach, the agent's decision variable is indirectly affected by the decisions of other agents, through their impact on an available resource. The utility function thus becomes:

$$U(y_i, s_i, x_i), \tag{6}$$

where s_i is the contribution of the resource used by the agent to the agent's utility. Interaction is induced through the effect of the actions of other agents on the resource use. Specifically, each agent's resource use is constrained by the actions taken by other agents, as in

$$s_i = H(y_i, y_{-i}; x_i). \tag{7}$$

After substituting this into the utility function, the same reaction function as in (4) obtains, again suggesting a spatial lag specification. Several tax competition and related strategic interaction models can be categorized as resource flow

[4]Note that both global or local interaction can be implemented in this framework, by specifying y_{-i} either as a population mean or as the mean in a subset of the population.

interactions (for further examples, see Brueckner 1998, 2003, Saavedra 2000, Brueckner and Saavedra 2001).

In these frameworks, the use of the spatial lag specification follows from formal economic reasoning. However, the model itself cannot distinguish between which mechanism led to the equilibrium under consideration. The two processes are observationally equivalent and cannot be identified empirically. This is one example of many identification problems associated with spatial econometric model specifications (see Section 2.3). In the statistical literature, it is referred to as the *inverse problem*, pervasive in the analysis of cross-sectional spatial data. It is also known as the problem of distinguishing *true* from *apparent* contagion, the latter being a case of spatial heterogeneity. Clusters in space may occur as the result of various mechanisms, including some that are not based on interaction or contagion. The information contained in the cross-sectional sample as such is insufficient to allow for the identification of the underlying mechanism.[5] This is a basic limitation of the cross-sectional setup, which is somewhat (but not completely) eliminated in spatial panel data settings.

2.2 Measurement Error and Common Shocks

Unlike spatial lag models, spatial error specifications are typically not motivated by a theoretical economic model, but instead are formulated to deal with data problems. In other words, it is the cross-sectional nature of the data that causes correlation problems, not necessarily the "spatial" nature of the model. Therefore, error autocorrelation is more relevant in practice than lag correlation, since it pertains to both spatial and non-spatial models.

In many instances in applied econometrics using cross-sectional data, there is a mismatch between the spatial scale of the process under study (e.g., a housing market or labor market) and the spatial unit of observation (e.g., a census tract or county). As a result, measurement errors are likely to systematically vary across space (Anselin 1988b, p. 12). In addition, the data integration (combination of observations from different spatial scales) and spatial interpolation that is facilitated by modern geographic information systems will tend to result in patterns of spatial correlation (Anselin 2001c). The statistical issues related to the joint consideration of observations at different spatial scales is referred to as the "change of support problem" (COSP). The incorporation of these complexities into spatial econometric methodology is still limited to date (for an overview of the statistical issues, see Gotway and Young 2002).

A theoretical framework for spatially correlated errors is provided by the general approach based on "common shocks" outlined by Andrews (2003). This focuses on the properties of estimators in a regression of a dependent variable $Y(\gamma)$ on explanatory variables $X(\gamma)$, where γ is a spatial sample from an arbitrary topological space Γ. Y and X are part of a random vector $W(\gamma) = (Y(\gamma), X(\gamma), S(\gamma))$ that contains observed and unobserved terms $S(\gamma)$ that are

[5] See, e.g., Johnson and Kotz (1969), Chapter 9, for a formal discussion of contagious distributions.

common to some units in the population. Common shocks are defined as random variables $C(\gamma)$ measurable on a σ-field such that, conditional on this field, a spatial sample $W_i : i = 1, \ldots$ is i.i.d. This general framework includes a wide range of forms of spatial correlation, as well as group effects and factor structures.[6] Factor structures are increasingly used in recently formulated models to deal with general forms of cross-sectional correlation in panel data settings (see Section 3.4).

2.3 Identification Issues

The degree of simultaneity and endogeneity inherent in spatial correlation creates a number of difficult identification problems. The best known of these is arguably the "reflection problem" outlined in Manski (1993), where it is shown that the parameters in models of social/spatial interaction are only identified under strict conditions. In Manski's model, three types of social interaction are considered, referred to as endogenous effects (interaction effects among individual agents), contextual effects (exogenous group characteristics) and correlated effects (observed or unobserved characteristics that agents have in common).

A general linear model for social interaction is then (using Manski's notation):

$$y = \alpha + \beta \ \mathrm{E}(y|x) + \mathrm{E}(z|x)'\gamma + z'\eta + u, \tag{8}$$

where y is the dependent variable (e.g., participation in gangs), x are exogenous group characteristics (e.g., census tract median income), and z and u are observed and unobserved variables that directly affect y (e.g., age of the individual as observed, and "culture" as unobserved). The endogenous effect is specified in terms of the group mean, $\mathrm{E}(y|x)$. The unobserved characteristics are further assumed to be correlated across individuals in the group, with $\mathrm{E}(u|x, z) = x'\delta$. The resulting conditional expectation, or regression, is then:

$$\mathrm{E}(y|x, z) = \alpha + \beta \ \mathrm{E}(y|x) + \mathrm{E}(z|x)'\gamma + +x'\delta + z'\eta. \tag{9}$$

Endogenous effects are present when $\beta \neq 0$, contextual effects when $\gamma \neq 0$, and correlated effects when $\delta \neq 0$. The reduced form of this regression yields:

$$\mathrm{E}(y|x) = \alpha/(1-\beta) + \mathrm{E}(z|x)'(\gamma + \eta)/(1-\beta) + x'\delta/(1-\beta), \tag{10}$$

which demonstrates how, without further restrictions, the different social effects cannot be separately identified. Manski (1993) outlines the types of constraints and instruments needed to obtain identification.[7]

In the spatial lag model, the reflection problem is avoided with a judicious specification of the spatial weights matrix. An important distinction between the endogenous effect $\beta \ \mathrm{E}(y|x)$ in Manski's model and the spatially lagged effect $\rho W y$ is that the latter only pertains to a subset of the sample, not to the

[6] For technical details, see Andrews (2003).

[7] See also Manski (1995, 2000) and Brock and Durlauf (2001a,b) for technical details.

population conditional expectation (see also Irwin 2004, for further discussion). Complications remain when both endogenous and correlated effects are present. An alternative approach to separately identify endogenous effects in the presence of correlated effects consists of a natural experiment design (Irwin 2004).

While much of the discussion of social interaction is couched in terms of models for individual agents, in practice estimation is often carried out for aggregate spatial units of observation (e.g., using the device of a representative agent). This can lead to an ecological fallacy or cross-level bias problem, where aggregate results are interpreted as if they pertained to individual behavior (for reviews of this issue, see, e.g., Stoker 1993, King 1997, Greenland 2002). When both individual characteristics as well as group effects are contained in a regression specification, estimation at the group level does not allow for the separate identification of individual and contextual effects (Greenland 2002). Moreover, the coefficients in a spatial lag specification at the individual level are not estimated consistently by the coefficients in a spatially aggregate counterpart, as shown in Anselin (2002). Consequently, considerable care is needed in the specification as well as the interpretation of models for spatial and social interaction.

3 Specification of Spatial Error Dependence

In the presence of spatial correlation, the error variance-covariance matrix $\mathrm{E}[\varepsilon\varepsilon'] = \Sigma$ contains n variance terms and $n \times (n-1)/2$ off-diagonal parameters. These cannot be separately estimated in a single cross-section with n observations. A fundamental specification problem in spatial econometrics is the need to impose constraints or structure on the elements of Σ such that estimation becomes possible. Several approaches have been suggested. They can be categorized as spatial stochastic process models, direct representation, nonparametric distance models, and error component models. They are considered in turn.

3.1 Spatial Stochastic Processes

A spatial stochastic process, or spatial random field, is a collection of random variables, indexed by location. Using regression error terms (ε_i) as an example, this can be formally expressed as

$$\{\varepsilon_i, \in D\}, \tag{11}$$

where D is a spatial index set, either a continuous surface or a finite set of discrete locations. The specification of the spatial process for the regression error terms gives rise to a particular covariance structure, or pattern of spatial autocorrelation. A crucial element in the specification is the determination of the neighbor structure, or, formally, which locations will be represented in the spatial lag term on the right hand side of the equation. The connectedness structure is expressed by means of a spatial weights matrix, which is considered first. This is followed by a brief review of commonly used process specifications.

3.1.1 Spatial Weights

The spatial weights matrix ia a $n \times n$ positive matrix, W, through which the "neighborhood set" is specified for each observation. A location (observation) appears both as row and column, with non-zero matrix elements w_{ij} indicating a neighbor relation between observation (row) i and (column) j. By convention, self-neighbors are excluded, such that the diagonal elements $w_{ii} = 0$. Also, the weights matrix is often used in row-standardized form, with weights $w^s_{ij} = w_{ij}/\sum_j w_{ij}$.[8] The row-standardization facilitates the interpretation of the weights as constructing an average of the neighboring values in the so-called *spatial lag operator*, $\sum_j w_{ij} z_j$ (Anselin 1988b, pp. 22–24). Since the traditional notion of a spatial shift is non-operational in irregular spatial layouts, this averaging is used instead (see Anselin and Bera 1998, pp. 245–246, for a more in-depth discussion and further references). In matrix notation, the spatial lag of a random vector z is then Wz. In spatial econometric regression models, a range of specifications can be obtained by applying the spatial lag operator to the dependent variable, Wy, the explanatory variables, WX, or the error term, $W\varepsilon$.

There is very little formal guidance in the choice of the "correct" spatial weights for a given model specification. Typically, the definition of neighbor is based on geographic criteria, such as polygons having a common boundary (contiguity) or points being within a critical distance band. Other geographic criteria include combinations of contiguity and border length, or k-nearest neighbors. The literature on the specification of spatial weights is extensive, with early discussions in Cliff and Ord (1973, 1981), Upton and Fingleton (1985), Anselin (1988b, Chapter 3), and recent technical reviews in Dietz (2002), Leenders (2002) and Anselin (2002, pp. 256–260).

Spatial weights need not be limited to geographic considerations, and extensions to abstract spaces have been suggested, such as those arising from social networks or economic distance (e.g., Case et al. 1993, Conley and Ligon 2002, Conley and Topa 2002). For example, the choice of a suitable metric can yield general measures of distance (d_{ij}), which are then converted into inverse distance spatial weights, as $w_{ij} = 1/d_{ij}$. More complex functional forms have been suggested as well, including negative exponential and gravity-like measures that combine a "mass" measure at origin and destination with a distance decay effect (e.g., Anselin 1988b, Ferrándiz et al. 1995, 1999)

An interesting form of "economic" weights follows when a hierarchical or group structure is suggested by the data, as in Case (1991, 1992). All observations belonging to the same group (e.g., a region) are considered to be neighbors, but the neighbor relation does not extend across groups. This results in a block-diagonal weights matrix with each block consisting of weights $1/(n_g - 1)$ (with n_g as the number of observations belonging to the group). This structure has different properties from the distance or contiguity based weights, with important implications for estimation (see Lee 2002, Kelejian and Prucha 2002).

[8] In what follows, the notation w_{ij} will be used to denote row-standardized weights.

In order to obtain suitable asymptotic properties for estimators and specification tests, the weights matrix is subject to regularity conditions. These regularity conditions boil down to bounds on the weights and sums of the weights. Specifically, as pointed out in the work of Kelejian and Prucha (e.g., Kelejian and Prucha 1999, Assumption 3, p. 515), the sums $\sum_{i=1}^{n} |w_{ij}|$ and $\sum_{j=1}^{n} |w_{ij}|$ must be bounded by $c_w < \infty$. The bounds are related to summability conditions typically required to ensure that the error variance-covariance matrix is proper (see also Mandy and Martins-Filho 1994). Spatial weights based on notions of contiguity easily satisfy these regularity conditions. For more complex weights, such as those derived from distance bands in economic space, careful consideration of the regularity conditions is required, especially when the weights also contain parameters.

3.1.2 Spatial Autoregressive Process – SAR

One of the earliest and simplest specifications considered for a spatial process is the simultaneous spatial autoregressive model (SAR), suggested by Whittle (1954). This specification applies to the special case where the data are located on a regular rectangular lattice. The right hand side of the equation is simply a weighted sum of the North, South, East and West neighbors in the lattice, each assigned a separate parameter.

Using ε for the regression error term, and with the locations on the rectangular lattice labeled by row (r) and column (c), this model is:

$$\varepsilon_{r,c} = \xi_1 \varepsilon_{r-1,c} + \xi_2 \varepsilon_{r+1,c} + \xi_3 \varepsilon_{r,c-1} + \xi_4 \varepsilon_{r,c+1} + u_{r,c}, \tag{12}$$

with $\xi_k,\ k = 1, \ldots, 4$, as the spatial autoregressive parameters and u as an idiosyncratic error term.[9] In irregular lattice layouts (such as a cross-section of counties or city neighborhoods), this simple specification is not applicable, since the number of neighbors typically varies by observation and the concept of a spatial shift has no meaning. Instead, a spatial weights matrix is used, as in:

$$\varepsilon_i = \lambda \sum_j w_{ij} \varepsilon_j + u_i, \tag{13}$$

with λ as the autoregressive coefficient. Alternatively, in matrix notation, for the $n \times 1$ vector of error terms, ε:

$$\varepsilon = \lambda W \varepsilon + u. \tag{14}$$

Since

$$\varepsilon = (I - \lambda W)^{-1} u, \tag{15}$$

the error variance-covariance matrix follows as

$$\mathrm{E}[\varepsilon \varepsilon'] = (I - \lambda W)^{-1} \mathrm{E}[u u'] (I - \lambda W')^{-1}. \tag{16}$$

[9] See also Cressie (1993, p. 405–7, 456–7) for further details.

Under the standard assumption of i.i.d. errors u, with $\mathrm{E}[uu'] = \sigma^2 I$, this expression simplifies to (after rearranging terms):

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2[(I - \lambda W)'(I - \lambda W)]^{-1}. \tag{17}$$

In the more general case where $\mathrm{E}[uu'] = \sigma^2\Omega$, e.g., to allow heteroskedastic errors, the full expression is:

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2(I - \lambda W)^{-1}\Omega(I - \lambda W')^{-1}. \tag{18}$$

Note that in practice, the weights matrix W is typically not symmetric (e.g., after row-standardization, or using k-nearest neighbors), so that $(I - \lambda W)^{-1} \neq (I - \lambda W')^{-1}$. In order to get better insight into the structure of the covariance matrix induced by this process, consider the case where W corresponds to row-standardized first order contiguity. With $|\lambda| < 1$, the following expansion holds (also known as the Leontief expansion):

$$(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \ldots, \tag{19}$$

and a similar expression with W'. The product of the two inverse terms then yields a sum of cross products of the form

$$\begin{aligned} &I + \lambda(W + W') + \lambda^2(WW + WW' + W'W') \\ &\quad + \lambda^3(WWW + WWW' + WW'W') + \ldots \end{aligned} \tag{20}$$

The resulting covariance is quite different from the familiar band structure for first order autoregressive processes in time. It is qualitatively similar, in that the strength of the covariance decreases (higher powers of λ) for higher orders of contiguity (powers of W), but the pattern is considerably more complex (due to the products of W and W'). Also, it is important to note that even though the W in the SAR process may only pertain to the first order neighbors, the resulting spatial covariance reaches far beyond those first order neighbors, in fact inducing *global* spatial autocorrelation.

Most importantly, when the number of neighbors (the nonzero elements in W) across locations is not constant (the typical case in practice), the diagonal elements in (20) are no longer constant, inducing heteroskedasticity. This heteroskedasticity is present even when u is i.i.d., greatly complicating estimation and specification testing.[10]

3.1.3 Conditional Autoregressive Process – CAR

The probabilistic model behind the SAR specification is based on the joint distribution of a random vector, such as ε. An alternative approach focuses on the conditional distribution. While originally outlined in Besag (1974), and very little used in the spatial econometric literature, this perspective is prevalent in

[10] For further technical details, see Haining (1990, pp. 81–83), Cressie (1993, pp. 405–407), Anselin and Bera (1998, pp. 248–249), and Anselin (2003a, pp. 155–156), among others.

the recent literature on Bayesian hierarchical spatial modeling (e.g., Banerjee et al. 2004, pp. 76–84). It is used both as a model for spatial correlation itself, and as a spatial prior for parameters in a hierarchical setup.

In essence, the conditional approach consists of establishing conditions under which a density for the random variable at a given location, conditional on its neighbors, yields a proper joint density (Kaiser and Cressie 2000). The formal result is contained in the Hammersley-Clifford theorem (Cressie 1993, pp. 410–419).

Assuming normality, and ignoring complications from heteroskedasticiy, the conditional expectation at location i, conditional upon all the other locations in the system ($j \neq i$), is expressed the CAR model as a linear function of the neighbors:

$$\mathrm{E}[\varepsilon_i | \varepsilon_j,\ j \neq i] = \lambda \sum_j w_{ij} \varepsilon_j. \tag{21}$$

Under the proper constraints on the elements of W (e.g., Cliff and Ord 1981, p. 148, 179–183), this yields a joint multivariate normal distribution with variance

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2 (I - \lambda W)^{-1}. \tag{22}$$

Extension to other distributions have been suggested as well, such as auto-logistic and auto-Poisson, but these only obtain under very restrictive conditions (Banerjee et al. 2004, pp. 84-84).

Two important characteristics of the variance structure in (22) should be noted. First, it is required that the spatial weights matrix W be symmetric, which excludes the use of row-standardized weights (or k-nearest neighbor weights) and affects the parameter space for the autoregressive coefficient λ (typically making λ much smaller than for corresponding row-standardized weights). Second, the range and strength of spatial covariance induced by this model is much smaller than for a corresponding SAR model. This can be seen by comparing the expansion in (19), which corresponds to the CAR model, to that in (20) for the SAR model (see also Wall 2004)

3.1.4 Spatial Moving Average Process – SMA

A third type of spatial stochastic process used to model the structure of spatial correlation is based on a moving average specification (see Haining 1988, 1990, pp. 83–84).

For the individual error terms, this model is:

$$\varepsilon_i = \gamma \sum_j u_j + u_i, \tag{23}$$

where γ is the moving average parameter. In matrix notation, for the random vector ε, the corresponding expression is:

$$\varepsilon = \gamma W u + u. \tag{24}$$

Assuming i.i.d. errors u, the resulting error covariance matrix takes on the form:

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2[I + \gamma(W + W') + \gamma^2 WW']. \tag{25}$$

In contrast to the SAR and CAR specifications, this represents a *local* pattern of autocorrelation. Inspection of (25) reveals that the only non-zero off-diagonal elements correspond to the non-zero elements of W (W') and WW'. In the case of first order contiguity for W, the matrix product WW' contains non-zero entries for the first and second order neighbors only. As a result, there is no spatial covariance beyond the second neighbor, hence the local characteristic (Anselin 2003a, pp. 156–157).

As in the SAR model, a non-constant number of neighbors across observations will induce heteroskedasticity, even when u is i.i.d.

3.2 Direct Representation

In a direct representation, the covariance between each pair or error terms is specified as an inverse function of the distance between them. Formally,

$$\mathrm{E}[\varepsilon_i\varepsilon_j] = \sigma^2 f(d_{ij}, \phi), \tag{26}$$

with $\varepsilon_{i,j}$ as the regression error terms, σ^2 the error variance, d_{ij} the distance separating i and j, and f a distance decay function. The function f should be a distance *decay* function and ensure a positive definite covariance matrix. This requires $\partial f/\partial d < 0$ and $|f(d_{ij}, \phi)| \leq 1$, with $\phi \in \Phi$ as a $p \times 1$ vector of parameters on an open subset Φ of $\mathbb{R}^p$. This approach is closely related to variogram models used in geostatistics, and requires assumptions of stationarity and isotropy (see Cressie 1993, for an extensive review). The complete variance-covariance is then:

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2\Omega(d_{ij}, \phi). \tag{27}$$

Note that, in contrast to the SAR and SMA specifications, the spatial autocorrelation in this model does *not* induce heteroskedasticity, since the diagonal terms of Ω are constant.

The direct representation specification of spatial autocovariance dates back to early papers by Cook and Pocock (1983), Mardia and Marshall (1984), Warnes and Ripley (1987), and Mardia and Watkins (1989), among others. Its application in spatial econometrics has been primarily in studies of real estate markets, such as Dubin (1988, 1992) and Basu and Thibodeau (1998) (see also the review in Dubin et al. 1999).

The choice of the distance decay function is not arbitrary and the requirement of ensuring a positive definite covariance matrix imposes constraints on the functional form and the parameter space, as well as on the metric and scale used for the distance measure (for technical details, see, e.g., Mardia and Marshall 1984, pp. 138–139). The extent to which global rather than local autocorrelation is modeled depends on the slope of the distance decay function.

A commonly used specification is based on a negative exponential distance decay:

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma^2[I + \gamma\Psi], \tag{28}$$

with the off-diagonal elements of Ψ, $\Psi_{ij} = e^{-\phi d_{ij}}$, and γ as a non-negative scaling parameter. In order to facilitate interpretation and specification testing, the diagonal elements of Ψ are often set to zero (the variance is captured by the term $\sigma^2 I$). The distance metric and parameter space must be such that elements of $e^{-\phi d_{ij}}$ yield a valid spatial correlation matrix.[11]

3.3 Nonparametric Distance Models

An alternative to the parametric specification of spatial covariance as a function of a distance metric is a nonparametric approach. This is an extension to the spatial domain of the principle behind heteroskedastic and serially autocorrelation consistent covariance matrix estimation of Newey and West (1987) and Andrews (1991), among others.

As in the direct representation approach, the spatial covariance is a function of the distance separating two observations, but the functional form is left unspecified. For example, for the regression error terms:[12]

$$\mathrm{E}[\varepsilon_i\varepsilon_j] = f(d_{ij}), \tag{29}$$

where d_{ij} is a "proper" positive and symmetric distance metric (for regularity conditions on the distance metric, see Conley 1999, Kelejian and Prucha 2003).

This approach was first used in Conley (1999), and further elaborated upon in Conley and Ligon (2002) and Conley and Topa (2002). In its simplest form, the estimator for the familiar term $V = n^{-1}X'\Sigma X$ is:

$$\hat{V} = n^{-1} \sum_{d_{ij}<\delta} x_i x_j' \hat{\varepsilon}_i \hat{\varepsilon}_j, \tag{30}$$

where δ is a distance cutoff, x_i is a column vector of observations at i, and $\hat{\varepsilon}_i$ is a residual at i. This estimator follows essentially the same principle as in the time series domain by adding up sample spatial autocovariances. In order to ensure positive definiteness of the estimator, a kernel is applied to the cross-products. For example, in the recent paper by Kelejian and Prucha (2003), a general covariance matrix estimator takes on the form:[13]

$$\hat{V} = n^{-1} \sum_i \sum_j x_i x_j' \hat{\varepsilon}_i \hat{\varepsilon}_j K(d_{ij}/d), \tag{31}$$

[11] For further details, see Mardia and Marshall (1984, p. 141) and Anselin (2001a, pp. 128–130).

[12] Since $\mathrm{E}[\varepsilon] = 0$, the usual deviations from the mean can be ignored in this example.

[13] The estimator presented in Kelejian and Prucha (2003) is slightly more general than the illustration given here in that it pertains to instrumental variable estimation.

where $K()$ is a kernel function and d a suitable cutoff distance. Formal properties of the kernel estimator for a spatial autocovariance were also established in Hall and Patil (1994) (see also Conley and Ligon 2002, Conley and Topa 2002, for further elaboration and examples).

Note that this structure results in zero spatial autocovariance beyond the distance cut-off. The latter is bounded but allowed to grow with the sample size, potentially yielding more global patterns of spatial correlation in larger samples.[14]

3.4 Error Component Models

In panel data models, it is customary to assume the existence of an unobserved error component, shared by all cross-sectional units in a given time period. This gives rise to a particular form of spatial autocorrelation, where all error terms are equicorrelated (see Section 7.2). In a pure cross-section, such an approach is not feasible, but alternative concepts of error components are applicable.

In Kelejian and Robinson (1995), an error decomposition was proposed that combines a location-specific or local component with a regional or spillover component. Formally, the error term in this spatial error component process is:

$$\varepsilon = W\psi + \xi, \tag{32}$$

with ψ as an $n \times 1$ vector of errors that incorporate the spillover across neighbors, defined through a spatial weights matrix W, and ξ as a matching vector of location-specific disturbances. Each error component is i.i.d. with:

$$\begin{aligned} \mathrm{E}[\psi] &= \mathrm{E}[\xi] = 0, & (33)\\ \mathrm{E}[\psi\psi'] &= \sigma_\psi^2 I,\ \mathrm{E}[\xi\xi'] = \sigma_\xi^2 I, & (34)\\ \mathrm{E}[\psi_i \xi_j] &= 0, \forall\ i, j. & (35) \end{aligned}$$

The error variance-covariance matrix then becomes:

$$\mathrm{E}[\varepsilon\varepsilon'] = \sigma_\psi^2 WW' + \sigma_\xi^2 I, \tag{36}$$

with σ_ψ^2 as the spatial variance component, and σ_ξ^2 the remainder variance term. Since the matrix product WW' is positive definite, the error components variance-covariance matrix is positive definite with $\sigma_\psi^2 \geq 0$ and $\sigma_\xi^2 > 0$.

This particular specification is a model for *local* spatial correlation, as can be seen by comparing (36) to the terms in the SMA variance (25). The two models share the term in I and WW', but the spatial error components model does not have a term in W or W'. Using a first order contiguity W as an example, this

[14] In Conley (1999), where the setting is a rectangular grid, the critical distance in each direction is $o(n^{1/3})$ in the dimension of that direction (see assumption C1 in Conley 1999, p. 12). In Kelejian and Prucha (2003, Assumption 4a), the constraint is through the maximum number of neighbors within the critical distance, $l_n = o(n^{1/3})$. This implies that for a data set with $n = 100$, the distance criterion would result in no more than 4 spatially covarying neighbors allowed per observation, whereas with $n = 1000$, the number of spatially correlated neighbors increases to 10.

implies that the spatial error covariance only contains first order neighbors that are also contained in the non-zero terms in WW'. More importantly, there is no spatial correlation beyond the second order neighbors.

The general framework outlined in Andrews (2003) provides a way to conceive of a mechanism for a broad range of common shocks. This includes the familiar error components from the panel data literature, where the error term is decomposed into a term associated with the location/individual (i), a term associated with the time period (t) and an idiosyncratic error term. Recent extensions to this model have been offered in the literature on heterogeneous panels. The time component can be generalized and expressed in the form of an unobserved common effect or *factor* f_t to which all cross-sectional units are exposed. However, unlike the standard error component model, each cross-sectional unit has a distinct factor *loading* on this factor. The simplest form is the so-called one factor structure, where the error term at location i and time t is specified as:

$$\epsilon_{it} = \delta_i f_t + u_{it}, \tag{37}$$

with δ_i as the cross-sectional-specific loading on factor f_t, and u_{it} as an $i.i.d$ zero mean error term. Consequently, cross-sectional (spatial) covariance between the errors at i and j follows from the the inclusion of the common factor f_t in both error terms:

$$\mathrm{E}[\epsilon_{it}\epsilon_{jt}] = \delta_i \delta_j \sigma_f^2. \tag{38}$$

The common factor model has been extended to include multiple factors. In these specifications, a wide range of covariance structures can be expressed by including sufficient factors and through cross-sectional differences among the factor loadings (for further details, see Driscoll and Kraay 1998, Coakley et al. 2002, Pesaran 2002, Hsiao and Pesaran 2004).

4 Specification of Spatial Lag Dependence

Spatial lag models include terms Wy and/or WX on the right hand side of the equation. This leads to spatial multiplier effects, similar in nature to the social multiplier of Glaeser et al. (1996, 2002). In this section, these specifications are further explored, with particular attention to identification issues in higher order spatial models.

4.1 Mixed Regressive Spatial Autoregressive Model

The basic spatial lag specification is the mixed regressive, spatial autoregressive model, introduced in Section 2, equation (2):

$$y = \rho W y + X\beta + \varepsilon, \tag{39}$$

with the i.i.d. error vector ε. After some matrix algebra, the reduced form is obtained as:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon. \tag{40}$$

This is a nonlinear model in ρ and β, with an error term that follows a SAR process. Consequently, the error variance will be as in (17).

The reduced form also illustrates how the spatially lagged dependent variable in (39), Wy, is *endogenous*, violating the assumptions of the standard regression model:

$$Wy = W(I - \rho W)^{-1} X\beta + W(I - \rho W)^{-1}\varepsilon, \tag{41}$$

such that $\mathrm{E}[(Wy)'\varepsilon] \neq 0$ (see also Anselin 1988b, pp. 58–59).

It is worthwhile to further consider the similarity between the spatial lag model and the linear regression with a SAR error. Inserting the error equation (15) into the usual linear regression specification $y = X\beta + \varepsilon$ yields:

$$y = X\beta + (I - \lambda W)^{-1} u, \tag{42}$$

or, alternatively,

$$y = \lambda W y + X\beta - \lambda W X\beta + u. \tag{43}$$

This latter specification is referred to as a *spatial Durbin* model or spatial common factor model, in analogy to the time series case (Anselin 1980, 1988b, Burridge 1981).

Comparison of (39) and (43) demonstrates how the spatial error model is a special case of a spatial lag model, but with additional nonlinear constraints on the parameters (the common factor constraints). Alternatively, comparison of (40) and (42) shows how the spatial lag model is a special case of a spatial error model, that is nonlinear in the parameters. Also note how the two specifications are not nested. Setting either ρ or λ to zero in the respective models yields an ordinary linear regression equation, not one of the spatial models. The non-nested nature of the two specifications, together with their similarity, creates considerable problems in identification, estimation and diagnostic testing (e.g., Anselin 1988b, Kelejian and Prucha 1997).

4.1.1 Spatial Multipliers

An important aspect of the spatial lag model is the spatial multiplier, which can be illustrated by means of the reduced form (40). Consider the conditional expectation $\mathrm{E}[y|X]$. Since $\mathrm{E}[(I - \rho W)^{-1}\varepsilon|X] = 0$, it follows that

$$\mathrm{E}[y|X] = (I - \rho W)^{-1} X\beta. \tag{44}$$

Using the expansion of the inverse term as in (19) yields:

$$\mathrm{E}[y|X] = X\beta + \rho W X\beta + \rho^2 W^2 X\beta + \ldots, \tag{45}$$

which demonstrates how the value of y at i depends not only on x_i, but on the x at other locations as well, with locations further removed (higher powers of W) discounted by powers of the autoregressive parameter. This illustrates the *global* nature of the spatial multiplier effect in the spatial lag model. Specifically, if a unit change were introduced in a given explanatory variable X_k, the effect on y would amount to $[1/(1-\rho)]\beta_k$ (Kim et al. 2003, p. 35). More generally, for any vector of non-uniform changes in a given explanatory variable, ΔX_k, the resulting spatial pattern of changes in the dependent variable is:

$$\Delta y = (I - \rho W)^{-1} \Delta X_k \beta_k. \tag{46}$$

The global nature of the spatial multiplier in the spatial lag model contrasts with the local multiplier in a model with spatially lagged explanatory variables. Specifically, in a spatial cross-regressive model (Florax and Folmer 1992),

$$y = X\beta + WX\gamma + \varepsilon, \tag{47}$$

with WX as the spatially lagged explanatory variables (excluding the constant term) and γ a matching vector of parameters. The impact on y of any change in a given explanatory variable, ΔX_k is limited to the immediate effect $\Delta X_k \beta_k$ and the matching spatial lag effect (including only those neighbors as specified in W), $W \Delta X_k \gamma_k$ (for a review of a wide range of models incorporating spatial multipliers, see Anselin 2003b).

4.2 Higher Order Models

In analogy to higher order autoregressive models in the time series domain, multiple spatial lagged dependent variables may be included in a spatial lag model. For example, Brandsma and Ketellapper (1979) suggested a biparametric spatial autoregressive model:

$$y = \rho_1 W_1 y + \rho_2 W_2 y + X\beta + \varepsilon, \tag{48}$$

where W_1 and W_2 are different spatial weights. Higher order SAR models were considered in Blommestein (1983, 1985):

$$y = \rho_1 W_1 y + \rho_2 W_2 y + \cdots + \rho_p W_p y + X\beta + \varepsilon, \tag{49}$$

with $W_i, i = 1, \ldots, p$ as the associated spatial weights. The weights are typically (but not necessarily) associated with increasing orders of contiguity.

In contrast to time series models, care must be taken to avoid redundancies and circularities in the higher order spatial weights. As shown in Blommestein (1985), simple powers of the first order contiguity weights W_1, e.g., $W_p = (W_1)^p$ result in weights matrices that include lower order neighbors as well. These redundancies affect the properties of estimators and should be avoided (Blommestein 1985, Blommestein and Koper 1998). In order to ensure that the parameters in a higher order model are properly identified and to avoid biased estimates, it is typically assumed that the spatial weights do not overlap. More

formally, with w_{i*} as the i-th row of the weights matrix, and for any two orders of contiguity h, l:

$$(w_{i*}^h)(w_{i*}^l)' = 0. \tag{50}$$

In time series analysis, the problem of redundancy and circularity of higher order lags does not occur, due to the lack of two-way feedback in such models. In spatial analysis, consider the simple biparametric specification in (48), with $W_1 = W_{11} + W_0$ and $W_2 = W_{22} + W_0$, where W_0 is a weights matrix containing the common elements of W_1 and W_2. Taking into account this overlap yields:

$$y = \rho_1 W_{11} y + \rho_2 W_{22} y + (\rho_1 + \rho_2) W_0 y + X\beta + \varepsilon, \tag{51}$$

which illustrated the potential problems of identification and interpretation when the induced parameter constraints are not satisfied (for an extensive technical discussion, see Anselin and Smirnov 1996).

A more encompassing specification, that includes both higher order spatially lagged dependent variables as well as spatial moving average error terms, was suggested by Huang (1984). The so-called SARMA(p, q) model takes the same form as equation (49) for the spatial autoregressive part, with the error term following:

$$\varepsilon = \lambda_1 W_1 u + \lambda_2 W_2 u + \cdots + \lambda_q W_q u + u, \tag{52}$$

a spatial moving average process of order q. The full order p, q has seen no application in practice. However, a SARMA(1,1) model is commonly considered as an alternative in specification testing (see, e.g., Anselin 2001a).

A different higher order specification that has received significant attention consists of a spatial lag model with a spatially autoregressive error term (Anselin 1988b, p. 61):

$$y = \rho W_1 y + X\beta + \varepsilon \tag{53}$$

$$\varepsilon = \lambda W_2 \varepsilon + u, \tag{54}$$

with $\mathrm{E}[uu'] = \sigma^2 \Omega$ to allow for heteroskedasticity. While similar to a first order autoregressive model with serially correlated errors in time series, the spatial model is much more complex and requires great care to ensure identification of parameters.

After substituting (15) for the error term in the spatial lag specification and some further algebra, the following result is obtained:

$$y = \rho W_1 y + \lambda W_2 y - \rho\lambda W_2 W_1 y + X\beta - \lambda W_2 X\beta + u, \tag{55}$$

which simplifies when $W_2 W_1 = 0$ (non-overlapping weights):

$$y = \rho W_1 y + \lambda W_2 y + X\beta - \lambda W_2 X\beta + u. \tag{56}$$

However, in practice, the same weights are often used for both the lag and the error part, $W_1 = W_2$, which yields:

$$y = (\rho + \lambda) W y - \rho\lambda W^2 y + X\beta - \lambda W X\beta + u. \tag{57}$$

When all $\beta = 0$, this model will not be identified (for a technical discussion, see Kelejian and Prucha 1998, Kelejian et al. 2004, Das et al. 2003, Lee 2003). But even with some $\beta \neq 0$, there will be difficulty disentangling the role of ρ and λ. For example, consider the covariance that is induced by this model for the dependent variable y. From the reduced form, ignoring heteroskedasticity ($\Omega = I$), and setting $W_1 = W_2$, some algebra yields the variance-covariance matrix for y as:

$$\mathrm{Var}[y] = \sigma^2 (I - \rho W)^{-1} (I - \lambda W)^{-1} (I - \lambda W')^{-1} (I - \rho W')^{-1}. \tag{58}$$

Using the familiar expansion (e.g., equation 19) for the first two matrix inverses in this expression gives the following leading terms:

$$I + (\lambda + \rho) W + (\rho^2 + \lambda\rho + \lambda^2) W^2 + \ldots, \tag{59}$$

and a similar expression in the transpose W'. It is clear that λ and ρ are completely interchangeable, suggesting that the same spatial covariance structure for y can be obtained by a range of combinations of lag and error dependence.

In practice, higher order models are typically used as alternatives in diagnostic tests. Rejection of the null may not necessarily imply that the higher order model is the proper alternative. Since the specification of the weights matrix is an additional source of uncertainty, it is possible (likely) that a different specification of the weights eliminates the need for the higher order model (see Florax and Rey 1995, for a discussion of the effects of misspecified weights).

5 Estimation

The estimation problems associated with spatial regression models are distinct for the spatial lag and spatial error case. Spatial error models are special instances of specifications with a non-spherical error. On the other hand, the inclusion of a spatially lagged dependent variable results in a form of endogeneity. Each of these complications can be tackled with the customary econometric methods. However, the two-directional (feed-back) nature of spatial dependence and the reliance on a distance metric or spatial weights require specialized techniques to handle the associated probabilistic and computational aspects.

In this section, the two main approaches towards estimation of spatial models are considered in turn. They are based on the maximum likelihood principle and the generalized methods of moments. Other methods, such as the coding approach (Besag 1974, Haining 1990), and Markov Chain Monte Carlo estimation in a Bayesian framework (LeSage 1997, 1999) have been suggested as well, but they are not in common use and will not be further considered here (for an application of MCMC to the spatial probit model, see Section 8.2). The review of estimation methods is prefaced by a brief discussion of the complexities of asymptotic reasoning in the spatial domain.

5.1 Asymptotics in Space

The properties of estimators (and test statistics) in spatial econometrics are based on asymptotic considerations that use laws of large numbers (LLN) and central limit theorems (CLT) to establish consistency and asymptotic normality. These are not simple generalizations to two dimensions of results for the time domain, but there are differences in three important respects.

First, many spatial processes (e.g., SAR and SMA) also induce heteroskedasticity, requiring a joint treatment of dependence and heterogeneity. Specifically, this non-stationarity precludes reliance on central limit theorems for stationary mixing spatial random fields (e.g., Bolthausen 1982), which have been used to obtain the properties of estimators in models where the dependence is based on a distance metric (Conley 1999). The treatment of heterogeneous spatially dependent processes that include spatial weights is further complicated by the necessity to consider CLT and LLN for triangular arrays. This is caused by the fact that the weights depend on the sample size, and precludes "standard" results (e.g., for maximum likelihood estimation) from being directly applicable. The importance of triangular arrays in the asymptotics of spatial econometrics is highlighted in the work of Kelejian and Prucha (1998, 1999).

A second important distinguishing characteristics is the way in which spatial sampling is conceptualized and the type of population it pertains to. Spatial data can be viewed as either constituting a continuous surface, or as a collection of discrete objects. The former view is prevalent in the physical sciences and geostatistical approaches (e.g., Cressie 1993, Stein 1999). The latter is common in spatial econometrics. When considering a population of spatial objects, the mechanism through which observations or locations are selected is not arbitrary, and can be deterministic (fixed grid cells or points) or stochastic. In the stochastic case, an important concept is that of a *directing process*, which is the set of random indices (i.e., locations) used to draw samples from the population. As a consequence, the random spatial process itself becomes *subordinated* to the directing process, a complication which is typically not encountered in the time domain (see Conley 1999, Andrews 2003, for a technical discussion).

A third distinction pertains to the way in which the sample increases to reach the asymptotic limit, $n \to \infty$. A pure increasing domain structure is obtained when the minimum distance between neighboring locations remains bounded away from zero as the sample size grows (Cressie 1993). One can conceive of this situation as a sampling structure where new data points are added at the edge such that the observation "region" becomes unbounded. In contrast, infill asymptotics are obtained when the sample region is bounded, but the number of data points increases. This yields a denser and denser sampling surface, with the minimum distance between sample locations approaching zero as $n \to \infty$. Asymptotic results that hold in the increasing domain case are often not transferable to the infill case, as shown in Lahiri (1996). Mixed situations can be considered as well, where an increasing domain "grid" is combined with infill in subregions (see, e.g., Fazekas and Kukush 2000, Lahiri and Mukherjee 2004).

All asymptotic properties are based on regularity conditions that restrict the range and strength of spatial dependence and heterogeneity. In models using spatial weights, these conditions translate into constraints on the structure of the weights (see Section 3.1.1). In practice, these are readily satisfied by contiguity-based weights, but not necessarily by more complex specifications. Similar regularity conditions on direct representation models are formulated in Mardia and Marshall (1984).

5.2 Maximum Likelihood Estimation

5.2.1 Spatial Lag Model

The point of departure for maximum likelihood (ML) estimation in spatial regression models is an assumption of normality for the error term. In general, allowing for heteroskedasticity and/or error correlation, the $n \times 1$ error vector has a multivariate normal distribution, $\varepsilon \sim N(0, \Sigma_\theta)$, with the subscript θ denoting that Σ may be a function of a $p \times 1$ vector θ of parameters. In the commonly considered i.i.d. case, this simplifies to $\varepsilon \sim N(0, \sigma^2 I)$, with $\theta = \sigma^2$.

To move from the likelihood for the error vector to a likelihood for the observed dependent variable, a *Jacobian* of the transformation needs to be inserted. In the spatial lag model (39), this corresponds to:

$$|\partial\varepsilon/\partial y| = |\partial(y - \rho Wy - X\beta)/\partial y| = |I - \rho W|, \tag{60}$$

the determinant of a $n \times n$ matrix. Note that this same Jacobian reduces to a scalar 1 in the standard regression model, since $|\partial(y - X\beta)/\partial y| = |I| = 1$. The presence of the Jacobian term constitutes a major computational complication (see Section 5.2.3).

Using the standard result for a multivariate normal distribution, and taking into account the Jacobian term, the log-likelihood for the spatial lag model follows as:

$$\begin{aligned} L &= -(n/2)(\ln 2\pi) - (1/2)\ln|\Sigma_\theta| + \ln|I - \rho W| \\ &\quad -(1/2)(y - \rho Wy - X\beta)'\Sigma_\theta^{-1}(y - \rho Wy - X\beta). \end{aligned} \tag{61}$$

Maximizing the log-likelihood is *not* equivalent to minimizing weighted least squares (the last term in L), as in the standard linear regression model. The main difference is in the presence of the log-Jacobian term $\ln|I - \rho W|$. This illustrates informally how weighted least squares will not yield a consistent estimator in the spatial lag model, due to the endogeneity in the Wy term (see Section 4.1).[15] The log-Jacobian also implies constraints on the parameter space for ρ, which must be such that $|I - \rho W| > 0$.

[15] An exception is the case where all observations in a "region" are considered to be neighbors, yielding weights of $1/(n_g - 1)$ for each (with n_g as the number of members in the group). In this case, OLS can be shown to yield consistent estimates (Lee 2002, Kelejian and Prucha 2002).

ML estimates for β, ρ and θ are obtained as solutions to the usual first order conditions (for technical details, see Ord 1975, Cliff and Ord 1981, Anselin 1980, 1988b, Anselin and Bera 1998, among others):

$$\begin{aligned} \partial L/\partial\beta &= e'\Sigma_\theta^{-1}X = 0, && (62)\\ \partial L/\partial\rho &= -tr[W(I-\rho W)^{-1}] + e'\Sigma_\theta^{-1}Wy = 0, && (63)\\ \partial L/\partial\theta_i &= -(1/2)tr[\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)] \\ &\quad +(1/2)e'\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)\Sigma_\theta^{-1}e = 0, \text{for } i=1,\ldots,p, && (64) \end{aligned}$$

with $e = y - \rho Wy - X\beta$ and tr as the matrix trace operator. Solutions to these conditions need to be obtained through numerical optimization.

Inference is based on an asymptotic variance matrix, the inverse of the information matrix. In the general case considered here, the information matrix for $[\rho, \beta, \theta]$ is:

$$\begin{bmatrix} tr[W_\rho]^2 + tr[\Sigma_\theta W_\rho'\Sigma_\theta^{-1}W_\rho] + (\widehat{Wy})'\Sigma_\theta^{-1}(\widehat{Wy}) & (X'\Sigma_\theta^{-1}\widehat{Wy})' & \psi' \\ X'\Sigma_\theta^{-1}\widehat{Wy} & X'\Sigma_\theta^{-1}X & 0 \\ \psi & 0 & (1/2)\Psi \end{bmatrix} \tag{65}$$

with $W_\rho = W(I-\rho W)^{-1}$, $\widehat{Wy} = W(I-\rho W)^{-1}X\beta$, and ψ as a $p \times 1$ vector of matrix traces of the form:

$$\psi_i = tr[\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)W_\rho], \tag{66}$$

one for each parameter in θ. Similarly, Ψ is a $p \times p$ matrix containing matrix traces of the form:

$$\Psi_{i,j} = tr[\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_j)]. \tag{67}$$

Note how the presence of the vectors ψ implies that the information matrix is not block-diagonal between the model parameters $[\rho, \beta]$ and the error parameters θ. This is an important distinguishing characteristic of the spatial lag model, and leads to some interesting results on the structure of specification tests (Anselin and Bera 1998).

Consider the special case of groupwise heteroskedasticity, which has considerable appeal in practice.[16] The error variance-covariance matrix Σ is diagonal, with $g = 1, \ldots p$ subdiagonals, each corresponding to a group (sorting the observations by group, without loss of generality):

$$\Sigma = \begin{bmatrix} \sigma_1^2 I_1 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 I_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \sigma_p^2 I_p \end{bmatrix} \tag{68}$$

[16] For example, this model is relevant when different subregions in the data are allowed to have a different error variance, or when pooling observations in space-time, see Section 7.1.2.

with $I_g, g = 1, \ldots, p$ as an identity matrix, with n_g elements matching the number of observations in each group and $\sum_g n_g = n$.

Plugging these expressions into the first order conditions (62)–(64) yields analyical solutions for β_{ML} and $\sigma_{g,ML}$, conditional upon the value of ρ, as:

$$\beta_{ML} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(y - \rho Wy) \tag{69}$$

$$\sigma^2_{g,ML} = (y - \rho Wy - X\beta)'H_g(y - \rho Wy - X\beta)/n_g, \; g = 1, \ldots, p \tag{70}$$

where H_g is an identity matrix with all diagonal elements except those corresponding to group g set to zero, making (70) the average sum of squared residuals for the respective group. The estimate of β_{ML} is a weighted sum of a FGLS estimation of X on y and X on Wy, with $-\rho$ as the weight. However, unlike the i.i.d. situation (Anselin 1980, Chapter 4), this does not lead to a simple concentrated likelihood, since the FGLS depends on the σ^2_g, which in turn depend on the unknown ρ. Estimates can be obtained in an iterative fashion by substituting values for β and Σ from a previous iteration into (63), solving this condition for ρ, which in turn yields new values for β and Σ from (69)-(70). Note that Σ^{-1} is a diagonal matrix with elements $1/\sigma^2_g$ corresponding to each group. Asymptotic inference can be based on the information matrix (65), with the simpler Σ^{-1} substituted for the general form, and with $\partial\Sigma/\partial\sigma^2_g = H_g$.

5.2.2 Spatial Error Models

Maximum likelihood estimation of the parameters in models with spatially dependent error terms follows as a special case of the results in Magnus (1978). For a general non-spherical error term Σ_θ, with θ as the parameters, the ML estimator for β is the familiar generalized least squares expression:

$$\hat{\beta}_{ML} = (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}y. \tag{71}$$

This follows as the solution of the first order conditions, applied to the log-likelihood:

$$L = -(n/2)\ln(2\pi) - (1/2)\ln|\Sigma_\theta| - (y - X\beta)'\Sigma_\theta^{-1}(y - X\beta). \tag{72}$$

The estimators for the θ_i are obtained from the following first order conditions (Magnus 1978, p. 283):

$$tr\left[\left(\partial\Sigma_\theta^{-1}/\partial\theta_i\right)\Sigma_\theta\right] = e'\left(\partial\Sigma_\theta^{-1}/\partial\theta_i\right)e, \tag{73}$$

with $e = y - X\hat{\beta}_{ML}$ as the residuals. With a consistent estimate for the parameters θ_i, consistent estimates for β are obtained through FGLS (71).

Asymptotic inference is based on the inverse of the information matrix, which is block-diagonal between the β and the parameters of the error variance-covariance (Breusch 1980). The asymptotic variance for $\hat{\beta}_{ML}$ takes on the familiar GLS form:

$$AsyVar[\hat{\beta}_{ML}] = (X'\Sigma_\theta^{-1}X)^{-1}, \tag{74}$$

while the block corresponding to the error variance-covariance parameters is of the form $2\Psi^{-1}$, with the elements of Ψ as in (67).

The range of spatial error processes considered in Section 3 will result in specialized forms for Σ_θ, some of which may simplify the expressions in (71)–(73). For example, consider the SAR error process outlined in Section 3.1.2, without heteroskedasticity. The corresponding parameter vector is $\theta = [\sigma^2, \lambda]$, and the error variance-covariance matrix is as in (18).

As a result, the FGLS estimator in this model simplifies to:

$$\hat{\beta}_{ML} = [X'(I - \hat{\lambda}W)'(I - \hat{\lambda}W)X]^{-1}X'(I - \hat{\lambda}W)'(I - \hat{\lambda}W)y, \tag{75}$$

or, a regression of spatially filtered $X_L = X - \hat{\lambda}WX$ on spatially filtered $y_L = y - \hat{\lambda}Wy$, sometimes referred to as spatially weighted least squares. The partial derivative for use in condition (73) is:

$$\partial\Sigma^{-1}/\partial\lambda = -W - W' + \lambda W'W. \tag{76}$$

Unlike the time series counterpart, a consistent estimate for λ cannot be obtained from a simple auxiliary regression, but the first order condition (73) must be solved by numerical means. As for the spatial lag model, asymptotic inference is based on the inverse of the information matrix (for technical details, see Anselin 1988b, Chapter 6).

Other spatial error processes do not yield a simple expression for $\hat{\beta}_{ML}$ as in the spatially weighted regression (75). For example, in the SMA process (Section 3.1.4), the error variance-covariance matrix is as in (25), and:

$$\Sigma^{-1} = (1/\sigma^2)[I + \gamma(W + W') + \gamma^2 WW']^{-1}, \tag{77}$$

requiring the inverse of a $n \times n$ matrix to carry out FGLS. Similarly, the partial derivatives required for condition (73) involve inverses of this order. Models for direct representation or error components suffer from the same problem. In addition, the functional forms and distance metric used in the direct representation approach may cause problems with multiple optima in the log-likelihood function (see, among others, Mardia and Marshall 1984, Warnes and Ripley 1987, Mardia and Watkins 1989).

5.2.3 Computational Issues

Maximum likelihood estimation in spatial regression models involves the application of nonlinear optimization techniques to the log-likelihood function. A main computational obstacle follows from the presence of the log-Jacobian term $\ln|I - \rho W|$ in the log-likelihood. In addition, the first order conditions and information matrix involve the traces of matrix products such as $W(I - \rho W)^{-1}$. For even medium-sized data sets, the computation of these terms by "brute force" is impractical.

An early solution was suggested by Ord (1975), who exploited the decomposition of the Jacobian in terms of the eigenvalues of the spatial weights matrix:

$$|I - \rho W| = \prod_{i=1}^{n}(1 - \rho\omega_i), \tag{78}$$

where the $\omega_i, i = 1, \ldots, n$ are the eigenvalues of W. The log-Jacobian then follows as:

$$\ln|I - \rho W| = \sum_{i=1}^{n} \ln(1 - \rho\omega_i). \tag{79}$$

This facilitates computation greatly, since the eigenvalues only need to be calculated once, and iterating over values of ρ in (79) is straightforward. The trace terms used in the information matrix can be expressed in terms of the eigenvalues as well (Anselin 1980).

The eigenvalue decomposition also suggests a simple set of constraints for the parameter space. For (79) to be valid, the condition $(1-\rho\omega_i) > 0$ must hold $\forall\ i$. This results in a parameter space of $(1/\omega_{min},\ 1/\omega_{max})$, where ω_{min} and ω_{max} are, respectively, the smallest (most negative) and largest eigenvalues of W (Anselin 1980). For row-standardized spatial weights, $\omega_{max} = 1$, but ω_{min} depends on the structure of the weights matrix and is typically < -1. While it is often suggested in the literature to constrain the parameter space to the interval $(-1, +1)$, this may be overly restrictive.[17]

The computation of eigenvalues becomes impractical and computationally unstable for medium and large-sized data sets ($n > 1000$). This precludes the application of the Ord approach. Several alternatives have been suggested that either approximate our bound the Jacobian or log-Jacobian term (e.g., Martin 1993, Griffith and Sone 1995, Barry and Pace 1999, Pace and LeSage 2002, 2004), or exploit the sparse nature of spatial weights.[18] The latter include factorization methods for sparse matrices, such as Cholesky decomposition (Pace and Barry 1997a,b), and a characteristic polynomial approach (Smirnov and Anselin 2001). The characteristic polynomial approach in particular allows very large spatial regression models ($n > 1$ million) to be estimated in a realistic time.

A second important computational problem pertains to the presence of terms like $tr[W(I-\rho W)^{-1}]^2$ in the information matrix, such as in (65). The calculation of these inverse matrices is impractical in large data settings. As a result, most large data ML methods developed so far have not based inference on the asymptotic variance matrix, but instead use a sequence of Likelihood Ratio tests. Recently, Smirnov (2005) developed a solution to this problem, based on the use of a conjugate gradient approach.

While much progress has been made, several issues remain to be resolved to allow the full range of spatial models (e.g., not just SAR models) to be estimated by maximum likelihood methods in large samples.

[17] In contrast to the time series literature, there is relatively little attention to unit root and cointegration issues in the spatial domain. An exception is the work of Fingleton (1999).

[18] For example, a first order contiguity weights matrix for the 3,140 U.S. counties contains only 0.19% non-zero elements.

5.3 Method of Moments/Instrumental Variables

5.3.1 Spatial Two Stage Least Squares

An alternative to maximum likelihood estimation is the use of the method of moments (including instrumental variables, generalized method of moments, and generalized moments). This approach does not require an assumption of normality and it avoids some of the computational problems associated with ML for very large data sets.

Consider the spatial lag model (2), rewritten as:

$$y = Z\gamma + \varepsilon, \tag{80}$$

with $Z = [Wy,\ X]$ and $\gamma = [\rho,\ \beta]$. This is a general specification of a linear model that contains endogenous variables (Wy) as well as exogenous variables (X).

A classic solution to the endogeneity problem is to use instrumental variables. In a two-stage least squares approach, the predicted value of Z in a regression on the instruments is obtained in a first stage, as:

$$\widehat{Z} = Q(Q'Q)^{-1}Q'Z, \tag{81}$$

with Q as an $n \times q$ matrix of instruments (including the exogenous variables X), with $q \geq k + 1$. Note how this has no impact on the exogenous variables X, but it yields:

$$\widehat{Wy} = Q(Q'Q)^{-1}Q'Wy. \tag{82}$$

The instrument $\widehat{Z}$ replaces Z in the second stage, resulting in the spatial two stage least squares estimator:

$$\hat{\gamma}_{2SLS} = [\widehat{Z}'\widehat{Z}]^{-1}\widehat{Z}'y, \tag{83}$$

or, in full:

$$\hat{\gamma}_{2SLS} = [Z'Q(Q'Q)^{-1}Q'Z]^{-1}Z'Q(Q'Q)^{-1}Q'y. \tag{84}$$

Inference on the γ_{2SLS} is based on the asymptotic variance matrix:

$$AsyVar[\hat{\gamma}_{2SLS}] = \hat{\sigma}^2[Z'Q(Q'Q)^{-1}Q'Z]^{-1}, \tag{85}$$

with $\hat{\sigma}^2 = (y - Z\hat{\gamma}_{2SLS})'(y - Z\hat{\gamma}_{2SLS})/n$.

The application of instrumental variables to the spatial lag model was outlined in Anselin (1980, 1988b, pp. 82–86), where some ad hoc suggestions were made for the selection of the instruments (see also Land and Deane 1992, for an early discussion). Specifically, the choice of a spatial lag of the predicted values of the y (using only the exogenous variables) or of spatially lagged exogenous variables was considered.

In Kelejian and Robinson (1993), the consistency of $\hat{\gamma}_{2SLS}$ is derived formally and the selection of instruments is couched in terms of the reduced form (44). From this, it follows that:

$$\mathrm{E}[Wy|X] = W(I - \rho W)^{-1} X\beta, \tag{86}$$

or, using the expansion (19):

$$\mathrm{E}[Wy|X] = WX\beta + \rho W^2 X\beta + \rho^2 W^3 X\beta + \dots. \tag{87}$$

Based on this expansion, Kelejian and Robinson (1993) suggest the use of a subset of columns from $\{X, WX, W^2X, W^3X, \dots\}$ as the instruments (see also Kelejian and Prucha 1998).

In the context of a model with both a spatial lag and spatial error dependence, recent work has focused on the selection of "optimal" instruments (see also Section 5.3.3). First consider this for the simpler specification of the pure spatial lag model without error dependence. Lee (2003) suggested using the optimal instrument matrix:

$$Q = [X, W(I - \hat{\rho}W)^{-1} X\hat{\beta}], \tag{88}$$

where the values for $\hat{\rho}$ and $\hat{\beta}$ are obtained in a first round estimation, using WX as the instrument (possibly augmented with W^2X).[19]

To avoid the inverse matrix operation, Kelejian et al. (2004) introduce a series approximation, with the instrument matrix (for the pure spatial lag case) as:

$$Q = [X, \sum_{s=0}^{r} \hat{\rho}^s W^{s+1} X\hat{\beta}\ ], \tag{89}$$

and the values for $\hat{\rho}$ and $\hat{\beta}$ from a first round estimation, as in the Lee (2003) approach. The highest power in the approximation is related to the sample size, with $r = o(n^{1/2})$.[20]

Recent extensions of the instrumental variables approach to systems of simultaneous equations are considered in Rey and Boarnet (2004) and Kelejian and Prucha (2004). In Pinkse et al. (2002), an instrumental variable estimator is applied to a semi-parametric spatial lag model. Instead of using a spatial weights matrix, the spatial dependence is specified in a generic way, as in equation (1). The function g is approximated by a polynomial series expansion in distance measures, the coefficient of which are estimated jointly with the other parameters in the model.

[19] Since this optimal instrument involves the inverse of a $n \times n$ matrix, it will be computationally prohibitive in large samples. Lee (2003) suggests a workaround based on Cholesky decomposition.

[20] In their simulation experiments, Kelejian et al. (2004) set $r = n^{\alpha}$, with $\alpha = .25,\ .35$ and .45.

5.3.2 Spatial Correlation as a Nuisance Parameter

A basic property of FGLS is that it suffices to obtain a consistent estimate for the parameters of the error variance-covariance in order to obtain consistent estimates for the β in the second step, as in (71). This general approach refers to the error variance-covariance parameters as *nuisance* parameters, in that they are not of interest in and of themselves, but are only useful in terms of improving the properties of the $\hat{\beta}$.

Kelejian and Prucha (1999) suggest a generalized moments approach to obtain a consistent estimate of the parameter λ in a spatial autoregressive error process $\varepsilon = \lambda W\varepsilon + u$. The idiosyncratic error term u is assumed to be i.i.d with variance σ^2. Using the property that $trW = 0$, three moment conditions on the u and their spatial lags Wu follow as:

$$\begin{aligned} \mathrm{E}[(1/n)u'u] &= \sigma^2 && (90)\\ \mathrm{E}[(1/n)u'W'Wu] &= (1/n)\sigma^2 tr(W'W) && (91)\\ \mathrm{E}[(1/n)u'W'u] &= 0. && (92) \end{aligned}$$

Since $u = \varepsilon - \lambda W\varepsilon$ and $Wu = W\varepsilon - \lambda WW\varepsilon$, these conditions can be expressed in terms of the regression errors ε. They are made operational by using the sample counterparts, i.e., the residuals e, their spatial lags We and double spatial lags WWe. After some algebra, this yields a system of three equations in the unknowns λ, λ^2 and σ^2:

$$\begin{bmatrix} (2/n)e'\bar{e} & (-1/n)\bar{e}'\bar{e} & 1 \\ (2/n)\bar{\bar{e}}'\bar{e} & (-1/n)\bar{\bar{e}}'\bar{\bar{e}} & (1/n)tr(W'W) \\ (1/n)(e'\bar{\bar{e}} + \bar{e}'\bar{e}) & (-1/n)\bar{e}'\bar{\bar{e}} & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} (1/n)e'e \\ (1/n)\bar{e}'\bar{e} \\ (1/n)e'\bar{e} \end{bmatrix} \tag{93}$$

with, for notational simplicity, $\bar{e} = We$ and $\bar{\bar{e}} = WWe$. This system can be solved for λ and σ^2 using nonlinear least squares. The resulting estimate $\hat{\lambda}$ is then plugged into the expression for the spatially weighted least squares (75) (for technical details, see Kelejian and Prucha 1999). Note that the nuisance parameter approach precludes inference about the coefficient λ, since there is no asymptotic variance for $\hat{\lambda}$.

5.3.3 Spatial Lag with Spatial SAR Errors

The general moments estimator can also be applied to the residuals from a two stage least squares regression. A special case consists of the specification with both a spatial lag and a spatial autoregressive error term (53), considered in Kelejian and Prucha (1998). This model can also be expressed as:

$$y = Z\gamma + (I - \lambda W_2)^{-1}u, \tag{94}$$

with $Z = [W_1 y,\ X]$ and $\gamma = [\rho,\ \beta]$, as before. Removing the inverse term yields:

$$y - \lambda W_2 y = (Z - \lambda W_2 Z)\gamma + u, \tag{95}$$

or,

$$y_L = Z_L\gamma + u, \tag{96}$$

with y_L and Z_L as spatially filtered variables, analogous to the transformed variables in the spatially weighted least squares regression. Note how the first term in Z_L is $(W_1y)_L = W_1y - \lambda W_2W_1y$.

The generalized spatial two stage least squares (GS2SLS) estimator developed in Kelejian and Prucha (1998) consists of three steps. The first step is a spatial two stage least squares estimation, as in (84). The second step consists of substituting the residuals $e = y - Z\hat{\gamma}_{2SLS}$ in the system of equations (93), using the weights W_2 from the error process. The solution of the system by nonlinear least squares yields a consistent estimate $\hat{\lambda}$ for the autoregressive error parameter.

Next, the values for $\hat{\lambda}$ are substituted into the spatial filter to obtain y_L and Z_L. The final stage consists of a second two stage least squares estimation, using the spatially filtered variables and the instruments:

$$\widehat{Z_L} = Q(Q'Q)^{-1}Q'Z_L, \tag{97}$$

with the estimator as:

$$\hat{\gamma}_{GS2SLS} = [\widehat{Z_L}'\widehat{Z_L}]^{-1}\widehat{Z_L}'y. \tag{98}$$

The estimation procedure can also be iterated, substituting the residuals from the GS2SLS estimation in the equation system (93) to yield a new value of $\hat{\lambda}$, etc. The asymptotic variance for the GS2SLS estimator follows in the usual fashion.

Note how the instruments boil down to replacing the $(W_1y)_L$ in Z_L by the predicted values:

$$\widehat{(W_1y)_L} = Q(Q'Q)^{-1}Q'(W_1y - \hat{\lambda}W_2W_1y). \tag{99}$$

Initially, Kelejian and Prucha (1998, p. 104) suggested that the instruments be selected as a subset of from $\{X, W_1X, W_1^2X, \ldots, W_2X, W_2W_1X, W_2W_1^2X, \ldots\}$, such as $[X, W_1X, W_2X, W_2W_1X]$. More recently, optimal instruments have been considered, as discussed in Section 5.3.1. Specifically, the optimal instruments (88) of Lee (2003) in this general model are:

$$Q = (I - \hat{\lambda}W_2)[X,\ W_1(I - \hat{\rho}W_1)^{-1}X\hat{\beta}]. \tag{100}$$

For the Kelejian et al. (2004) series estimator (89), this is:

$$Q = (I - \hat{\lambda}W_2)[X,\ \sum_{s=0}^{r}\hat{\rho}^s W_1^{s+1}X\hat{\beta}\ ]. \tag{101}$$

Monte Carlo experiments to assess the finite sample properties of the GS2SLS estimator relative to ML are reported in Das et al. (2003). In general, very small

differences are obtained. Interestingly, there does not seem to be a worthwhile payoff from iterating on the GS2SLS (Das et al. 2003, p. 12). This finding does not transfer to the methods using the optimal instruments, evaluated in Monte Carlo experiments in Kelejian et al. (2004). There, a suggestion is made that iterating will tend to improve efficiency. Overall, the performance of the estimators with the three forms of instruments was found to be very similar.

5.3.4 Heteroskedastic and Autocorrelation Consistent Estimators

Up to this point, the estimators considered pertain to models in which the spatial dependence is specified in parameterized form, typically as a SAR process. An alternative is to avoid such specification, and estimate the error variance-covariance matrix in non-parametric fashion. This follows along the lines of the White (1980) heteroskedastic-consistent approach and its extension to both heteroskedasticity and serial correlation by Newey and West (1987), and others.

For example, a ready extension of the S2SLS estimator in the spatial lag model is to allow for heteroskedasticity of unspecified form, as a direct application of the White approach. In a first step, a S2SLS estimation is carried out using instrument matrix Q (as before), followed by:

$$\hat{\beta}_{HS2SLS} = [Z'Q(\widehat{Q'\Sigma Q})^{-1}Q'Z]^{-1}Z'Q(\widehat{Q'\Sigma Q})^{-1}Q'y, \tag{102}$$

with $(\widehat{Q'\Sigma Q})^{-1} = (Q'SQ)^{-1}$, where S is a diagonal matrix containing the squared S2SLS residuals. The asymptotic variance for $\hat{\beta}_{HS2SLS}$ is obtained in the usual fashion.

The incorporation of spatial dependence in this framework was first considered by Conley (1999) in the context of GMM estimation, and recently elaborated upon in Kelejian and Prucha (2003) (see also Section 3.3).[21] Using a consistent estimate for $(Q'\Sigma Q)^{-1}$ along the lines of (30) or (31) yields a heteroskedastic and spatial autocorrelation consistent (HAC) estimator for the spatial lag model as (102). Alternatively, this idea can be exploited to obtain robust inference in the standard linear regression model, using OLS to estimate the β, but a HAC estimator for the asymptotic variance:

$$AsyVar[\hat{\beta}_{OLS}] = (X'X)^{-1}\widehat{V}(X'X)^{-1}, \tag{103}$$

with $\widehat{V}$ based on (30) or (31) (omitting the term n^{-1}).

The HAC approach is asymptotic and in finite samples a major practical problem is to ensure that the estimated variance-covariance matrix is positive semidefinite. A number of suggestions have been formulated, but considerable research remains to be done to obtain insight into finite sample properties (see Kelejian and Prucha 2003, for some technical details).

[21] A slightly different approach towards estimating the error covariance matrix non-parametrically is offered in Pinkse et al. (2002, pp. 1126–1128).

6 Specification Testing

In applied work, diagnostic tests against the presence of spatial correlation are arguably more relevant than estimation itself. Ignoring spatial correlation when it is present may lead to biased and inconsistent estimates of the model parameters (spatial lag model), or inefficient estimates and biased t-test statistics (spatial error model). Also, given the complexities associated with estimating models that include spatial dependence, it is important to be able to assess whether this is in fact necessary. It is therefore reasonable that some check for autocorrelation should be as common in cross-sectional regression work as are tests for serial correlation in the time domain. In these tests, the null hypothesis is the absence of spatial autocorrelation, or the standard regression model. Alternatives include the many spatial models reviewed in Sections 3 and 4, as well as combinations of these spatial models with other sources of misspecification, such as heteroskedasticity and non-linearity.

The literature on specification tests against spatial correlation in cross-sectional regression is by now quite extensive (for recent reviews, see Anselin and Bera 1998, Anselin 2001a, Florax and de Graaff 2004). In the following sections, the main features of the most commonly used tests are outlined, organized along tests against spatial autocorrelation, tests based on the Maximum Likelihood principle, and tests against multiple sources of misspecification.[22] To close, a brief discussion is given of the specification search process.

6.1 Tests Against Spatial Autocorrelation

Tests against spatial autocorrelation are so-called "diffuse" tests (Florax and de Graaff 2004) in that the alternative is an unspecified form of spatial correlation. These test statistics are excellent tools as misspecification diagnostics. However, they offer little in terms of guiding a specification search, since they have power against multiple forms of spatial autocorrelation. Hence, it is not always clear which of these is the proper alternative.

6.1.1 Moran's I Test

Perhaps the best known test statistic against spatial autocorrelation is the application of Moran's I statistic (Moran 1948) to regression residuals (Moran 1950), popularized in the work of Cliff and Ord (Cliff and Ord 1972, 1973, 1981).

Formally, Moran's I for regression residuals is:

$$I = \frac{e'We/S_0}{e'e/n}, \tag{104}$$

where e is a $n \times 1$ vector of OLS residuals $y - X\hat{\beta}$, W is a spatial weights matrix,

[22] The focus is on parametric tests. For a general non-parametric approach to testing spatial error dependence, see Brett and Pinkse (1997).

and $S_0 = \sum_i \sum_i w_{ij}$, a normalizing factor.[23] Inference in a test against spatial autocorrelation is based on a normal approximation, using a standardized value, or z-value. This is obtained by subtracting the mean under the null and dividing by the square root of the variance. The first two moments were derived in Cliff and Ord (1972) as:

$$\mathrm{E}[I] = tr(MW)/(n-k), \tag{105}$$

and

$$\mathrm{Var}[I] = \frac{tr(MWMW') + tr(MWMW) + [tr(MW)]^2}{(n-k)(n-k+2)} - (E[I])^2, \tag{106}$$

where tr is a matrix trace operator and $M = I - X(X'X)^{-1}X'$. The normality of the z-value is an approximation, which works well in large samples.

Exact inference for Moran's I can be carried out under the assumption of normality for the error terms, by noting that under the null, the statistic consists of a ratio of quadratic forms of independent normal random variables (see Tiefelsdorf and Boots 1995). To obtain exact critical values for the statistic, it is necessary to compute the $n-k$ non-zero eigenvalues of MW, which is not practical in medium to large samples (see also Tiefelsdorf 2002, for an alternative approximation).

Moran's I has been shown to have certain optimality properties, similar to the Durbin-Watson test against serial correlation in the time domain. For example, King (1981) demonstrated that the test is locally best invariant. Also, Moran's I turns out to be asymptotically equivalent to a Likelihood Ratio (LR) test and to a Rao Score (RS) or Lagrange Multiplier (LM) test (Cliff and Ord 1972, Burridge 1980) and therefore shares the asymptotic properties of these statistics.

Moran's I is often interpreted as a test statistic against serial error correlation, but this is incorrect. The test has power against any alternative of spatial correlation, including spatial lag dependence, as demonstrated in a large number of Monte Carlo simulation experiments (see, e.g., Anselin and Rey 1991, Anselin and Florax 1995b, Florax and de Graaff 2004). In addition, not unlike the Durbin-Watson statistic, the test has power against heteroskedasticity as well (Anselin and Griffith 1988).

Moran's I test statistic is very general and can be applied in many contexts other than the classic regression model. For example, in Anselin and Kelejian (1997) the test is applied to the residuals of a model with endogenous variables, estimated by two stage least squares (2SLS). A distinction needs to be made between the case where the endogenous variables include a spatial lag and the standard non-spatial case. The general case can be written as:

$$y = Z\gamma + e, \tag{107}$$

[23] Note that for row-standardized weights, $S_0 = \sum_i 1 = n$, such that the normalizing factors cancel out and $I = e'We/e'e$.

with $Z = [Wy, Y, X]$ and $\gamma = [\rho, \delta, \beta]$, where Wy is the spatially lagged dependent variable, Y is a matrix of non-spatial endogenous variables, and X a matrix of exogenous variables.

The general result for the asymptotic distribution of Moran's I in this model is:

$$n^{1/2} I \xrightarrow{d} N(0, \phi^2), \tag{108}$$

with I as in (104), but using the 2SLS residuals from (107), and

$$\phi^2 = \frac{S_2}{2S_1^2} + \frac{4}{S_1^2 \hat{\sigma^2}} A. \tag{109}$$

The terms in the expression for the asymptotic variance include the error variance estimated from 2SLS residuals, $\hat{\sigma^2}$, functions of the elements of the weights matrix:

$$S_1 = (1/n) S_0 \tag{110}$$

$$S_2 = (1/n) tr[(W + W')(W + W')] \tag{111}$$

and

$$A = (e'WZ/n)(n[(Z'Q)(Q'Q)^{-1}Q'Z]^{-1})(Z'W'e/n), \tag{112}$$

with e as the 2SLS residuals, S_0 and in (104), Z as in (107), and Q as the matrix of instruments used in the 2SLS estimation (for details, see Anselin and Kelejian 1997, pp. 162–164). When no spatially lagged dependent variable is included in the model, the second term in (109) becomes zero and the test statistic simplifies to the Lagrange Multiplier statistic for spatial error autocorrelation (see Section 6.2.1).

A general framework for the asymptotic properties of Moran's I in a wide range of specifications is presented in Kelejian and Prucha (2001). This includes a spatial autocorrelation test in probit models, which is considered more closely in Section 8.3.

6.1.2 Kelejian-Robinson Test

A second test against an unspecified form of spatial correlation was suggested by Kelejian and Robinson (1992). The rationale for the test is intuitive: if the covariance between "neighboring" pairs of error terms shows a systematic variation, then the null hypothesis of no spatial autocorrelation should be rejected. Formally, the spatial covariance is specified as:

$$\text{Cov}[\varepsilon_i \varepsilon_j] = \sigma_{ij} = v_{ij}\gamma \neq 0, \tag{113}$$

where v_{ij} is a $1 \times q$ vector of covariates related to the pair of locations $i,\ j$. For example, a given element h of v_{ij} can be specified as $v_{ijh} = x_{ih}x_{jh}$, the product of the explanatory variable x_h for the pair of locations $i,\ j$. Under the null

hypothesis, there should be no covariance, hence $H_0 : \gamma = 0$. The test is made operational by selecting h_n pairs of cross-products of residuals, $\hat{C}_{ij} = e_i.e_j$ and regressing them on matching cross-products of the explanatory variables, v_{ij}. In matrix notation:

$$\hat{C} = V\gamma + u, \tag{114}$$

with u as an idiosyncratic error. The test statistic is a measure of goodness-of-fit in this auxiliary regression:

$$KR = \frac{\hat{\gamma}'V'V\hat{\gamma}}{\hat{\sigma}^4}, \tag{115}$$

which, under the null hypothesis, is distributed asymptotically as $\chi^2(q)$, with q as the number of explanatory variables in the auxiliary regression (i.e., the number of columns of V). The denominator term $\hat{\sigma}^4$ is any consistent estimator for σ^4, such as $\hat{\sigma}^4 = (\hat{C} - V\hat{\gamma})'(\hat{C} - V\hat{\gamma})/h_n$. Anselin and Bera (1998, p. 269) show how the test has an asymptotically equivalent expression in the form of the familiar NR^2 in an auxiliary regression:

$$KR = h_n.\frac{\hat{C}'V(V'V)^{-1}V'\hat{C}}{\hat{C}'\hat{C}}, \tag{116}$$

where h_n is the sample size of the auxiliary regression and the R^2 is the uncentered coefficient of determination in a regression of $\hat{C}$ on V.

The properties of the test statistic are obtained without assuming normality, and they do not require the regression model to be linear. Kelejian and Robinson (1992) suggest that the test is not based on the use of spatial weights, but only the existence of a "spatial ordering," which determines the pairs i, j to be included in the auxiliary regression. In practice, this often boils down to the use of contiguity. This may be problematic, since it is well established that the spatial covariance (on which the test is based) only corresponds to the spatial weights for *local* patterns of spatial correlation. For example, this is the case for the spatial error components model covered in Section 3.4 (see Kelejian and Robinson 1995, p. 89), but not for a SAR alternative (for further discussion, see Anselin and Moreno 2003, pp. 599–600).

The KR test is a large sample test, and its asymptotic properties are not well reflected in small sample situations. This is evidenced in a number of Monte Carlo simulation studies (e.g., Anselin and Florax 1995b, Anselin and Moreno 2003, Florax and de Graaff 2004).

6.2 Tests Based on Maximum Likelihood

In contrast to diffuse spatial autocorrelation tests, "focused" tests are constructed with a specific alternative in mind. In general, they boil down to a test of restrictions on the parameters of a model that includes spatial dependence, such as a spatial error model or a spatial lag model. The most commonly used

approach is based on the three classic test statistics obtained under Maximum Likelihood estimation: the Wald, Likelihood Ratio and Lagrange Multiplier (or, Rao Score) tests.[24]

Both Wald test and Likelihood Ratio test are standard, and require estimation of the unrestricted or spatial model (for technical details, see Anselin 1988b, Ch. 6). The Lagrange Multiplier tests on the other hand are based on estimation under the null, or restricted model, i.e., the classic linear regression and its OLS residuals. In contrast to results in mainstream econometrics, the LM test statistics for spatial dependence cannot be obtained in an NR^2 form based on a simple auxiliary regression (for a recent comprehensive overview, see Anselin 2001a).

6.2.1 Spatial Error Autocorrelation

The point of departure for a LM test for spatial error autocorrelation is the log-likelihood (72) for a specific data generating process for the regression error terms, such as SAR or SMA. In the usual fashion, a LM test statistic is then obtained as:

$$LM = [d(\theta)]'[I(\theta)]^{-1}[d(\theta)], \tag{117}$$

where $d(\theta)$ is the familiar score, $\partial L(\theta)/\partial\theta$, and $I(\theta)$ is the information matrix, $-\mathrm{E}[\partial^2 L(\theta)/\partial\theta\partial\theta']$. Both score and information matrix are derived for the unrestricted (spatial) model, but evaluated under the null, i.e., with the restricted parameters. For the SAR and SMA alternatives, this is $\lambda = 0$ in (14), or $\gamma = 0$ in (24). Both restrictions lead to the same test statistic (Burridge 1980, Anselin 1988a):

$$LM_\lambda = \frac{[e'We/(e'e/n)]^2}{tr[W'W + WW]}\ , \tag{118}$$

where e is a $n \times 1$ vector of OLS residuals. Apart for the scaling factor in the denominator, this statistic is essentially the square of Moran's I. It is asymptotically distributed as $\chi^2(1)$.

Since the LM test exploits the slope of the log-likelihood under the null, it is possible that different alternatives result in the same slope. This is the case for the SAR and SMA alternatives, which are termed "locally equivalent alternatives" or LEA (Godfrey 1981). As a consequence, it is not possible to distinguish between these two alternatives using a LM test.

Traditionally, the asymptotic properties for the LM test are obtained for a complete specification of the log-likelihood, including an assumption of normality. This turns out to be overly restrictive, and the same asymptotic properties can also be derived without assuming normality, using appropriate CLT (for technical details, see Anselin and Kelejian 1997, Kelejian and Prucha 2001).

[24] In the remainder, the LM/RS test statistics will be referred to as LM for notational simplicity, although they can equally be considered RS statistics.

The LM principle can also be applied to other one-directional spatial error alternatives, such as the error components model (32) or direct representation (28). The resulting test statistic for the error components model is (Anselin 2001a, Anselin and Moreno 2003):

$$LM_{SEC} = \left[\frac{e'WW'e}{(e'e/n)} - T_1\right]^2 / 2\left[T_2 - \frac{T_1^2}{n}\right] \tag{119}$$

where $T_1 = tr(WW')$ and $T_2 = tr(WW'WW')$. The statistic is asymptotically distributed as $\chi^2(1)$.[25]

A test statistics for the direct representation model derived from ML principles is more complex, since under the null, the parameter value is on the boundary of the parameter space. Moreover, the nuisance parameters are only identified under the alternative hypothesis. These non-standard conditions invalidate the use of LR or Wald statistics. The LM statistic can still be employed, but it requires an approximation due to Davies (1977, 1987), applied to the direct representation model in Anselin (2001a, pp. 130–132).

An alternative strategy, still based on ML principles, applies the idea of double length artificial regressions (DLR, Davidson and MacKinnon 1984, 1988) to the spatial models, as outlined in Baltagi and Li (2001a). The DLR approach consists of expressing the regression model as a function of standard normal error terms. In the spatial models, this follows as a simple standardization. Furthermore, Baltagi and Li (2001a) exploit the eigenvalue decomposition of the log-Jacobian term (79) to obtain the contribution of each observation to the log-likelihood as:

$$l_i(y_i, \phi) = -(1/2)\ln(2\pi) - (1/2)f_i^2(y_i, \phi) + k_i(y_i, \phi), \tag{120}$$

with $\phi = [\beta, \lambda, \sigma]'$ as a vector of parameters, f_i as the standard normal error term:

$$f_i(y_i, \phi) = (1/\sigma)[(y_i - x_i'\beta) - \lambda \sum_j w_{ij}(y_j - x_j'\beta)], \tag{121}$$

and k_i as the complete log-Jacobian (including the error variance):

$$k_i(y_i, \phi) = \ln(1 - \lambda\omega_i) - \ln\sigma, \tag{122}$$

where ω_i are the eigenvalues of the spatial weights matrix. The DLR consists of an artificial regression with $2n$ "observations":

$$\begin{bmatrix} f(y, \phi) \\ \iota_n \end{bmatrix} = \begin{bmatrix} -F(y, \phi) \\ K(y, \phi) \end{bmatrix} b + u, \tag{123}$$

in which the u are unspecified residuals, ι_n is a $n \times 1$ vector of ones, and F and K are $n \times (k+2)$ matrices of partial derivatives $F_{ij}(y_i, \phi) \equiv \partial f_i(y_i, \phi)/\partial\phi_j$,

[25] An alternative test statistic against the SEC model can be based on an auxiliary regression and a significance test on the estimated variance component (see Kelejian and Robinson 1993, p. 304).

and $K_{ij}(y_i, \phi) \equiv \partial k_i(y_i, \phi)/\partial \phi_j$. Using the expressions (121) and (122) for the spatial SAR error model, this yields:

$$\begin{bmatrix} e/\hat{\sigma} \\ \iota_n \end{bmatrix} = \begin{bmatrix} (1/\hat{\sigma})X & (1/\hat{\sigma})We & e/\hat{\sigma}^2 \\ 0 & -\omega & -\iota_n/\hat{\sigma} \end{bmatrix} b + u, \tag{124}$$

with e as the OLS residuals, $\hat{\sigma}$ as the OLS estimate for the disturbance standard error, and ω as a $n \times 1$ vector of the eigenvalues ω_i. The test statistic is computed as $2n$ less the residual sum of squares in the artificial regression. It is asymptotically distributed as $\chi^2(1)$. While the artificial regression is easy to carry out, in practice it suffers from the necessity to compute the eigenvalues of the spatial weights matrix, which is problematic in large samples (see Section 5.2.3).

6.2.2 Spatial Lag

Using the same principle as in (117), applied to the log-likelihood of the spatial lag model (61), and with the constraint $\rho = 0$, yields a LM test statistic for spatial lag dependence, as shown in Anselin (1988a):

$$LM_\rho = [e'Wy/(e'e/n)]^2/D, \tag{125}$$

with e as the OLS residuals, and the denominator term:

$$D = [(WX\hat{\beta})'[I - X(X'X)^{-1}X'](WX\hat{\beta})/\hat{\sigma}^2] + tr(W'W + WW), \tag{126}$$

where the estimates for $\hat{\beta}$ and $\hat{\sigma}^2$ are from OLS. The test statistics is asymptotically distributed as $\chi^2(1)$.

Similarly, the application of the DLR principle to the log-likelihood and log-Jacobian terms of the spatial lag model, yields the artificial regression as (Baltagi and Li 2001a):

$$\begin{bmatrix} e/\hat{\sigma} \\ \iota_n \end{bmatrix} = \begin{bmatrix} (1/\hat{\sigma})X & (1/\hat{\sigma})Wy & e/\hat{\sigma}^2 \\ 0 & -\omega & -\iota_n/\hat{\sigma} \end{bmatrix} b + u, \tag{127}$$

where, again, all estimates are based on OLS. The test statistic ($2n$ less the the residual sum of squares) is also asymptotically distributed as $\chi^2(1)$.

6.2.3 Higher Order Models

So far, the tests considered are for one-directional alternatives. When the alternative is of a higher order, such as the models reviewed in Section 4.2, different test strategies may be pursued. One is a so-called *marginal* test, which only focuses on one source of misspecification at a time, ignoring the other, such as the tests covered so far. A second perspective is that of a *joint* test, where the null hypothesis is to set all spatial parameters equal to zero. For example, for the spatial lag model with a SAR error term (53), $H_0 : \rho = \lambda = 0$.[26]

[26] Note that since the SAR error and SMA error are LEA, either alternative may be considered.

In contrast to standard results, the joint test statistic is not simply the sum of the marginal test statistics, i.e., $LM_{\lambda\rho} \neq LM_\lambda + LM_\rho$, but it takes on a far more complex form (Anselin 1988a). To simplify notation, set $d_\rho = (e'Wy)/(e'e/n)$ and $d_\lambda = (e'We)/(e'e/n)$ and use D as in (126), but with W_1 is the spatial weights matrix pertaining to the lag model. As before, all estimates are from OLS. Further, using W_2 for the spatial weights in the error process, take $T_{ij} = tr[W_iW_j + W_i'W_j]$. The LM test statistic for the joint null hypothesis is given by:

$$LM_{\rho\lambda} = \frac{\left[d_\lambda^2 D + d_\rho^2 T_{22} - 2d_\lambda d_\rho T_{12}\right]}{DT_{22} - T_{12}^2}. \tag{128}$$

This test statistic is asymptotically distributed as $\chi^2(2)$. This will result in a loss of power relative to the proper one-directional test when only one source of misspecification is present. In the commonly used simplification where $W_1 = W_2$, the test statistic becomes:

$$LM_{\rho\lambda} = \frac{d_\lambda^2}{T} + \frac{(d_\lambda - d_\rho)^2}{D - T}. \tag{129}$$

Extensions of these principles to joint tests in SARMA(p, q) models are straightforward, and can be found in Anselin (2001a).

Yet a third testing strategy for higher order alternatives is to take a *conditional* approach, where a test on the null hypothesis $\rho = 0$ is carried out in a model with $\lambda \neq 0$, and vice versa. This can no longer be based on OLS estimates, but requires estimation of the proper spatial model by means of ML. For example, a test statistic for $H_0 : \lambda = 0$ in the spatial lag model is obtained as:

$$LM_{\lambda|\rho} = \frac{d_\rho^2}{T_{22} - (T_{21A}^2 \widehat{Var(\hat{\rho})})}, \tag{130}$$

where all the estimates are obtained using ML in the spatial lag model, $T_{21A} = tr[(W_2W_1 + W_2'W_1)(I - \hat{\rho}W)^{-1}]$, and the other notation is as before. The test statistic is asymptotically distributed as $\chi^2(1)$. It forms the ML-based counterpart to the Moran's I for S2SLS residuals outlined in equations (108)–(109). A test statistic for $H_0 : \rho = 0$ in the spatial error model, $LM_{\rho|\lambda}$ can be developed along the same lines, but turns out to be quite complex (for technical details, see Anselin 1988a, Anselin et al. 1996, Anselin and Bera 1998).

6.2.4 Robust Tests

In a higher order model, the parameter that is not of interest can be considered to be a nuisance parameter. Unfortunately, the one-directional test statistics LM_λ and LM_ρ become non-central χ^2 in the presence of local misspecification in the form of the other type of spatial dependence. In other words, in the presence of spatial lag dependence, the LM_λ test against error correlation becomes biased, and, in the presence of spatial error dependence, the LM_ρ test against lag

dependence becomes biased. Using a result of Bera and Yoon (1993), a robust version of these test statistics can be developed, as shown in Anselin et al. (1996) (see also Anselin and Bera 1998, pp. 273–278).

Limiting attention to the case where $W_1 = W_2 = W$, the expression for the robustified test against spatial error dependence is:

$$LM_\lambda^* = \frac{\left(d_\lambda - TD^{-1}d_\rho\right)^2}{[T(1-TD)]}, \tag{131}$$

using the same notation as before, and with all estimates obtained from OLS in the classic regression. The essence of the robustification is to correct the LM_λ statistic for the covariance between d_λ and d_ρ. The new test statistic is asymptotically distributed as $\chi^2(1)$.

Similarly, a robustified test against a spatial lag alternative takes the form:

$$LM_\rho^* = \frac{(d_\rho - d_\lambda)^2}{(D-T)}, \tag{132}$$

and is also asymptotically distributed as $\chi^2(1)$.

An interesting result is the decomposition of the two-directional test statistic (129). As pointed out, this is not the sum of the two one-directional tests, but instead:

$$LM_{\rho\lambda} = LM_\rho + LM_\lambda^* = LM_\lambda + LM_\rho^*, \tag{133}$$

or, the sum of the one-directional statistic for one source of misspecification and the robustified version of the test statistic for the second source (for technical details, see Anselin et al. 1996).

6.3 Tests Against Multiple Sources of Misspecification

In Sections 6.1 and 6.2, the misspecifications considered consisted of forms of spatial dependence only. In practice, cross-sectional data are also likely to be affected by other sources of misspecification, such as heteroskedasticity and nonlinearity. Testing for heteroskedasticity in the presence of spatial dependence is straightforward. As shown in Anselin (1988b), a Breusch-Pagan (BP) type LM test statistic (Breusch and Pagan 1979, 1980) readily extends to the ML-residuals in a spatial lag model, or to the spatially filtered residuals in the ML estimation of a spatial error model.[27]

A *joint* test for heteroskedasticity and spatial error dependence consists of the sum of a BP statistic and the LM_λ (118) (Anselin 1988b). An alternative is the extended Kelejian-Robinson statistic outlined in Kelejian and Robinson (1998), which does not require an assumption of normality and applies equally to linear and nonlinear regressions. For notational simplicity, consider the scalar

[27] In Kelejian and Robinson (1998, p. 395) the estimated spatial autoregressive coefficient is included in a test statistic along the lines of the expression (114) in Section 6.1.2 (see also Kelejian and Robinson 2004).

case only, in which the heteroskedasticity is modeled as $\sigma_i^2 = g(z_i)$, where z is one of the explanatory variables in the regression specification (this generalizes to z being a subset of the explanatory variables). The resulting test statistic is obtained as a joint test of significance on the slope coefficients in an auxiliary regression (Kelejian and Robinson 1998, p. 393):

$$e_i e_j = a_0 \delta_{ij} + a_1(\delta_{ij} z_i) + a_2(z_i w_{ji} + z_j w_{ij}) + u, \tag{134}$$

with the terms $e_i e_j$ including all the squared residuals and all cross-products for $j > i$ and $(w_{ji} + w_{ij}) \neq 0$, $\delta_{ii} = 1$, and $\delta_{ij} = 0$ for $i \neq j$. The joint significance test on $\hat{a_1} = \hat{a_2} = 0$ uses a heteroskedastic-consistent (White) covariance estimator. The extension of the simple scalar case to a situation with multiple z variables requires nonlinear least squares (for details, see Kelejian and Robinson 1998, pp. 394–395).

A *conditional* test for spatial error autocorrelation in a heteroskedastic regression model can be derived from the LM-test principles. With $\hat{\Omega}$ as the estimated (diagonal) heteroskedastic variance matrix, e.g., with elements $\hat{\sigma}_i = g(\hat{\alpha}, z_i)$ (where the $\hat{\alpha}$ are consistent estimates, say obtained in a Maximum Likelihood estimation), the heteroskedastic LM statistic becomes (Anselin 1988b, p. 107):

$$LM_{H\lambda} = \frac{(e'\hat{\Omega}^{-1}We)^2}{tr(WW + W'\hat{\Omega}^{-1}W\hat{\Omega})}\ , \tag{135}$$

with e as the residuals in the heteroskedastic regression. The $LM_{H\lambda}$ test statistic is asymptotically distributed as $\chi^2(1)$.[28]

The joint presence of spatial correlation and functional misspecification has received some attention as well. In Baltagi and Li (2001b), the LM principle is applied to derive a test for a general Box-Cox alternative with spatial error dependence, as well as conditional tests for functional form given spatial error dependence and vice versa. The test statistics do not reduce to simple analytical forms, but require computing the expression (117) with the proper elements for the score and information matrix substituted in each special case. de Graaff et al. (2001) consider the extension of a general misspecification test from chaos theory in time series to the spatial domain in order to assess dependence, heterogeneity and nonlinearity.

6.4 Specification Search

In practice, the sheer number of available test statistics can seem overwhelming and a strategy needs to be developed to move from the null model to a superior alternative (when appropriate). Given that tests may be based on marginal, joint or conditional approaches, the results of a specification search may be subject to the order in which tests are carried out, and whether or not adjustments

[28] A similar, heteroskedasticity robust version of Moran's I is derived in Kelejian and Robinson (2004, pp. 89-90).

are made for pre-testing (see, e.g., Florax and Folmer 1992, Anselin and Florax 1995b, Florax and de Graaff 2004).

Based on a large number of simulation results, an ad hoc decision rule was suggested in Anselin and Rey (1991) for the simple case of choosing between a spatial lag or spatial (SAR) error alternative. There is considerable evidence that the proper alternative is most likely the one with the largest significant LM test statistic value. This was later refined in light of the robust forms of the statistics in Anselin et al. (1996) to a sequence where in the first step the significance of the LM_λ and LM_ρ test statistics was assessed. The robust forms LM_λ^* and LM_ρ^* are only considered when *both* LM_λ *and* LM_ρ are significant. At that point, the largest, most significant value of the robust statistics suggests the most likely alternative.

In a recent paper by Florax et al. (2003) this classic forward stepwise specification search is compared to a "general-to-simple" model selection rule. In the backward stepwise approach, the point of departure is a spatial common factor (or spatial Durbin) model, (43). Failure to reject the factor constraints suggests a spatial error model. Alternatively, rejection of the factor constraints suggests a spatial lag model. If the autoregressive coefficient in the lag model is not significant, the final model selection is the standard regression specification. In simulation experiments, the forward and backward specification searches are compared and some evidence is provided of better performance by the forward strategy (for further discussion, see also Florax et al. 2005, Hendry 2005).

7 Spatial Effects in Panel Data

Up to this point, the focus in this chapter has been on models for a single cross-section. Early efforts to incorporate spatial effects in a panel data setting, with observations over time as well as across space were described in Anselin (1988b, Chapter 10). More recently, this has received growing attention, also in the mainstream econometric literature. For example, the second edition of Baltagi's well known panel data text now includes a brief discussion of *spatial panels* (Baltagi 2001, pp. 195–197).

The topic of spatial panels is too broad ranging to be covered satisfactorily within a single chapter section, and by necessity the treatment here will be limited to a few salient issues (for a recent and extensive review, see Anselin et al. 2005). First, the generic aspects of incorporating spatial effects in pooled cross-section and time series models will be considered. This is followed by a brief review of specification, estimation and diagnostic testing in two special models that have received considerable attention in empirical practice, the error components model with spatial effects and the spatial seemingly unrelated regression (SUR) model.[29]

[29] More extensive reviews of the recent literature can also be found in Elhorst (2001), and Elhorst (2003).

7.1 Spatial Effects in Pooled Cross-Section and Time Series Models

7.1.1 Model Specification

The basic specification for a model containing observations in the time domain as well as across space can be given as:

$$y_{it} = x'_{it}\beta \ + \varepsilon_{it}, \tag{136}$$

where i is an index for the cross-sectional dimension, with $i = 1, \ldots, n$, and t is an index for the time dimension, with $t = 1, \ldots, T$. Using customary notation, y_{it} is an observation on the dependent variable at i and t, x_{it} a $k \times 1$ vector of observations on the (exogenous) explanatory variables, β a matching $k \times 1$ vector of regression coefficients, and ε_{it} an error term. The setting considered here is where the cross-sectional dimension dominates, with $n \gg T$. Also, even though the basic design is referred to as "space" and "time," the second dimension could equally pertain to different cross-sections, such as in a study of industrial sectors or household types. In stacked matrix form, the simple pooled regression then becomes:

$$y = X\beta + \varepsilon, \tag{137}$$

with y as a $nT \times 1$ vector, X as a $nT \times k$ matrix and ε as a $nT \times 1$ vector. Note that in order to incorporate spatial effects, the stacking is for a complete cross-section at a time, and not for each individual cross-section over time.

The models (136)–(137) refer to homogeneous panels, in that the intercept and slope coefficients are assumed to be constant over the cross-sectional units. In contrast, in a heterogeneous panel, these coefficients are specific to each cross-sectional unit:

$$y_{it} = \alpha_i + x'_{it}\beta_i + \varepsilon_{it}. \tag{138}$$

Common approaches to deal with the heterogeneity are fixed effects and random effects models, which are not further considered in detail here (specific examples are covered in Sections 7.3 and 7.2). For a technical discussion of spatial effects in fixed and random effects models, see, e.g., Elhorst (2003) and Anselin et al. (2005).

Spatial effects can easily be introduced into the homogeneous panel by straightforward extension of the models considered in Sections 3 and 4. This is readily accomplished by generalizing the notion of a cross-sectional n-dimensional spatial weights matrix, W_n, to the panel dimension nT. Typically, it is assumed that the weights remain constant over time. Consequently, the $nT \times nT$ dimension simplifies to:

$$W_{nT} = I_T \otimes W_n, \tag{139}$$

where the subscripts refer to the matrix dimension and $\otimes$ is the Kronecker product.

A spatial lag model can then be expressed as:

$$y = \rho(I_T \otimes W_N)y + X\beta + \epsilon, \tag{140}$$

where ρ is the spatial autoregressive parameter (constant over the time dimension), and the other notation is as before. Similarly, a model with spatial SAR error dependence results in an $nT \times nT$ non-spherical error variance-covariance matrix of the form:

$$\Sigma_{NT} = \sigma_u^2 \left[I_T \otimes (B_N' B_N)^{-1} \right], \tag{141}$$

where $B_n = I_n - \lambda W_n$, σ_u^2 is a common variance term, and the spatial autoregressive coefficient λ is assumed to be constant over the time dimension. More complex model specifications can be introduced in the same fashion (see Anselin et al. 2005).

7.1.2 Estimation

Estimation of the spatial lag and spatial error specifications for homogeneous panels can be carried out by direct extension of the methods covered in Sections 5.2 and 5.3.

For example, consider maximum likelihood estimation and the log-likelihood for the spatial lag model, (61), as the point of departure. Its counterpart in the panel setting requires the generalization of the log-Jacobian term to $\ln |I_T \otimes (I_n - \rho W_N)| = T \ln |I_n - \rho W_n|$, which yields:

$$L = T \ln |I_n - \rho W_n| - \frac{1}{2} \ln |\Sigma_{nT}| - \frac{1}{2} \varepsilon' \Sigma_{nT}^{-1} \varepsilon, \tag{142}$$

with $\varepsilon = y - \rho(I_T \otimes W_N)y - X\beta$, and Σ_{nT} as a general $nT \times nT$ error variance-covariance matrix. A special case of particular interest in practice is the group-wise heteroskedasticity, considered in Section 5.2.1. This allows each time period to have a separate error variance. The extension of ML estimation for panel spatial error models follows in the same fashion, for example, using (141) as the expression for the error variance-covariance matrix (see Anselin et al. 2005).

The principles of instrumental variables estimation and method of moments outlined in Section 5.3 can be extended to the pooled case as well, taking advantage of the spatial weights $I_T \otimes W_n$. For example, the instruments in an IV estimation of the spatial lag model would be $(I_T \otimes W_n)X$ (with X as a stacked $nT \times (k-1)$ matrix, excluding the constant term). Similarly, the Kelejian-Prucha GM estimator (Kelejian and Prucha 1999), considered in Section 5.3.2 can be generalized, by replacing the single equation spatial weights by their pooled counterparts. Specifically, consider a stacked vector of SAR errors:

$$\varepsilon = \lambda(I_T \otimes W_n)\varepsilon + u, \tag{143}$$

where both ε and u are $nT \times 1$ vectors, and $u \backsim \text{IID}[0, \sigma_u^2 I_{nT}]$.

Extending the moment conditions for the idiosyncratic errors u from (90) to this case yields:

$$\begin{aligned} \mathrm{E}[\frac{1}{nT}u'u] &= \sigma_u^2 && (144) \\ \mathrm{E}[\frac{1}{nT}u'(I_T \otimes W_n')(I_T \otimes W_n)u] &= \frac{1}{N}\sigma_u^2 tr(W_n'W_n) && (145) \\ \mathrm{E}[\frac{1}{nT}u'(I_T \otimes W_n)u] &= 0, && (146) \end{aligned}$$

where tr is the matrix trace operator and use is made of $tr(I_T \otimes W_n'W_n) = TtrW_n'W_n$, and $tr(I_T \otimes W_n) = 0$.

The estimator is made operational by substituting $u = \epsilon - \lambda(I_t \otimes W_n)\epsilon$, and replacing ϵ by the regression residuals. The result is a system of three equations in λ, λ^2 and σ_u^2, which can be solved in the same fashion as for the single cross-section case.

Extension to the estimation of spatial effects in heterogeneous panel models are outlined in Elhorst (2003) and Anselin et al. (2005).

7.1.3 Specification Testing

Specification tests in pooled models also generalize directly from the single cross-section case. For example, the LM statistics against a spatial error and spatial lag alternative are readily obtained as:

$$LM_\lambda = \frac{[e'(I_T \otimes W_n)e/(e'e/nT)]^2}{Ttr(W_nW_n + W_n'W_n)}, \tag{147}$$

and

$$LM_\rho = \frac{[e'(I_T \otimes W_N)y/(e'e/nT)]^2}{[(W\hat{y})'M(W\hat{y})/\hat{\sigma}^2] + Ttr(W_nW_n + W_n'W_n)}, \tag{148}$$

with e as a $nT \times 1$ vector of OLS residuals, $W\hat{y} = (I_T \otimes W_n)X\hat{\beta}$ as the spatially lagged predicted values in the regression, and $M = I_{nT} - X(X'X)^{-1}X'$. Both statistics are asymptotically distributed as $\chi^2(1)$, since the spatial parameter is constrained to remain constant over time.

Extensions incorporating groupwise heteroskedasticity can be obtained using the principles outlined in Section 6.3, and robustified versions can be derived using extensions of the results from Section 6.2.4.

In heterogeneous panels, a test statistic that does not require the specification of a spatial weights matrix was recently suggested by Pesaran (2004). It consists of an average of cross-sectional residual correlations, based on the residuals of individual-specific regressions, $e_{it} = y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}_i$. The test statistic is obtained as:

$$CD = \sqrt{\frac{2T}{n(n-1)}}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\gamma}_{ij}\right), \tag{149}$$

with the pairwise correlation coefficients as:

$$\hat{\gamma}_{ij} = \sum_{t=1}^{T} \frac{e_{it}e_{jt}}{(e_i'e_i)^{1/2}(e_j'e_j)^{1/2}}, \tag{150}$$

where e_i and e_j are $T \times 1$ residual vectors for each cross-sectional unit. The statistic is asymptotically distributed as standard normal. Extensions to "local" forms of cross-sectional dependence are given as well (see Pesaran 2004, for details).

7.2 Spatial Effects in Error Component Models

In the textbook case of the two-way error component regression model, heterogeneity is introduced to the panel setup through random effects. The error term ε_{it} contains a cross-sectional unobserved random effect (α_i), a random time effect (ϕ_t), as well as an idiosyncratic component u_{it} (e.g., Baltagi 2001, p. 31):

$$\varepsilon_{it} = \alpha_i + \phi_t + u_{it}. \tag{151}$$

The cross-sectional component has variance σ_α^2, the time component has variance σ_ϕ^2, and the idiosyncratic error term is assumed to be $i.i.d$ with variance σ_u^2. Furthermore, the three random components are assumed to be zero mean and to be uncorrelated with each other. The random components α_i are assumed to be uncorrelated across cross-sectional units, and the components ϕ_t are assumed to be uncorrelated across time periods. This model is standard, except that in a spatial econometric context, the data are stacked as cross-sections for different time periods. For each time period $t = 1, \ldots, T$, the $n \times 1$ cross-sectional error vector ε_t is:

$$\varepsilon_t = \alpha + \phi_t \iota_n + u_t, \tag{152}$$

where α is a $n \times 1$ vector of cross-sectional error components, ϕ_t is the scalar time component for time t, ι_n is a $n \times 1$ vector of ones, and u_t is a $n \times 1$ vector of idiosyncratic errors. The common time component (and not the cross-sectional component) yields a particular type of cross-sectional correlation of the form $\sigma_\phi^2 \iota_n \iota_n'$. This equicorrelation across space is different from the usual spatial autocorrelation, since no distance decay is present. Stacking the error vectors by cross-section for each time period yields the complete $nT \times 1$ error vector as (see also Anselin 1988b, p. 153):

$$\varepsilon = (\iota_T \otimes I_n)\alpha + (I_T \otimes \iota_n)\phi + u, \tag{153}$$

where the subscripts indicate the dimensions, ϕ is a $T \times 1$ vector of time error components, u is a $nT \times 1$ vector of idiosyncratic errors, and the other notation is as before. The overal error variance-covariance matrix then follows as:

$$\Sigma_{nT} = \sigma_\alpha^2(\iota_T \iota_T' \otimes I_n) + \sigma_\phi^2(I_T \otimes \iota_n \iota_n') + \sigma_u^2 I_{nT}. \tag{154}$$

This differs slightly from the standard textbook notation, due to the stacking of observations by cross-section.

A spatially lagged dependent variable can be added to the regression part of this model in the usual fashion. However, given possible identification problems between the spatial correlation induced by the lag term and the spatial correlation that results from the time random component, this is typically only considered for the one way error component model:

$$\varepsilon = (\iota_T \otimes I_n)\alpha + u. \tag{155}$$

This model does not contain the time component ϕ_t and its error variance-covariance is a subset of (154), without the second term.

A one-way error component model spatial lag model is a special case of the spatial lag specification for pooled data considered in Section 7.1.1. Using the appropriate expression for the error variance-covariance (and exploiting some standard results on matrix Kronecker products), the log likelihood for this model follows as (omitting constants):

$$\begin{aligned} L = \quad & T \ln |I_n - \rho W_n| - \tfrac{1}{2} \ln |\sigma_\alpha^2 (\iota_T \iota_T' \otimes I_n) + \sigma_u^2 I_{nT}| \\ & -\tfrac{1}{2} \epsilon' \left[\sigma_\alpha^2 (\iota_T \iota_T' \otimes I_n) + \sigma_u^2 I_{nT}\right]^{-1} \epsilon. \end{aligned} \tag{156}$$

Inference can be based on the maximum likelihood estimates.

Another way to introduce spatial dependence into the one-way error component specification is through the error term. Two different approaches have been suggested in the literature, each leading to a particular estimation strategy.

The first approach consists of the specification of a SAR process for the error component u_t in each cross-section, for $t = 1, \ldots, T$ (Anselin 1988b, p. 153,):

$$u_t = \lambda W_n u_t + \xi_t, \tag{157}$$

with λ as the spatial autoregressive parameter (constant over time), W_n as the spatial weights matrix, and ξ_t as an $i.i.d$ idiosyncratic error term with variance σ_ξ^2.

Using the notation $B_n = I_n - \lambda W_n$, and $u_t = B_n^{-1} \xi_t$, the complete $nT \times 1$ error vector ε for the stacked observations is:

$$\varepsilon = (\iota_T \otimes I_n)\alpha + (I_T \otimes B_n^{-1})\xi. \tag{158}$$

The matching variance-covariance matrix for ϵ is then:

$$\Sigma_{nT} = \mathrm{E}[\epsilon\epsilon'] = \sigma_\alpha^2 (\iota_T \iota_T' \otimes I_n) + \sigma_\xi^2 [I_T \otimes (B_n' B_n)^{-1}]. \tag{159}$$

Note that the first component induces correlation in the time dimension, but not in the cross-sectional dimension, whereas the opposite holds for the second component (correlation only in the cross-sectional dimension).

Maximum likelihood estimation is again a special case of the linear model with a nonspherical variance-covariance matrix, treated in Section 5.2.2. To operationalize this method, use is made of some matrix Kronecker product properties to simplify the complex structure in (159). Specifically, with $\eta = \sigma_\alpha^2 / \sigma_\xi^2$,

the error variance-covariance can be rewritten as $\Sigma_{nT} = \sigma_\xi^2 \Psi_{nT}$, where:

$$\Psi_{nT} = \iota_T \iota_T' \otimes \eta I_n + [I_T \otimes (B_n' B_n)^{-1}]. \tag{160}$$

This allows the determinant and inverse to be obtained as (see Anselin 1988b, p. 154, for details):

$$|\Psi_{nT}| = |(B_n' B_n)^{-1} + (T\eta) I_n| |B_n|^{-2(T-1)} \tag{161}$$

and,

$$\Psi_{nT}^{-1} = \frac{\iota_T \iota_T'}{T} \otimes [(B_n' B_n)^{-1} + (T\eta) I_n]^{-1} + (I_T - \frac{\iota_T \iota_T'}{T}) \otimes (B_n' B_n). \tag{162}$$

The log-likelihood thus becomes:

$$\begin{aligned} L \quad = \quad & -\frac{nT}{2} \ln \sigma_\xi^2 - (T-1) \ln |B_n| \\ & -\frac{1}{2} \ln |(B_n' B_n)^{-1} + (T\eta) I_n| \\ & -\frac{1}{2\sigma_\xi^2} \varepsilon' [\frac{\iota_T \iota_T'}{T} \otimes [(B_n' B_n)^{-1} + (T\eta) I_n]^{-1}] \varepsilon \\ & -\frac{1}{2\sigma_\xi^2} \varepsilon' [(I_T - \frac{\iota_T \iota_T'}{T}) \otimes (B_n' B_n)] \varepsilon, \end{aligned} \tag{163}$$

with $\varepsilon = y - X\beta$. Note how, in contrast to the standard SAR error model, the log-likelihood involves the inverse $(B_n' B_n)^{-1}$, which constitutes a serious computational challenge.

The likelihood framework for the one-way error component model with spatial error autocorrelation can be exploited to derive LM tests for a range of misspecifications (Anselin 1988b, Baltagi et al. 2003a,b). Taking as a point of departure the standard pooled specification (137), both the error component and the spatial parameter can be considered as part of the null hypothesis, and a number of interesting combinations result. The corresponding tests can be classified as marginal, joint or conditional, depending on which combinations of parameters restrictions are considered (Baltagi et al. 2003b).

Specifically, marginal tests would be on either $\mathrm{H}_0 : \lambda = 0$ (the spatial parameter) or on $\mathrm{H}_0 : \sigma_\alpha^2 = 0$ (the error component), based on the residuals of the pooled model. A joint test is on $\mathrm{H}_0 : \lambda = \sigma_\alpha^2 = 0$, and conditional tests are for $\mathrm{H}_0 : \lambda = 0$ (assuming $\sigma_\alpha^2 \geq 0$), or $\mathrm{H}_0 : \sigma_\alpha^2 = 0$ (assuming λ may or may not be zero). Each case yields a LM statistic using the standard principles applied to the proper likelihood function (for details, see Baltagi et al. 2003b). This rationale can be further extended to include a time-wise dependent process, as in Baltagi et al. (2003a). LM test statistics robust to local misspecification of one of the other forms can be developed along these lines as well. The derivation and evaluation of effective diagnostics for spatial effects in error component models is still an area of active research.

A second approach to introduce spatial error autocorrelation into an error components model was recently suggested by Kapoor et al. (2003). Instead of specifying a spatial process for the idiosyncratic error u in (155), the error vector (in the pooled model) is first subject to a SAR process:

$$\varepsilon = \lambda(I_T \otimes W_n)\varepsilon + u, \tag{164}$$

or:

$$\varepsilon = (I_T \otimes B_n^{-1})u, \tag{165}$$

and, as before, $B_n = I - \lambda W_n$. The error components are introduced into the error u, as:

$$u = (\iota_T \otimes I_n)\alpha + \nu. \tag{166}$$

This results in an overall error variance-covariance matrix:

$$\Sigma_{nT} = \mathrm{E}[\epsilon\epsilon'] = (I_T \otimes B_n^{-1})[\sigma_\alpha^2(\iota_T\iota_T' \otimes I_n) + \sigma_\nu^2 I_{nT}](I_T \otimes B_n^{-1'}). \tag{167}$$

Again, this model combines both time-wise as well as cross-sectional correlation. The middle term in (167) is the standard expression from the error components literature, which allows a simple solution to obtain the inverse of the matrix. This leads to a system of six moment conditions in three unknowns (the spatial parameter λ and the variance components) that can be solved to obtain consistent estimators for the parameters. This readily yields FGLS estimates for the model coefficients β (for technical details, see Kapoor et al. 2003).

Further work is needed to compare the relative merits of the two spatial error component formulations in empirical practice.

7.3 Spatial Seemingly Unrelated Regression Models (SUR)

A special case of temporal heterogeneity (different parameter values over time, but constant values across space) that has received considerable attention in empirical work (e.g., in regional economics, Rey and Montouri 1999) consists of a system of cross-sectional equations connected through cross-equation error correlation. This is a special case of fixed effects in that the number of cross-sections T is typically fixed. In the mainstream literature, such seemingly unrelated regressions are considered primarily for the case where $T \gg n$. In the spatial context, however, the more interesting design is where $n \gg T$ and the equations are stacked one cross-section at a time:

$$y_t = X_t\beta_t + \epsilon_t, \text{ for } t = 1, \ldots, T, \tag{168}$$

where β_t is a time period specific $k \times 1$ vector of regression coefficients. In practice, interest centers on the null hypothesis of homogeneity, $H_0 : \beta_t = \beta$ for $t = 1, \ldots, T$. In each cross-section, the variance-covariance matrix is

$\mathrm{E}[\epsilon_t \epsilon_t'] = \sigma_t^2 I_n,$ for $t = 1, \ldots, T$. The temporal covariance matrix between the cross-section at t and the cross-section at s takes the general form:

$$\mathrm{E}[\epsilon_t \epsilon_s'] = \sigma_{ts} I_n, \text{for} \quad s \neq t, \tag{169}$$

where σ_{ts} is the temporal covariance between s and t (by convention, the variance terms are expressed as σ_t^2). In stacked form (T cross-sections), the model becomes:

$$y = X\beta + \epsilon, \tag{170}$$

with

$$\mathrm{E}[\epsilon \epsilon'] = \Sigma_T \otimes I_n \tag{171}$$

and Σ_T is the $T \times T$ temporal covariance matrix with elements σ_{ts}.

A spatial lag SUR model is obtained by introducing a spatially lagged dependent variable with a time specific autoregressive coefficient ρ_t for each period, with the cross-equation error covariance as in (171). Each individual cross-section for time period t follows the model:

$$y_t = \rho_t W_n y_t + X_t \beta_t + \epsilon_t, \text{ for } t = 1, \ldots, T. \tag{172}$$

The full system can be expressed concisely in stacked form as:

$$y = (R_T \otimes W_n) y + X\beta + \epsilon, \tag{173}$$

where R_T is a $T \times T$ diagonal matrix containing the time-specific autoregressive coefficients ρ_t on the diagonal, and β is a $kT \times 1$ vector of stacked time-specific coefficient vectors. The matrix X is a $nT \times kT$ block diagonal matrix:

$$X = \begin{pmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & X_T \end{pmatrix}. \tag{174}$$

In a spatial lag SUR model, special interest goes to testing the null hypothesis $H_0 : \rho_t = \rho \ \forall \ t = 1, \ldots, T$, i.e., whether the spatial process governing the cross-sectional dependence in each time period is stable over time.

Maximum likelihood estimation of the spatial lag SUR model can be considered a special case of the general log-likelihood (61), but with a log-Jacobian of $\ln |I_{nT} - (R_T \otimes W_n)|$ and an error variance-covariance matrix of $\Sigma_T \otimes I_n$. The block diagonal structure of the Jacobian can be exploited to simplify this expression to $\sum_t \ln |I_n - \rho_t W_n|$ (with the sum over $t = 1, \ldots, T$). The log-likelihood then follows as (omitting the constant part):

$$L = \sum_t \ln |I_n - \rho_t W_n| - \frac{n}{2} \ln |\Sigma_T| - \frac{1}{2} \varepsilon' \left(\Sigma_T^{-1} \otimes I_n \right) \varepsilon, \tag{175}$$

with $\varepsilon = [I_{nT} - (R_T \otimes W_n)]y - X\beta$ (for further details, see Anselin 1988b, pp. 145–146). Instrumental variables estimation of this model is a special case of three stage least squares (see Anselin et al. 2005).

Spatial error autocorrelation is introduced into the SUR specification in each time period, with the $n \times 1$ error vector ε_t as:

$$\varepsilon_t = \lambda_t W_n \varepsilon_t + u_t, \text{for} \quad t = 1, \ldots, T, \tag{176}$$

or,

$$\varepsilon_t = (I_n - \theta_t W_n)^{-1} u_t. \tag{177}$$

The cross-equation covariance is specified through the remainder error term u_t, with $\mathrm{E}[uu'] = \Sigma_T \otimes I_N$, where Σ_T contains elements σ_t^2 on the diagonal, and has σ_{ts} as off-diagonal entries. As a result, the overall error covariance becomes:

$$\mathrm{E}[\epsilon\epsilon'] = B_{nT}^{-1}(\Sigma_T \otimes I_n)B_{nT}^{-1'}. \tag{178}$$

with $B_{nT} = [I_{nT} - (\lambda_T \otimes W_N)]$ and λ_T as a $T \times 1$ vector of time specific spatial autoregressive coefficients λ_t.

Maximum likelihood estimation follows as a special case of a linear model with non-spherical error variance-covariance, along the lines presented in Section 5.2.2. Similarly, specification tests for the presence of spatial error autocorrelation in the SUR model can be based on the LM principle (Anselin 1988c).

8 Spatial Effects in Models with Limited Dependent Variables

In applied econometrics, interest often focuses on models where the dependent variable is not continuous, as considered so far in this chapter, but instead takes on a limited number of values, or is truncated or censored in some fashion. The extension of spatial effects to such models is not straightforward and still a very active area of research. In spatial econometrics, the main interest in this context has been in the *spatial probit* model, where the multivariate normal distribution provides a flexible framework to incorporate spatial correlation. This will be the focus of the current section, where specification, estimation and diagnostic testing in this model will be briefly reviewed in turn. Recent, more in-depth treatment of spatial probit can be found in Pinkse and Slade (1998), Novo (2001), Fleming (2004), and Beron and Vijverberg (2004). Illustrative applications are Holloway et al. (2002), Beron et al. (2003) and Murdoch et al. (2003).

Before proceeding with the spatial probit, it is worthwhile to briefly point out some alternative approaches, mostly formulated in the statistical literature. The almost exclusive emphasis on the normal distribution as the stochastic framework to handle spatial correlation for discrete dependent variables is because most other distributions do not provide an analytical link between the marginal

distribution (at each location) and the joint distribution, without impractical constraints on the correlation structure (Johnson and Kotz 1969, Chapter 11). Early on, some conditional models were suggested by Besag (1974), such as auto-logistic and auto-Poisson, but these constrain the spatial range of the dependence as well as the parameter space. For example, the auto-Poisson model only allows negative spatial autocorrelation, precluding its use for modeling diffusion or contagion processes. Approaches to allow a relaxation of this requirement have only recently been developed (for a technical discussion and extensions, see Kaiser and Cressie 1997, 2000).

Alternatives to the auto models that incorporate spatial correlation in specifications for limited dependent variables have been suggested that are based on generalized linear models (GLM) and generalized linear mixed models (GLMM). These utilize specialized estimators, such as quasi-likelihood and estimating equations (see, e.g., Gotway and Stroup 1997, Gotway and Wolfinger 2003, Zhang 2002). In addition, there is a growing literature on hierarchical random coefficient models from a Bayesian perspective, where spatial correlation is often introduced through a conditional autoregressive (CAR) prior on one of the model parameters (such as the mean in a Poisson count model). A detailed discussion of these methods is beyond the scope of the current chapter. Recent reviews can be found in Lawson (2001), Banerjee et al. (2004), and Waller and Gotway (2004), among others.

8.1 Spatial Probit

The textbook model for a linear latent variable regression specifies an unobserved (latent) dependent variable y_i^* as a linear function of an "index function" and a random error term:

$$y_i^* = x_i'\beta + \varepsilon_i, \tag{179}$$

with $x_i\beta$ as the index function, where x_i is a $k \times 1$ vector of observations on the explanatory variables, and β is a matching vector of coefficients. The observed counterpart of y_i^*, the discrete dependent variable y_i, equals one for $y_i^* > 0$ and zero otherwise. Interest therefore centers on $\mathrm{P}[y_i^* > 0] = \mathrm{P}[x_i'\beta + \varepsilon_i > 0]$. By specifying a distribution for the random error term, estimates for β can be obtained. In the probit model, the standard normal distribution is used, which due to its symmetry, yields the familiar result:

$$\mathrm{P}[y_i = 1] = \mathrm{P}[y_i^* > 0] = \mathrm{P}[\varepsilon_i < x_i'\beta] = \Phi[x_i'\beta], \tag{180}$$

where Φ is the cumulative density function for a standard normal random variate.

Spatial dependence is introduced into this model either through the latent variable itself, with $\mathrm{Cov}[y_i^* y_j^*] \neq 0$ for "neighboring" i, j, or through the error term, with $\mathrm{Cov}[\varepsilon_i \varepsilon_j] \neq 0$ for "neighboring" i, j. Following standard practice, the neighbors are defined by means of a spatial weights matrix.

A spatial lag model in the latent variable becomes:

$$y_i^* = \rho \sum_j w_{ij} y_j^* + x_i' \beta + \varepsilon_i, \tag{181}$$

or, in matrix notation, using the counterpart of the reduced form (40):

$$y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon. \tag{182}$$

In this *simultaneous* model, the latent variables are jointly determined, both by the values for x at their own location and by the values for x at all other locations, subject to a distance decay effect. This is quite distinct from a *conditional* specification, where the observed actions of the neighbors enter on the right hand side, not the unobserved (and unobservable) latent values.

The conditional counterpart of (181) is:

$$y_i^* = \rho \sum_j w_{ij} \mathrm{E}[y_j^* | X] + x_i' \beta + \varepsilon_i, \tag{183}$$

where $\mathrm{E}[y_j^* | X]$ can be estimated by $y_i^L = \sum_j w_{ij} y_j$, the average of the observed decisions y_j at the neighboring locations. The new variable y_i^L can be included on the right hand side of the model as an exogenous variable. Consequently, the neighboring observations cannot be jointly determined with the observations a i. While the standard probit model remains valid, coding methods must be employed to ensure that the sample does not contain these neighbors.

In contrast, the simultaneous model (182) causes two major complications. First, the error term is no longer independent, nor homoskedastic, since $u = (I - \rho W)^{-1} \varepsilon$ has variance-covariance matrix $[(I - \rho W)'(I - \rho W)]^{-1}$. The induced heteroskedasticity (when the number of neighbors is not constant across observations) will cause standard probit estimators to be inconsistent. A standardization must be carried out that takes into account the location-specific variance. The correlation structure means that to obtain the marginal cumulative density for each individual error term u_i, the remaining $n - 1$ dimensions need to be integrated out. Second, the inequality constraint on the random error does not pertain to $x_i \beta$, but to a series expansion, containing both ρ and β, as well as the spatially lagged x. With

$$G(X, W, \beta, \rho) = x_i' \beta + \rho [WX\beta]_i + \rho^2 [W^2 X\beta]_i + \ldots \tag{184}$$

as the familiar series expansion, the censoring condition becomes:

$$\mathrm{P}[y_i = 1] = \mathrm{P}\left[u_i < \frac{G_i(X, W, \beta, \rho)}{\sigma_i} \right], \tag{185}$$

where the cumulative density is for the marginal distribution of u_i obtained by integrating out the $n-1$ other dimensions in a multivariate normal density, and σ_i is the square root of the heteroskedastic variance at i.

The probit model with error terms that follow a SAR process is represented by the latent variable specification (179) with the error terms as:

$$\varepsilon = \lambda W \varepsilon + u \tag{186}$$

This also requires the marginal density for a multivariate error term in the censoring condition. However, unlike (185), this does not involve a function G, but instead:

$$\mathrm{P}[y_i = 1] = \mathrm{P}\left[\, u_i < \frac{x_i'\beta}{\sigma_i} \right], \tag{187}$$

where the probability for u_i is for the marginal density from a multivariate normal distribution with variance-covariance matrix $[(I - \lambda W)'(I - \lambda W)]^{-1}$.

8.2 Estimation of Spatial Probit

The textbook solution to estimation in the probit model consists of a straightforward application of the maximum likelihood principle. Each observation on the discrete dependent variable y_i can be considered an independent draw for a binomial random variable with probability $\Phi(x_i'\beta)$. The log-likelihood readily follows as:

$$\ln L = \sum_i [\, y_i \ln \Phi(x_i'\beta) + (1 - y_i) \ln(1 - \Phi(x_i'\beta))], \tag{188}$$

with $y_i = 0, 1$. Since $1 - \Phi(x_i'\beta) = \Phi(-x_i'\beta)$, and using $q_i = 2y_i - 1$, this can be expressed succintly as:

$$\ln L = \sum_i \ln \Phi(q_i x_i'\beta). \tag{189}$$

In the spatial case, the simple summation as in (189) is no longer appropriate and the full multivariate density must be evaluated to obtain the log-likelihood. Consider u as the $n \times 1$ vector of multivariate normal random variables with variance-covariance matrix Σ. In order to generalize the censoring conditions for both values of y_i, set Q as a diagonal matrix with diagonal elements q_i defined above. The multivariate censoring conditions (or, the bounds on the integrals in the evaluation of the multivariate cumulative density) are:[30]

$$u < QX\beta, \tag{190}$$

for the spatial error probit model, and:

$$u < Q(I - \rho W)^{-1} X\beta \tag{191}$$

for the spatial lag probit model. These must be evaluated in a multivariate normal distribution with Σ respectively as $[(I - \lambda W)'(I - \lambda W)]^{-1}$ or $[(I - \rho W)'(I - \rho W)]^{-1}$. The corresponding log-likelihood can be expressed as:

$$\ln L = \ln \Phi_n[QX\beta; 0, \Sigma_\lambda] \tag{192}$$

[30] The notation used is similar to the suggestion made in Beron and Vijverberg (2004).

for the spatial error probit model, and:

$$\ln L = \ln \Phi_n[Q(I - \rho W)^{-1} X\beta; 0, \Sigma_\rho], \tag{193}$$

for the spatial lag probit model, where Φ_n stands for an n-dimensional multivariate normal cumulative distribution function with upper bounds as the first term, mean 0 and variance-covariance matrix Σ. The evaluation of this log-likelihood involves the computation of n-dimensional integrals, which is infeasible in practice.

A crucial step in any numerical optimization of the log-likelihood is the evaluation of the probabilities involved. Beron and Vijverberg (2004) outline a simulation estimator for the spatial probit model based on the relative importance sampler (RIS) for an n-dimensional multivariate normal density (see also Vijverberg 1997, Beron et al. 2003).[31]

The problem consists of evaluating the multivariate normal probability $\mathrm{P}[u < V]$, with V as either $QX\beta$ or $Q(I - \rho W)^{-1} X\beta$, and $u \sim \mathrm{MVN}(0, \Sigma)$. Since Σ is positive definite, there exists a Choleski decomposition $A'A = \Sigma^{-1}$, where A is an upper triangular matrix. This is particularly useful for SAR models, since Σ^{-1} does not involve an inverse operation. For example, in the spatial error probit, this requires a Choleski decomposition of $(I - \lambda W)'(I - \lambda W) = I + \lambda(W + W') + \lambda^2 WW'$.

The transformation $\eta = Au$ yields a vector of independent standard normal random variables whose joint density can be computed as a simple product of the marginal densities. With $B = A^{-1}$ (an upper triangular matrix with positive diagonal elements), the random vector u can be replaced by $u = A^{-1}\eta = B\eta$. Since the matrix B is upper triangular and its diagonal elements are positive, the bounds on the random vector η can be written, starting from the bottom (i.e., for $i = n$), as:

$$\eta_n < b_{nn}^{-1} V_n \equiv \nu_n \tag{194}$$

$$\eta_{n-1} < b_{n-1n-1}^{-1}[V_{n-1} - b_{n-1n}\eta_n] \equiv \nu_{n-1}, \tag{195}$$

and, in general:

$$\eta_j < b_{jj}^{-1} \left[V_j - \sum_{i=j+1}^{n} b_{ji}\eta_i \right] \equiv \nu j. \tag{196}$$

The values for V_j are based on the current value of the model parameters, including the spatial autoregressive parameter. For a given choice of a proper density function for the importance sampler, the simulator operates in a recursive fashion, by first drawing a random variable $\bar{\eta}_n$ that satisfies the condition $\bar{\eta}_n < \nu_n$. The value $\bar{\eta}_n$ is next substituted in the expression for the bound ν_{n-1} to obtain $\bar{\nu}_{n-1}$. A a second random variable, $\bar{\eta}_{n-1}$ is drawn, satisfying this new bound. This is continued until a $n \times 1$ vector of bounds $\bar{\nu}$ is obtained. This

[31] The RIS estimator is a generalized version of the well known GHK simulation estimator. For a review, see Stern (1997).

vector is then used to evaluate the joint density, as a product of the individual cumulative densities (technically, these are ratios of densities in the importance sampler). In the case of a normal sampler, this simplifies to $\prod_{j=1}^{n} \Phi(\bar{\nu}_j)$. This process is repeated R times to obtain an estimate for the log-likelihood as the average of the sampled joint probabilities (for technical details, see Beron and Vijverberg 2004, pp. 176–177):

$$\hat{p} = (1/R) \sum_{r=1}^{R} \left[\prod_{j=1}^{n} \Phi(\bar{\nu}_{j,r}) \right]. \tag{197}$$

This is then used in a nonlinear optimization algorithm to obtain a new set of parameter values, and the process is repeated until convergence. Two important computational challenges in this procedure are the Choleski decomposition of the inverse variance matrix, and the need for the inverse $(I - \rho W)^{-1}$, required for the bounds in the spatial lag model. The RIS estimator for the spatial probit model was applied in empirical work by Beron et al. (2003) and Murdoch et al. (2003).

An alternative simulation estimator for the spatial probit model was formulated by LeSage (1999, 2000). This uses Markov Chain Monte Carlo (MCMC) simulation methods, such as the Gibbs sampler and the Metropolis-Hastings sampler, and is couched in a Bayesian context (see also Holloway et al. 2002). The basic principle of a Gibbs sampler is to generate draws for the joint posterior distribution of the parameters by sampling from the full conditional distributions. For example, in a model with two parameters, $[\theta_1, \theta_2]$, a sample from the joint density can be obtained by alternating repeated draws from $f(\theta_1|\theta_2)$ and $f(\theta_2|\theta_1)$, using the sampled variate at each iteration to condition.[32] In this process it is important to establish a proper posterior conditional density from which it easy to sample.

In the spatial probit model, two key aspects need to be considered. First, the analysis proceeds similar to the Bayesian framework for the standard spatial regression models, conditional upon the latent dependent variable y^*. The latter is obtained from its posterior conditional density, which can be shown to be a truncated normal distribution (given values for the other parameters in the model). Sampling from this distribution generates "observations" on the latent variable, which can then be conditioned on in the remainder of the analysis.[33] A second aspect pertains to the conditional density for the spatial parameters. While the conditional densities for the β parameters (in a homoskedastic probit model, $\sigma^2 = 1$ to avoid identification problems) are the usual multivariate normal, the conditional density for the spatial parameters is of unknown form (see

[32] A technical discussion of the Gibbs sampler and other MCMC techniques is beyond the current scope. A classic reference is Geman and Geman (1984). For introductory overviews, see, e.g., Casella and George (1992), Gilks et al. (1996), and Gelman et al. (2004, Chapter 11). A basic reference for application to the probit model is Albert and Chib (1993).

[33] This is similar in spirit to the EM algorithm (expectation-maximization) which was applied in the spatial probit context by McMillen (1992). See LeSage (2000) and Fleming (2004) for a critical assessment.

also LeSage 1997). For example, in the spatial lag model (ignoring σ^2), this is:

$$f(\rho|\beta, y^*) \propto |I - \rho W| \exp[(y^* - \rho W y^* - X\beta)'(y^* - \rho W y^* - X\beta)]. \quad (198)$$

Sampling from this conditional density requires a specialized approach, such as a Metropolitan-Hastings algorithm. This also involves the evaluation of the Jacobian term (for technical details, see LeSage 1997, 1999, 2000).

Alternatives to these simulation estimators can be based on GMM, as in Pinkse and Slade (1998). However, these approaches only tackle the additional heteroskedasticity in the spatial models, but ignore the off-diagonal elements in the variance-covariance matrix. Specifically, the heteroskedasticity is a function of the spatial autoregressive parameters, which can be exploited to adjust the correction by σ_i in (185) and (187).[34]

To date, relatively little is known about the relative performance of these estimators in finite samples (some preliminary findings are given in Novo 2001).

8.3 Specification Tests in Spatial Probit

Testing for the presence of spatial effects in the probit model is hampered by the fact that neither the true residual nor the dependent variable in the latent variable model are observed. Instead of $e_i^* = y_i^* - x_i'\hat{\beta}$, either a naive residual must be used, $e_i = y_i - \Phi(x_i'\hat{\beta})$, or the notion of a generalized residual (Cox and Snell 1968). The latter follows from the first order conditions in the probit log-likelihood (188):

$$\partial L/\partial\beta = \sum_i \phi_i \left[\frac{y_i}{\Phi_i} - \frac{1-y_i}{1-\Phi_i}\right] x_i = 0, \quad (199)$$

where Φ_i and ϕ_i are, respectively, the cumulative standard normal distribution and the standard normal density, evaluated at $x_i'\beta$. This takes on the general form of the familiar orthogonality condition $\sum_i \hat{\varepsilon}_i x_i = 0$, used as the basis for GMM estimation. After some algebra, the *generalized residual* can be expressed as:

$$\hat{\varepsilon}_i = \hat{\phi}_i \left[\frac{y_i - \hat{\Phi}_i}{\hat{\Phi}_i(1-\hat{\Phi}_i)}\right], \quad (200)$$

with a consistent estimate $\hat{\beta}$ used to evaluate $\hat{\Phi}_i$ and $\hat{\phi}_i$.

The main results obtained so far are developed in papers by Pinkse and Slade (1998), Pinkse (1999, 2004), and Kelejian and Prucha (2001). A test statistic against spatial error correlation in the probit model takes on the form of a generalized Moran's I statistic or a LM statistic, asymptotically distributed as, respectively, standard normal or $\chi^2(1)$:

$$I^* = e'We/\Gamma \xrightarrow{d} N(0,1), \quad (201)$$

[34] A GMM approach as a weighted non-linear estimator of the linear probability model is outlined in Fleming (2004).

or:

$$LM_\lambda = [e'We]^2/\Gamma \xrightarrow{d} \chi^2(1), \tag{202}$$

where e is a vector of "residuals" and Γ is a standardization factor, containing weights matrix traces and variance terms.[35]

In Pinkse and Slade (1998), the residuals are standardized generalized residuals, to correct for the inherent heteroskedasticity in the model. Since the variance of the generalized residual $\hat{\varepsilon}_i$ is $\phi_i^2/[\Phi_i(1-\Phi_i)]$, dividing (200) by the square root of this variance yields the standardized residual as:

$$\hat{\varepsilon}_i^s = \frac{y_i - \hat{\Phi}_i}{\sqrt{\hat{\Phi}_i(1-\hat{\Phi}_i)}}. \tag{203}$$

The Pinkse-Slade LM test statistic is then:

$$LM_{PS} = \frac{(\hat{\varepsilon}_i^s{\prime}W\hat{\varepsilon}_i^s)^2}{tr(WW + W'W)}. \tag{204}$$

The asymptotic distribution is not formally derived, but instead a bootstrap procedure is suggested to carry out inference (see Pinkse and Slade 1998, p. 131). In contrast, both Pinkse (1999) and Kelejian and Prucha (2001) obtain formal asymptotic results, under slightly different sets of assumptions.[36] Pinkse (1999) takes a LM perspective and demonstrates that:

$$LM_\lambda = \frac{(\hat{\varepsilon}_i{\prime}W\hat{\varepsilon}_i)^2}{\hat{\sigma}^4 tr(WW + W'W)} \xrightarrow{d} \chi^2(1), \tag{205}$$

where $\hat{\varepsilon}_i$ are the generalized residuals from (200), and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i \frac{\hat{\phi}_i^2}{\hat{\Phi}_i(1-\hat{\Phi}_i)}. \tag{206}$$

Kelejian and Prucha (2001)derive the asymptotic properties of the generalized Moran's I statistic in a wide range of specifications. For the probit model, the corresponding test statistic is (Kelejian and Prucha 2001, pp. 234–236):

$$I^* = \frac{e_i{\prime}We_i}{\sqrt{tr(W\Sigma W\Sigma + W'\Sigma W\Sigma)}} \xrightarrow{d} N(0,1), \tag{207}$$

[35]Tests against a spatial lag alternative have received much less attention. In Pinkse (1999), a statistic of the form $y^*{\prime}We/\Gamma$ is suggested, along the lines of the LM_ρ statistic in the linear regression. The latent dependent variable is estimated as $\hat{y}_i^* = x_i'\hat{\beta} + \hat{\varepsilon}_i$, with $\hat{\varepsilon}_i$ as in (200). The standardizing factor Γ is a complex expression in matrix traces and variance terms (see Pinkse 1999, p.11). To date, this test statistics has seen little application.

[36]For example, Pinkse (1999) assumes the explanatory variables to be independently distrubuted, whereas Kelejian and Prucha (2001) focus on moment conditions and constraints on the spatial weights.

with $e_i = y_i - \hat{\Phi}_i$, and Σ as a diagonal matrix containing $\hat{\sigma}_i^2 = \hat{\Phi}_i(1 - \hat{\Phi}_i)$.

The I^* and LM_λ test statistics differ in the regularity conditions under which they are developed. There are operational distinctions as well, in the way in which the residuals are estimated (generalized vs. naive) and in the expression for the variance used in the standardization factor. Little is known about the relative performance of these test statistics in finite samples (for limited evidence, see Novo 2001).

9 Future Directions

The objective of this chapter was to review the current state of the art in spatial econometrics, while emphasizing recent results. There is now a solid body of results in the literature to deal with spatial effects in the linear regression model. However, considerable work remains to be done in the context of panel data and limited dependent variable models. To close the chapter, it may be useful to outline a number of directions where progress is most desired.

There is still no encompassing theoretical framework to handle asymptotics in the spatial domain that translates into readily verifiable conditions. Whereas several useful results have been obtained to date, they tend to be case-specific and depend on particular views of the spatial sampling process. As a result, properties that hold for one conceptual framework do not readily translate into another.

There also seems to be considerable room for further integration of the insights obtained in the statistical literature on hierarchical Bayesian spatial models into spatial econometric formulations with random coefficients. Along the same lines, the simulation estimators that are crucial in obtaining results in these models require efficient computational algorithms. In order to become useful in the empirical (large data) context commonly encountered in applied economics, important computational issues remain to be tackled, especially to handle space-time dynamics.

Finally, the dissemination of theoretical results to the empirical practice of econometricians crucially depends on the availability of software. While considerable progress has been made (see the review in Anselin et al. 2004b, pp. 10–11), this has been primarily concerned with the linear regression setup, and much remains to be done to develop useful software tools to tackle panel data and limited dependent variable settings.

It is hoped that the current chapter will stimulate further work in both the theory and practice of spatial econometrics.

References

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, 65:1005–1027.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

Andrews, D. W. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:953–966.

Andrews, D. W. (2003). Cross-section regression with common shocks. Technical report, Cowles Foundation Discussion Paper No. 1428, Yale University, New Haven, CT.

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

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

Anselin, L. (1988b). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin, L. (1988c). A test for spatial autocorrelation in seemingly unrelated regressions. *Economics Letters*, 28:335–341.

Anselin, L. (1990). Spatial dependence and spatial structure instability in applied regression analysis. *Journal of Regional Science*, 30:185–207.

Anselin, L. (1992). Space and applied econometrics. Introduction. *Regional Science and Urban Economics*, 22:307–316.

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. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267.

Anselin, L. (2003a). Spatial externalities. *International Regional Science Review*, 26(2):147–152.

Anselin, L. (2003b). Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

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

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

Anselin, L. and Florax, R. J. (1995b). Small sample properties of tests for spatial dependence in regression models: Some further results. In Anselin, L. and Florax, R. J., editors, *New Directions in Spatial Econometrics*, pages 21–74. Springer-Verlag, Berlin.

Anselin, L., Florax, R. J., and Rey, S. J. (2004a). *Advances in Spatial Econometrics. Methodology, Tools and Applications*. Springer-Verlag, Berlin.

Anselin, L., Florax, R. J., and Rey, S. J. (2004b). Econometrics for spatial models, recent advances. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics. Methodology, Tools and Applications*, pages 1–25. Springer-Verlag, Berlin.

Anselin, L. and Griffith, D. A. (1988). Do spatial effects really matter in regression analysis. *Papers, Regional Science Association*, 65:11–34.

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., Le Gallo, J., and Jayet, H. (2005). Spatial panel econometrics. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice (3rd Edition)*. Kluwer, Dordrecht. (forthcoming).

Anselin, L. and Moreno, R. (2003). Properties of tests for spatial error components. *Regional Science and Urban Economics*, 33(5):595–618.

Anselin, L. and Rey, S. J. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, 23:112–131.

Anselin, L. and Smirnov, O. (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, 36:67–89.

Baltagi, B. H. (2001). *Econometric Analysis of Panel Data (Second Edition)*. John Wiley & Sons, Chichester, United Kingdom.

Baltagi, B. H. and Li, D. (2001a). Double length artificial regressions for testing spatial dependence. *Econometric Reviews*, 20(1):31–40.

Baltagi, B. H. and Li, D. (2001b). LM tests for functional form and spatial error correlation. *International Regional Science Review*, 24(2):194–225.

Baltagi, B. H., Song, S. H., Jung, B. C., and Koh, W. (2003a). Testing for serial correlation, spatial autocorrelation and random effects using panel data. Working paper, Texas A&M University, College Station, TX.

Baltagi, B. H., Song, S. H., and Koh, W. (2003b). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117:123–150.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton, FL.

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

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

Bera, A. and Yoon, M. J. (1993). Specification testing with misspecified alternatives. *Econometric Theory*, 9:649–658.

Beron, K. J., Murdoch, J. C., and Vijverberg, W. P. (2003). Why cooperate? Public goods, economic power, and the Montreal Protocol. *The Review of Economics and Statistics*, 85(2):286–297.

Beron, K. J. and Vijverberg, W. P. (2004). Probit in a spatial context: a Monte Carlo analysis. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*, pages 169–195. Springer-Verlag, Heidelberg.

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

Besley, T. and Case, A. (1995). Incumbent behavior: Vote-seeking, tax-setting and yardstick competition. *American Economic Review*, 85:25–45.

Bivand, R. and Szymanski, S. (1997). Spatial dependence through local yardstick competition: Theory and testing. *Economics Letters*, 55:257–265.

Blommestein, H. (1983). Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modeling. *Regional Science and Urban Economics*, 13:250–271.

Blommestein, H. (1985). Elimination of circular routes in spatial dynamic regression equations. *Regional Science and Urban Economics*, 15:121–130.

Blommestein, H. J. and Koper, N. A. (1998). The influence of sample size on the degree of redundancy in spatial lag operators. *Journal of Econometrics*, 82(2):317–333.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10:1047–1050.

Brandsma, A. and Ketellapper, R. H. (1979). A biparametric approach to spatial autocorrelation. *Environment and Planning A*, 11:51–58.

Brett, C. and Pinkse, J. (1997). Those taxes all over the map! A test for spatial independence of municipal tax rates in British Columbia. *International Regional Science Review*, 20:131–151.

Breusch, T. (1980). Useful invariance results for generalized regression models. *Journal of Econometrics*, 13:327–340.

Breusch, T. and Pagan, A. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, 47:1287–1294.

Breusch, T. and Pagan, A. (1980). The Lagrange Multiplier test and its applications to model specification in economics. *Review of Economic Studies*, 67:239–253.

Brock, W. and Durlauf, S. (2001a). Discrete choice with social interactions. *Review of Economic Studies*, 59:235–260.

Brock, W. A. and Durlauf, S. N. (2001b). Interactions-based models. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics, Volume 5*, pages 3297–3380. North-Holland, Amsterdam.

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(2):175–188.

Brueckner, J. K. and Saavedra, L. A. (2001). Do local governments engage in strategic tax competition? *National Tax Journal*, 54:203–229.

Burridge, P. (1980). On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B*, 42:107–108.

Burridge, P. (1981). Testing for a common factor in a spatial autoregressive model. *Environment and Planning A*, 13:795–800.

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.

Case, A. C., Rosen, H., and Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52:285–307.

Casella, G. and George, E. (1992). Explaining the Gibbs sampler. *American Statistician*, 46:167–174.

Casetti, E. (1997). The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review*, 20:9–33.

Cliff, A. and Ord, J. K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4:267–284.

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

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

Coakley, J., Fuentes, A.-M., and Smith, R. (2002). A principal components approach to cross-section dependence in panels. Working Paper, Department of Economics, Birkbeck College, University of London, London, United Kingdom.

Conley, T. G. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics*, 92:1–45.

Conley, T. G. and Ligon, E. (2002). Economic distance, spillovers and cross country comparisons. *Journal of Economic Growth*, 7:157–187.

Conley, T. G. and Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17:303–327.

Cook, D. and Pocock, S. (1983). Multiple regression in geographic mortality studies, with allowance for spatially correlated errors. *Biometrics*, 39:361–371.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society Series B*, 39:248–275.

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

Das, D., Kelejian, H. H., and Prucha, I. R. (2003). Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, 82:1–27.

Davidson, R. and MacKinnon, J. G. (1984). Model specification tests based on artificial regressions. *International Economic Review*, 25:485–502.

Davidson, R. and MacKinnon, J. G. (1988). Double-length artificial regression. *Oxford Bulletin of Economics and Statistics*, 50:203–217.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254.

Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74:33–43.

de Graaff, T., Florax, R. J. G. M., Nijkamp, P., and Reggiani, A. (2001). A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity. *Journal of Regional Science*, 41(2):255–276.

Dietz, R. D. (2002). The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research*, 31:539–575.

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, Steven, 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.

Durlauf, Steven, N. (2004). Neighborhood effects. In Henderson, J. and Thisse, J.-F., editors, *Handbook of Regional and Urban Economics, Volume 4*, pages 2173–2242. North Holland, Amsterdam.

Elhorst, J. P. (2001). Dynamic models in space and time. *Geographical Analysis*, 33:119–140.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3):244–268.

Fazekas, I. and Kukush, A. G. (2000). Infill asymptotics inside increasing domains for the least squares estimator in linear models. *Statistical Inference for Stochastic Processes*, 3:199–223.

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.

Fingleton, B. (1999). Spurious spatial regression: Some Monte Carlo results with spatial unit root and spatial cointegration. *Journal of Regional Science*, 39(1):1–19.

Fleming, M. (2004). Techniques for estimating spatially dependent discrete choice models. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*, pages 145–168. Springer-Verlag, Heidelberg.

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. and de Graaff, T. (2004). The performance of diagnostic tests for spatial dependence in linear regression models: A meta-analysis of simulation studies. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics. Methodology, Tools and Applications*, pages 29–65. Springer-Verlag, Berlin.

Florax, R. J., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33(5):557–579.

Florax, R. J., Folmer, H., and Rey, S. J. (2005). A comment on specification searches in spatial econometrics: The relevance of Hendry's methodology: A reply. *Regional Science and Urban Economics*, 35. (forthcoming).

Florax, R. J. and Rey, S. J. (1995). The impacts of misspecified spatial interaction in linear regression models. In Anselin, L. and Florax, R. J., editors, *New Directions in Spatial Econometrics*, pages 111–135. Springer-Verlag, Berlin.

Florax, R. J. G. M. and van der Vlist, A. (2003). Spatial econometric data analysis: Moving beyond traditional models. *International Regional Science Review*, 26(3):223–243.

Fotheringham, A., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression*. John Wiley and Sons, Chichester, UK.

Gamerman, D., Moreira, A. R., and Rue, H. (2003). Space-varying regression models: Specifications and simulation. *Computational Statistics and Data Analysis*, 42(3):513–533.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98:387–396.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis, Second Edition*. Chapman and Hall, Boca Raton, FL.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Getis, A., Mur, J., and Zoller, H. G. (2004). *Spatial Econometrics and Spatial Statistics*. Palgrave Macmillan, London.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, UK.

Glaeser, E. L., Kallal, H., Scheinkman, J., and Schleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100:1126–1152.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. (1996). Crime and social interactions. *Quarterly Journal of Economics*, 111:507–548.

Glaeser, E. L., Sacerdote, B. I., and Scheinkman, J. A. (2002). The social multiplier. Technical Report 9153, NBER, Cambridge, MA 02138.

Godfrey, L. (1981). On the invariance of the Lagrange Multiplier test with respect to certain changes in the alternative hypothesis. *Econometrica*, 49:1443–1455.

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 Wolfinger, R. D. (2003). Spatial prediction of counts and rates. *Statistics in Medicine*, 22:1415–1432.

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.

Griffith, D. A. and Sone, A. (1995). Trade-offs associated with normalizing constant computational simplications for estimating spatial statistical models. *Journal of Statistical Computation and Simulation*, 51:165–183.

Haining, R. (1988). Estimating spatial means with an application to remotely sensed data. *Communications in Statistics, Theory and Methods*, 17:573–597.

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

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, 99:399–424.

Hendry, D. F. (2005). A comment on specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics*, 35. (forthcoming).

Holloway, G., Shankar, B., and Rahman, S. (2002). Bayesian spatial probit estimation: A primer and an application to HYV rice adoption. *Agricultural Economics*, 27(3):383–402.

Hordijk, L. (1974). Spatial correlation in the disturbances of a linear interregional model. *Regional Science and Urban Economics*, 4:117–140.

Hordijk, L. (1979). Problems in estimating econometric relations in space. *Papers, Regional Science Association*, 42:99–115.

Hordijk, L. and Paelinck, J. (1976). Some principles and results in spatial econometrics. *Recherches Économiques de Louvain*, 42:175–197.

Hsiao, C. and Pesaran, M. H. (2004). Random coefficient panel data models. Working paper, University of Cambridge, Cambridge, United Kingdom.

Huang, J. (1984). The autoregressive moving average model for spatial analysis. *Australian Journal of Statistics*, 26:169–178.

Irwin, E. G. (2004). Identifying interaction effects among spatially distributed agents: An application to land use spillovers. Working paper, Department of Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH.

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

Jones, J.-P. and Casetti, E. (1992). *Applications of the Expansion Method*. Routledge, New York, NY.

Kaiser, M. S. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics and Probability Letters*, 35:423–432.

Kaiser, M. S. and Cressie, N. (2000). The construction of multivariate distributions from Markov random fields. *Journal of Multivariate Analysis*, 73:199–220.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2003). Panel data models with spatially correlated error components. Working paper, University of Maryland, College Park, MD.

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 Prucha, I. R. (2002). 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics*, 32(6):691–707.

Kelejian, H. H. and Prucha, I. R. (2003). HAC estimation in a spatial framework. Working paper, Department of Economics, University of Maryland, College Park, MD.

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118:27–50.

Kelejian, H. H., Prucha, I. R., and Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In LeSage, J. P. and Pace, R. K., editors, *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, pages 163–198. Elsevier Science Ltd., Oxford, UK.

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.

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

Kelejian, H. H. and Robinson, D. P. (2004). The influence of spatially correlated heteroskedacity on tests for spatial correlation. In Anselin, L. and Florax, R. J., editors, *Advances in Spatial Econometrics*, pages 79–97. Springer-Verlag, Heidelberg.

Kim, C.-W., Phipps, T. T., and Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management*, 45:24–39.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press, Princeton, NJ.

King, M. (1981). A small sample property of the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Association B*, 43:263–264.

Lahiri, S. (1996). On the inconsistency of estimators under infill asymptotics for spatial data. *Sankhya A*, 58:403–417.

Lahiri, S. and Mukherjee, K. (2004). Asymptotic distributions of M-estimators in a spatial regression model under some fixed and stochastic spatial sampling designs. *Annals Institute of Statistical Mathematics*, 56:225–250.

Land, K. and Deane, G. (1992). On the large-sample estimation of regression models with spatial or network-effect terms: A two stage least squares approach. In Marsden, P., editor, *Sociological Methodology*, pages 221–248. Jossey-Bass, San Francisco.

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.

Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22:307–335.

Leenders, R. T. A. J. (2002). Modeling social influence through network autocorrelation: Constructing the weights matrix. *Social Networks*, 24:21–47.

LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20:113–129.

LeSage, J. P. (1999). *Spatial Econometrics*. The Web Book of Regional Science, Regional Research Institute, West Virginia University, Morgantown, WV.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32:19–35.

LeSage, J. P. and Pace, R. K. (2004). *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*. Elsevier Science Ltd., Oxford, UK.

LeSage, J. P., Pace, R. K., and Tiefelsdorf, M. (2004). Methodological developments in spatial econometrics and statistics. *Geographical Analysis*, 36:87–89.

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.

Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, MA.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146.

Mardia, K. and Watkins, A. (1989). On multimodality of the likelihood for the spatial linear model. *Biometrika*, 76:289–295.

Martin, R. (1993). Approximations to the determinant term in Gaussian maximum likelihood estimation of some spatial models. *Communications in Statistics: Theory and Methods*, 22:189–205.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32:335–348.

Moran, P. (1948). The interpretation of statistical maps. *Biometrika*, 35:255–260.

Moran, P. (1950). A test for the serial dependence of residuals. *Biometrika*, 37:178–181.

Murdoch, J. C., Rahmatian, M., and Thayer, M. A. (1993). A spatially autoregressive median voter model of recreational expenditures. *Public Finance Quarterly*, 21:334–350.

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.

Murdoch, J. C., Sandler, T., and Vijverberg, W. P. (2003). The participation decision versus the level of participation in an environmental strategy: A Spatial Probit analysis. *Journal of Public Economics*, 87:337–362.

Nelson, G. C. (2002). Introduction to the special issue on spatial analysis. *Agricultural Economics*, 27(3):197–200.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Novo, A. A. (2001). *Spatial Probit Models: Statistical Inference and Estimation. Monte Carlo Simulations and an Application to Contagious Currency Crises*. PhD thesis, University of Illinois, Urbana-Champaign, Champaign, IL. Department of Economics.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126.

Pace, R. K. and Barry, R. (1997a). Quick computation of spatial autoregressive estimators. *Geographical Analysis*, 29:232–246.

Pace, R. K. and Barry, R. (1997b). Sparse spatial autoregressions. *Statistics and Probability Letters*, 33:291–297.

Pace, R. K. and LeSage, J. P. (2002). Semiparametric Maximum Likelihood estimates of spatial dependence. *Geographical Analysis*, 34:76–90.

Pace, R. K. and LeSage, J. P. (2004). Chebyshev approximation of log-determinants of spatial weights matrices. *Computational Statistics and Data Analysis*, 45:179–196.

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

Pesaran, M. H. (2002). Estimation and inference in large heterogenous panels with cross section dependence. DAE Working Paper 0305 and CESifo Working Paper no. 869, University of Cambridge, Cambridge, United Kingdom.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. Working paper, University of Cambridge, Cambridge, United Kingdom.

Pinkse, J. (1999). Asymptotics of the Moran test and a test for spatial correlation in Probit models. Working paper, Department of Economics, University of British Columbia, Vancouver, BC.

Pinkse, J. (2004). Moran-flavored tests with nuisance parameter. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*, pages 67–77. Springer-Verlag, Heidelberg.

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.

Pinkse, J., Slade, M. E., and Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, 70(3):1111–1153.

Rey, S. J. and Boarnet, M. G. (2004). A taxonomy of spatial econometric models for simultaneous equations systems. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg. 99-119.

Rey, S. J. and Montouri, B. D. (1999). US regional income convergence: A spatial econometrics perspective. *Regional Studies*, 33:143–156.

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

Royle, J. A. and Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological and Environmental Statistics*, 4(1):29–56.

Saavedra, L. A. (2000). A model of welfare competition with evidence from AFDC. *Journal of Urban Economics*, 47:248–279.

Smirnov, O. (2005). Computation of the information matrix for models with spatial interaction on a lattice. *Journal of Computational and Graphical Statistics*. (forthcoming).

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 Spatal Data, Some Theory for Kriging*. Springer-Verlag, New York.

Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, 35:2006–2039.

Stoker, T. M. (1993). Empirical approaches to the problem of aggregation over individuals. *Journal of Economic Literature*, 33:1827–1874.

Tiefelsdorf, M. (2002). The saddlepoint approximation of Moran's I and local Moran's I_i's reference distribution and their numerical evaluation. *Geographical Analysis*, 34:187–206.

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

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.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121:311–324.

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.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley, Hoboken, NJ.

Warnes, J. and Ripley, B. D. (1987). Problems with likelihood estimation of covariance functions of spatial gaussian processes. *Biometrika*, 74:640–642.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:434–449.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 56:129–136.