Chapter 18

SPATIAL PANEL ECONOMETRICS

Luc Anselin
*University of Illinois, Urbana-Champaign**

Julie Le Gallo
Université Montesquieu-Bordeaux IV

Hubert Jayet
Université de Lille

1. Introduction

Spatial econometrics is a subfield of econometrics that deals with the incorporation of *spatial effects* in econometric methods (Anselin, 1988a). Spatial effects may result from *spatial dependence*, a special case of cross-sectional dependence, or from *spatial heterogeneity*, a special case of cross-sectional heterogeneity. The distinction is that the *structure* of the dependence can somehow be related to location and distance, both in a geographic space as well as in a more general economic or social network space. Originally, most of the work in spatial econometrics was inspired by research questions arising in regional science and economic geography (early reviews can be found in, among others, Paelinck and Klaassen, 1979, Cliff and Ord, 1981, Upton and Fingleton, 1985, Anselin, 1988a, Haining, 1990, Anselin and Florax, 1995). However, more recently, spatial (and social) interaction has increasingly received more attention in mainstream econometrics as well, both from a theoretical as well as from an applied perspective (see the recent reviews and extensive references in Anselin and Bera, 1998, Anselin, 2001b, Anselin, 2002, Florax and Van Der Vlist, 2003, and Anselin et al., 2004a).

*The research by Luc Anselin and Julie Le Gallo was supported in part by the U.S. National Science Foundation Grant BCS-9978058 to the Center for Spatially Integrated Social Science (CSISS).

The central focus in spatial econometrics to date has been the single equation cross-sectional setting. However, as Arrelano argues in the introduction to his recent panel data econometrics text, “the field [of econometrics of panel data] has expanded to cover almost any aspect of econometrics” (Arellano, 2003, p. 2). It is therefore not surprising that this has included spatial econometrics as well. 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), and an increasing number of papers are devoted to the topic (see the reviews in Anselin, 2001b, Elhorst, 2001, and Elhorst, 2003, as well as the recent papers by Baltagi et al., 2003a, Baltagi et al., 2003b, Kapoor et al., 2003, and Pesaran, 2004, among others).

In this chapter, we review and organize this recent literature and emphasize a range of issues pertaining to the specification, estimation and diagnostic testing for spatial effects in panel data models. Since this encompasses a large and rapidly growing literature, we limit our attention to models with continuous dependent variables,[1] and to a design where the cross-sectional dimension (N) vastly exceeds the time dimension ($N \gg T$). We also avoid duplication by excluding aspects of the standard treatment of heterogeneity and dependence in panel data models, as well as the case where cross-sectional dependence is modeled by relying on the time dimension (e.g., as in the classic SURE case with fixed N, and some more recent extensions, such as Chen and Conley, 2001).

The chapter is organized into five remaining sections. First, we define the notion of *spatial effects* more precisely and provide a brief outline of how the traditional cross-sectional models can be extended to panel data model specifications. Next, we consider this more closely and develop a taxonomy of space-time models. We then turn to the issues of model estimation and diagnostic testing. We close with some concluding remarks.

2. Spatial Effects

As a point of departure, consider a simple pooled linear regression model:

$$y_{it} = x_{it}\beta + \epsilon_{it}, \tag{18.1}$$

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 $t = 1, \ldots, T$.[2] Using customary notation, y_{it} is an observation on the dependent variable at i and t, x_{it} a $1 \times K$ vector of observations on the (exogenous) explanatory variables, β a matching $K \times 1$ vector of regression coefficients, and ϵ_{it} an error term.

Given our interest in spatial effects, the observations will be stacked as successive cross-sections for $t = 1, \ldots, T$, referred to as y_t (a $N \times 1$ vector of cross-sectional observations for time period t), X_t (a $N \times K$ matrix of observations on a cross-section of the explanatory variables for time period t) and ϵ_t

(a $N \times 1$ vector of cross-sectional disturbances for time period t). In stacked form, the simple pooled regression then becomes:

$$y = X\beta + \epsilon, \tag{18.2}$$

with y as a $NT \times 1$ vector, X as a $NT \times K$ matrix and ϵ as a $NT \times 1$ vector.

In general, spatial dependence is present whenever correlation across cross-sectional units is non-zero, and the pattern of non-zero correlations follows a certain spatial *ordering*. When little is known about the appropriate spatial ordering, spatial dependence is reduced to simple cross-sectional dependence. For example, the error terms are spatially correlated when $\mathrm{E}[\epsilon_{it}\epsilon_{jt}] \neq 0$, for a given t and $i \neq j$, and the non-zero covariances conform to a specified *neighbor* relation. Note how the correlation is purely cross-sectional in that it pertains to the same time period t.

The neighbor relation is expressed by means of a so-called spatial weights matrix. We will briefly review the concept of spatial weights (and the associated spatial lag operator) and outline two classes of specifications for models with spatial dependence. In one, the spatial correlation pertains to the dependent variable, in a so-called *spatial lag* model, in the other it affects the error term, a so-called *spatial error* model. The two specifications can also be combined, resulting in so-called higher order spatial models. While these models and terms are by now fairly familiar in the spatial econometric literature, we thought it useful to briefly review them and to illustrate how they may be incorporated into a panel data setting.[3]

The second class of spatial effects, spatial heterogeneity, is a special case of the observed and unobserved heterogeneity which is treated prominently in the mainstream panel data econometrics literature. For example, a heterogeneous panel would relax the constant regression coefficient in equation 18.1, and replace it by:

$$y_{it} = x_{it}\beta_i + \epsilon_{it},$$

where the β_i is a $K \times 1$ vector of regression coefficients specific to the cross-sectional unit i.

This heterogeneity becomes *spatial* when there is a structure to the variability across the i that is driven by spatial variables, such as location, distance or region. In the spatial literature, discrete spatial variability is referred to as *spatial regimes* (Anselin, 1988a). The continuous case can be modeled as a special form of random coefficient variation (where the covariance shows a spatial pattern), or deterministically, as a function of extraneous variables (so-called spatial expansion, e.g., Casetti, 1997), or as a special case of local regression models (so-called geographically weighted regression, Fotheringham et al., 2002). Neither of these has seen application in panel data contexts.[4]

Since most econometric aspects of spatial heterogeneity can be handled by means of the standard panel data methods, we will focus the discussion that follows on spatial dependence and will only consider the heterogeneity when it is relevant to the modeling of the dependence.

2.1 Spatial Weights and Spatial Lag Operator

A spatial weights matrix W is a $N \times N$ positive matrix in which the rows and columns correspond to the cross-sectional observations. An element w_{ij} of the matrix expresses the prior strength of the interaction between location i (in the row of the matrix) and location j (column). This can be interpreted as the presence and strength of a link between nodes (the observations) in a network representation that matches the spatial weights structure. In the simplest case, the weights matrix is binary, with $w_{ij} = 1$ when i and j are neighbors, and $w_{ij} = 0$ when they are not. By convention, the diagonal elements $w_{ii} = 0$. For computational simplicity and to aid in the interpretation of the spatial variables, the weights are almost always standardized such that the elements in each row sum to 1, or, $w^s_{ij} = w_{ij} / \sum_j w_{ij}$.[5] A side effect of this standardization is that the sum of all elements in W equals N, the number of cross-sectional observations. Whereas the original weights are often symmetric, the row-standardized form is no longer, which is an unusual complication with significant computational consequences.

The specification of the spatial weights is an important problem in applied spatial econometrics.[6] Unless the weights are based on a formal theoretical model for social or spatial interaction, their specification is often ad hoc. In practice, the choice is typically driven by geographic criteria, such as contiguity (sharing a common border) or distance, including nearest neighbor distance (for examples and further discussion, see, e.g., Cliff and Ord, 1981, pp. 17–19, Anselin, 1988a, Chapter 3).

Generalizations that incorporate notions of "economic" distance are increasingly used as well (e.g., Case et al., 1993, Conley and Ligon, 2002, Conley and Topa, 2002). A slightly different type of economic weights are so-called block weights, where all observations in the same region are considered to be neighbors (and not only the adjoining observations). More formally, if there are N_g units in a block (such as counties in a state), they are all considered to be neighbors, and the spatial weights equal $1/(N_g - 1)$ for all observations belonging to the same block (see, e.g., Case, 1991, Case, 1992, and, more recently, Lee, 2002).

So far, the weights considered were purely cross-sectional. To extend their use in a panel data setting, they are assumed to remain constant over time.[7] Using the subscript to designate the matrix dimension, with W_N as the weights for the cross-sectional dimension, and the observations stacked as in equation 18.2, the full $NT \times NT$ weights matrix then becomes:

$$W_{NT} = I_T \otimes W_N, \tag{18.3}$$

with I_T as an identity matrix of dimension T.

Unlike the time series case, where "neighboring" observations are directly incorporated into a model specification through a shift operator (e.g., $t-1$), this is not unambiguous in a two dimensional spatial setting. For example, observations for irregular spatial units, such as counties or census tracts, typically do not have the same number of neighbors, so that a spatial shift operator cannot be implemented. Instead, in spatial econometrics, the neighboring observations are included through a so-called *spatial lag* operator, more akin to a distributed lag than a shift (Anselin, 1988a). In essence, a spatial lag operator constructs a new variable that consists of the weighted average of the neighboring observations, with the weights as specified in W. More formally, for a cross-sectional observation i for variable z, the spatial lag would be $\sum_j w_{ij} z_j$. In most applications, the bulk of the row elements in w_{ij} are zero (resulting in a *sparse* structure for W) so that in effect the summation over j only incorporates the "neighbors," i.e., those observations for which $w_{ij} \neq 0$. In matrix notation, this corresponds to the matrix operation $W_N y_t$, in which the $N \times N$ cross-sectional weights matrix is post-multiplied by a $N \times 1$ vector of cross-sectional observations for each time period $t = 1, \ldots, T$.

Spatial variables are included into a model specification by applying a spatial lag operator to the dependent variable, to the explanatory variables, or to the error term. A wide range of models for local and global spatial externalities can be specified in this manner (for a review, see Anselin, 2003). This extends in a straightforward manner to the panel data setting, by applying the $NT \times NT$ weights from equation 18.3 to the stacked y, X or ϵ from equation 18.2.

More precisely, in the same notation as above, a vector of spatially lagged dependent variables follows as:

$$Wy = W_{NT} y = (I_T \otimes W_N) y, \tag{18.4}$$

a matrix of spatially lagged explanatory variables as:

$$WX = W_{NT} X = (I_T \otimes W_N) X,$$

and a vector of spatially lagged error terms as:

$$W\epsilon = W_{NT} \epsilon = (I_T \otimes W_N) \epsilon.$$

The incorporation of these spatial lags into a regression specification is considered next.

2.2 Spatial Lag Model

A spatial lag model, or, mixed regressive spatial autoregressive model, includes a spatially lagged dependent variable on the RHS of the regression specification (Anselin, 1988a). While usually applied in a pure cross-sectional setting, it can easily be extended to panel models. Using the stacked equation 18.2

and the expression for the spatial lag from equation 18.4, this yields:

$$y = \rho(I_T \otimes W_N)y + X\beta + \epsilon, \tag{18.5}$$

where ρ is the spatial autoregressive parameter, and the other notation is as before.

In a cross-section, a spatial lag model is typically considered as the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of the neighboring agents.[8] This model is increasingly applied in the recent literature on social/spatial interaction, and is used to obtain empirical estimates for the parameters of a spatial reaction function (Brueckner, 2003) or social multiplier (Glaeser et al., 2002). It should be noted that other formulations to take into account social interaction have been suggested as well (e.g., Manski, 2000, Brock and Durlauf, 2001) mostly in the context of discrete choice. The modeling of complex neighborhood and network effects (e.g., Topa, 2001) requires considerable attention to identification issues, maybe best known from the work of Manski on the “reflection problem” (Manski, 1993). Because of this theoretical foundation, the choice of the weights in a spatial lag model is very important.

At first sight, the extension of the spatial lag model to a panel data context would presume that the equilibrium process at hand is stable over time (constant ρ and constant W). However, the inclusion of the time dimension allows much more flexible specifications, as outlined in section 18.3.

The essential econometric problem in the estimation of 18.5 is that, unlike the time series case, the spatial lag term is *endogenous*. This is the result of the two-directionality of the neighbor relation in space (“I am my neighbor’s neighbor”) in contrast to the one-directionality in time dependence (for details, see Anselin and Bera, 1998). The consequence is a so-called *spatial multiplier* (Anselin, 2003) which formally specifies how the joint determination of the values of the dependent variables in the spatial system is a function of the explanatory variables and error terms at all locations in the system.

The extent of the joint determination of values in the system can be seen by expressing equation 18.5 as a reduced form:

$$y = \left[I_T \otimes (I_N - \rho W_N)^{-1}\right] X\beta + \left[I_T \otimes (I_N - \rho W_N)^{-1}\right] \epsilon, \tag{18.6}$$

with the subscripts indicating the dimensions of the matrices. The inverse matrix expression can be expanded and considered one cross-section at a time, due to the block-diagonal structure of the inverse. A a result, for each $N \times 1$ cross-section at time $t = 1, \dots, T$:

$$y_t = X_t\beta + \rho W_N X_t\beta + \rho^2 W_N^2 X_t\beta + \dots + \epsilon_t + \rho W_N \epsilon_t + \rho^2 W_N^2 \epsilon_t \dots$$

The implication of this reduced form is that the spatial distribution of the y_{it} in each cross-section is determined not only by the explanatory variables and associated regression coefficients at each location ($X_t\beta$), but also by those at neighboring locations, albeit subject to a distance decay effect (the increasing powers of ρ and W_N). In addition, the unobserved factors contained in the error term are not only relevant for the location itself, but also for the neighboring locations ($W_N\epsilon$), again, subject to a distance decay effect. Note that in the simple pooled model, this spatial multiplier effect is contained within each cross-section and does not spill over into other time periods.[9]

The presence of the spatially lagged errors in the reduced form illustrates the joint dependence of the $W_N y_t$ and ϵ_t in each cross-section. In model estimation, this simultaneity must be accounted for through instrumentation (IV and GMM estimation) or by specifying a complete distributional model (maximum likelihood estimation).

Even without a solid theoretical foundation as a model for social/spatial interaction, a spatial lag specification may be warranted to spatially detrend the data. This is referred to as a *spatial filter*:

$$[I_T \otimes (I_N - \rho W_N)]\, y = X\beta + \epsilon, \tag{18.7}$$

with the LHS as a new dependent variable from which the effect of spatial autocorrelation has been eliminated. In contrast to time series, a simple detrending using $\rho = 1$ is not possible, since that value of ρ is not in the allowable parameter space.[10] As a consequence, the parameter ρ must be estimated in order for the spatial filtering to be operational (see Anselin, 2002).

2.3 Spatial Error Model

In contrast to the spatial lag model, a spatial error specification does not require a theoretical model for spatial/social interaction, but, instead, is a special case of a non-spherical error covariance matrix. An unconstrained error covariance matrix at time t, $\mathrm{E}[\epsilon_{it}\epsilon_{jt}], \forall\, i \neq j$ contains $N \times (N-1)/2$ parameters. These are only estimable for small N and large T, and provided they remain constant over the time dimension. In the panel data setting considered here, with $N \gg T$, structure must be imposed in order to turn the covariance matrix into a function of a manageable set of parameters.

Four main approaches have been suggested to provide the basis for a parsimonious covariance structure: direct representation, spatial error processes, spatial error components, and common factor models. Each will be reviewed briefly.

2.3.1 Direct Representation. The direct representation approach has its roots in the geostatistical literature and the use of theoretical variogram and covariogram models (Cressie, 1993). It consists of specifying the covariance between two observations as a *direct* function of the distance that separates them, $\forall\, i \neq j$ and $t = 1, \ldots, T$:

$$\mathrm{E}[\epsilon_{it}\epsilon_{jt}] = \sigma^2 f(\tau, d_{ij}), \tag{18.8}$$

where τ is a parameter vector, d_{ij} is the (possibly economic) distance between observation pairs i, j, σ^2 is a scalar variance term, and f is a suitable distance decay function, such as a negative exponential.[11] The parameter space for τ should be such that the combination of functional form and the distance metric ensures that the resulting covariance matrix is positive definite (for further discussion, see, e.g, Dubin, 1988).

An extension to a panel data setting is straightforward. With $\sigma^2\Omega_{t,N}$ as the error covariance matrix that results from applying the function 18.8 to the $N \times 1$ cross-sectional error vector in time period t, the overall $NT \times NT$ error variance-covariance matrix Σ_{NT} becomes a block diagonal matrix with the $N \times N$ variance matrix for each cross-section on the diagonal.[12] However, as specified, the function 18.8 does not vary over time, so that the result can be expressed concisely as:

$$\Sigma_{NT} = \sigma^2 \left[I_T \otimes \Omega_N \right],$$

with $\Omega_{t,N} = \Omega_N\ \forall t$.[13]

2.3.2 Spatial Error Processes. Whereas the direct representation approach requires a distance metric and functional form for the distance decay between a pair of observations, spatial error processes are based on a formal relation between a location and its *neighbors*, using a spatial weights matrix. The error covariance structure can then be derived for each specified process, but typically the *range* of neighbors specified in the model is different from the range of spatial dependence in the covariance matrix. This important aspect is sometimes overlooked in empirical applications.

In analogy to time series analysis, the two most commonly used models for spatial processes are the autoregressive and the moving average (for extensive technical discussion, see Anselin, 1988a, Anselin and Bera, 1998, Anselin, 2003, and the references cited therein).

A *spatial autoregressive* (SAR) specification for the $N \times 1$ error vector ϵ_t in period $t = 1, \ldots, T$, can be expressed as:

$$\epsilon_t = \theta W_N \epsilon_t + u_t,$$

where W_N is a $N \times N$ spatial weights matrix (with the subscript indicating the dimension), θ is the spatial autoregressive parameter, and u_t is a $N \times 1$ idiosyncratic error vector, assumed to be distributed independently across the cross-sectional dimension, with constant variance σ_u^2.

Continuing in matrix notation for the cross-section at time t, it follows that:

$$\epsilon_t = (I_N - \theta W_N)^{-1} u_t,$$

and hence the error covariance matrix for the cross-section at time t becomes:

$$\Omega_{t,N} = \mathrm{E}[\epsilon_t \epsilon_t'] = \sigma_u^2 (I_N - \theta W_N)^{-1} \left(I_N - \theta W_N'\right)^{-1},$$

or, in a simpler notation, with $B_N = I_N - \theta W_N$:

$$\Omega_{t,N} = \sigma_u^2 (B_N' B_N)^{-1}.$$

As before, in this homogeneous case, the cross-sectional covariance does not vary over time, so that the full $NT \times NT$ covariance matrix follows as:

$$\Sigma_{NT} = \sigma_u^2 \left[I_T \otimes (B_N' B_N)^{-1}\right]. \tag{18.9}$$

Note that for a row-standardized weights matrix, B_N will not be symmetric. Also, even though W_N may be sparse, the inverse term $(B_N' B_N)^{-1}$ will not be sparse and suggests a much wider range of spatial *covariance* than specified by the non-zero elements of the weights matrix. In other words, the spatial covariance structure induced by the SAR model is *global*.

A *spatial moving average* (SMA) specification for the $N \times 1$ error vector ϵ_t in period $t = 1, \ldots, T$, can be expressed as:

$$\epsilon_t = \gamma W_N u_t + u_t,$$

where γ is the moving average parameter, and the other notation is as before. In contrast to the SAR model, the variance covariance matrix for an error SMA process does not involve a matrix inverse:

$$\Omega_{t,N} = \mathrm{E}[\epsilon_t \epsilon_t'] = \sigma_u^2 \left[I_N + \gamma(W_N + W_N') + \gamma^2 W_N W_N'\right], \tag{18.10}$$

and, in the homogenous case, the overall error covariance matrix follows directly as:

$$\Sigma_{NT} = \sigma_u^2 \left(I_T \otimes \left[I_N + \gamma(W_N + W_N') + \gamma^2 W_N W_N'\right]\right).$$

Aso, in contrast to the SAR model, the spatial covariance induced by the SMA model is *local*.[14]

2.3.3 Spatial Error Components. A spatial error components specification (SEC) was suggested by Kelejian and Robinson as an alternative to the SAR and SMA models (Kelejian and Robinson, 1995, Anselin and Moreno, 2003). In the SEC model, the error term is decomposed into a local and a spillover effect.

In a panel data setting, the $N \times 1$ error vector ϵ_t for each time period $t = 1, \ldots, T$, is expressed as:

$$\epsilon_t = W_N \psi_t + \xi_t, \tag{18.11}$$

where W_N is the weights matrix, ξ_t is a $N \times 1$ vector of local error components, and ψ_t is a $N \times 1$ vector of spillover error components. The two component vectors are assumed to consist of *i.i.d* terms, with respective variances σ^2_ψ and σ^2_ξ, and are uncorrelated, $\mathrm{E}[\psi_{it}\xi_{jt}] = 0,\ \forall\, i, j, t$.

The resulting $N \times N$ cross-sectional error covariance matrix is then, for $t = 1, \ldots, T$:

$$\Omega_{t,N} = \mathrm{E}[\epsilon_t \epsilon_t'] = \sigma^2_\psi W_N W_N' + \sigma^2_\xi I_N. \tag{18.12}$$

In the homogeneous model, this is again unchanging across time periods, and the overall $NT \times NT$ error covariance matrix can be expressed as:

$$\Sigma_{NT} = \sigma^2_\xi I_{NT} + \sigma^2_\psi (I_T \otimes W_N W_N').$$

Comparing equations 18.10 and 18.12, it can be readily seen that the range of covariance induced by the SEC model is a subset that of the SMA model, and hence also a case of *local* spatial externalities.

2.3.4 Common Factor Models. In the standard two-way error component regression model, the cross-sectional units share an unobserved component due to the time period effect (e.g., Baltagi, 2001, p. 31). In our notation:

$$\epsilon_{it} = \mu_i + \lambda_t + u_{it},$$

with μ_i as the cross-sectional component, with variance σ^2_μ, λ_t as the time component, with variance σ^2_λ, and u_{it} as an idiosyncratic error term, assumed to be *i.i.d* with variance σ^2_u. Furthermore, the three random components are assumed to be zero mean and to be uncorrelated with each other. The random components μ_i are assumed to be uncorrelated across cross-sectional units, and the components λ_t are assumed to be uncorrelated across time periods. This model is standard, except that for our purposes, the data are stacked as cross-sections for different time periods. Consequently, the $N \times 1$ cross-sectional error vector ϵ_t for time period $t = 1, \ldots, T$, becomes:

$$\epsilon_t = \mu + \lambda_t \iota_N + u_t, \tag{18.13}$$

where μ is a $N \times 1$ vector of cross-sectional error components, λ_t is a scalar time component, ι_N is a $N \times 1$ vector of ones, and u_t is a $N \times 1$ vector of idiosyncratic errors.

The structure in equation 18.13 results in a particular form of cross-sectional (spatial) correlation, due to the common time component:

$$\mathrm{E}[\epsilon_t \epsilon_t'] = \sigma^2_\mu I_N + \sigma^2_\lambda \iota_N \iota_N' + \sigma^2_u I_N,$$

where the subscript N indicates the dimension of the identity matrices. Note that the second term in this expression indicates equicorrelation in the cross-sectional dimension, i.e., the correlation between two cross-sectional units i, j equals σ^2_λ, no matter how far these units are apart. While perfectly valid as a model for general (global) cross-sectional correlation, this violates the distance decay effect that underlies spatial interaction theory.

The complete $NT \times 1$ error vector can be written as (see also Anselin, 1988a, p. 153):

$$\epsilon = (\iota_T \otimes I_N)\mu + (I_T \otimes \iota_N)\lambda + u,$$

where the subscripts indicate the dimensions, λ 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:

$$\Sigma_{NT} = \sigma^2_\mu(\iota_T\iota_T' \otimes I_N) + \sigma^2_\lambda(I_T \otimes \iota_N\iota_N') + \sigma^2_u I_{NT}.$$

Note the how the order of matrices in the Kronecker products differs from the standard textbook notation, due to the stacking by cross-section.

A recent extension of the error component model can be found in the literature on heterogeneous panels. Here, the time component is generalized and expressed in the form of an unobserved common effect, common shock or *factor* f_t to which all cross-sectional units are exposed (for a recently developed general framework that includes factors as a special case, see Andrews, 2005). 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 is specified as:

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

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:

$$\mathrm{E}[\epsilon_{it}\epsilon_{jt}] = \delta_i\delta_j\sigma^2_f.$$

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. In general, the error term is:

$$\epsilon_{it} = \boldsymbol{\delta_i}'\mathbf{f_t} + u_{it},$$

where δ and $\mathbf{f}$ are vectors of dimension m, equal to the number of unobserved common factors. After standardizing the idiosyncratic error (e.g., as in Pesaran, 2004), we find the correlation between two cross-sectional terms as:

$$\rho_{ij} = \frac{\boldsymbol{\delta_i}'\boldsymbol{\delta_j}}{(1 + \boldsymbol{\delta_i}'\boldsymbol{\delta_i})^{1/2}(1 + \boldsymbol{\delta_j}'\boldsymbol{\delta_j})^{1/2}}.$$

In contrast to the covariance structures considered in sections 18.2.3.1-18.2.3.3, the multiple factor model allows for dependencies that are not direct functions of distance or of a spatial weights specification. Provided the data are sufficiently rich (and, typically, in large T contexts), the correlation is determined by the model residuals (for further details, see Driscoll and Kraay, 1998, Coakley et al., 2002, Coakley et al., 2005, Hsiao and Pesaran, 2004, Pesaran, 2005 and Kapetanios and Pesaran, 2005).

3. A Taxonomy of Spatial Panel Model Specifications

So far, we have considered the introduction of spatial effects for panel data in the form of spatial lag or spatial error models under extreme homogeneity. The point of departure was the pooled specification, equation 18.1, and lag and error models are obtained as outlined in sections 18.2.2 and 18.2.3. We now extend this taxonomy by introducing heterogeneity, both over time and across space, as well as by considering joint space-time dependence.

It should be noted that a large number of combinations of space-time heterogeneity and dependence are possible, although many of those suffer from identification problems and/or are not estimable in practice. In our classification here, we purposely limit the discussion to models that have seen some empirical applications (other, more extensive typologies can be found in Anselin, 1988a, Chapter 4, Anselin, 2001b, Elhorst, 2001, and Elhorst, 2003).

3.1 Temporal Heterogeneity

3.1.1 General Case. Temporal heterogeneity is introduced in the familiar way in fixed effects models, by allowing time-specific intercepts and/or slopes, and in random effects models, by incorporating a random time component or factor (see section 18.2.3.4). The addition of a spatially lagged dependent variable or spatially correlated error term in these models is straightforward. For example, consider a pooled model with time-specific intercept and slope coefficient to which a spatially autoregressive error term is added. The cross-section in each period $t = 1, \ldots, T$, is:

$$y_t = \alpha_t + X_t\beta_t + \epsilon_t, \tag{18.14}$$

with

$$\epsilon_t = \theta_t W_N \epsilon_t + u_t$$

where θ_t is a period-specific spatial autoregressive parameter, α_t is the period-specific intercept and β_t a $(K-1) \times 1$ vector of period-specific slopes. Since T is fixed (and the asymptotics are based on $N \to \infty$), this model is a straightforward replication of T cross-sectional models. A spatial lag specification is obtained in a similar way.

3.1.2 Spatial Seemingly Unrelated Regressions. A generalization of the fixed effects model that has received some attention in the empirical literature (e.g., Rey and Montouri, 1999) allows the cross-sectional error terms ϵ_t to be correlated over time periods. This imposes very little structure on the form of the temporal dependence and is the spatial counterpart of the classic SURE model. It is referred to as the *spatial* SUR model (see Anselin, 1988a, Chapter 10, and Anselin, 1988b). In matrix form, the equation for the cross-sectional regression in each time period $t = 1, \ldots, T$, is as in equation 18.14, but now with the constant term included in the vector β_t:

$$y_t = X_t\beta_t + \epsilon_t, \tag{18.15}$$

with the cross-equation (temporal) correlation in general form, as:

$$\mathrm{E}[\epsilon_t\epsilon_s'] = \sigma_{ts}I_N, s \neq t,$$

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 is:

$$y = X\beta + \epsilon, \tag{18.16}$$

with

$$\mathrm{E}[\epsilon\epsilon'] = \Sigma_T \otimes I_N \tag{18.17}$$

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

Spatial correlation can be introduced as a spatial lag specification or a spatial error specification. Consider the spatial lag model first (see Anselin, 1988a for details). In each cross-section (with $t = 1, \ldots, T$), the standard spatial lag specification holds, but now with a time-specific spatial autoregressive coefficient ρ_t:

$$y_t = \rho_t W_N y_t + X_t\beta_t + \epsilon_t.$$

To consider the full system, let β be a $TK \times 1$ vector of the stacked time-specific β_t, for $t = 1, \ldots, T$.[15] The corresponding $NT \times KT$ matrix X of observations on the explanatory variables then takes the form:

$$X = \begin{pmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & X_T \end{pmatrix}. \tag{18.18}$$

Also, let the spatial autoregressive coefficients be grouped in a $T \times T$ diagonal matrix R_T, as:

$$R_T = \begin{pmatrix} \rho_1 & 0 & \ldots & 0 \\ 0 & \rho_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \rho_T \end{pmatrix}.$$

The full system can then be expressed concisely as:

$$y = (R_T \otimes W_N)y + X\beta + \epsilon, \tag{18.19}$$

with the error covariance matrix as in equation 18.17.

In empirical practice, besides the standard hypothesis tests on diagonality of the error covariance matrix and stability of the regression coefficients over time, interest in the *spatial lag* SUR model will center on testing the hypothesis of homogeneity of the spatial autoregressive coefficients, or, $\mathrm{H}_0 : \rho_1 = \rho_2 = \ldots = \rho_T = \rho$. If this null hypothesis can be maintained, a simplified model can be implemented:

$$y = \rho(I_T \otimes W_N)y + X\beta + \epsilon.$$

Spatial error autocorrelation can be introduced in the spatial SUR model in the form of a SAR or SMA process for the error terms (see Anselin, 1988a, Chapter 10). For example, consider the following SAR error process for the cross-section in each time period $t = 1, \ldots, T$:

$$\epsilon_t = \theta_t W_N \epsilon_t + u_t. \tag{18.20}$$

The cross-equation covariance is introduced through the remainder error term u_t, for which it is assumed that $\mathrm{E}[u_t] = 0$, $\mathrm{E}[u_t u_t'] = \sigma_t^2 I_N$, and $\mathrm{E}[u_t u_s'] = \sigma_{ts} I_N$, for $t \neq s$. As a result, the covariance matrix for the stacked $NT \times 1$ error vector u becomes the counterpart of equation 18.17:

$$\mathrm{E}[uu'] = \Sigma_T \otimes I_N,$$

with, as before, Σ_T as a $T \times T$ matrix with elements σ_{ts}.

The SAR error process in equation 18.20 can also be written as:

$$\epsilon_t = (I_N - \theta_t W_N)^{-1} u_t,$$

or, using the simplifying notation $B_{t,N} = (I_N - \theta_t W_N)$, as:

$$\epsilon_t = B_{t,N}^{-1} u_t.$$

The overall cross-equation covariance between error vectors ϵ_t and ϵ_s then becomes:

$$\mathrm{E}[\epsilon_t \epsilon_s'] = B_{t,N}^{-1} \mathrm{E}[u_t u_s'] B_{s,N}^{-1\prime} = \sigma_{ts} B_{t,N}^{-1} B_{s,N}^{-1\prime},$$

which illustrates how the simple SUR structure induces space-time covariance as well (the $B_{t,N}^{-1}$ matrices are not diagonal).

In stacked form, the error process for the $NT \times 1$ error vector ϵ can be written as:

$$\epsilon = B_{NT}^{-1} u,$$

with B_{NT} as the matrix:

$$B_{NT} = [I_{NT} - (\Theta_T \otimes W_N)], \tag{18.21}$$

and Θ_T as a $T \times T$ diagonal matrix containing the spatial autoregressive coefficients $\theta_t, t = 1, \ldots, T$. The overall error covariance matrix for the stacked equations then becomes:

$$\mathrm{E}[\epsilon\epsilon'] = B_{NT}^{-1}(\Sigma_T \otimes I_N)B_{NT}^{-1'}. \tag{18.22}$$

As in the spatial lag SUR model, specific interest in the spatial error SUR model centers on the homogeneity of the spatial autoregressive parameters, $\mathrm{H}_0 : \theta_1 = \theta_2 = \ldots = \theta_T = \theta$. If the homogeneity holds, the expression for B_{NT} (equation 18.21) simplifies to:

$$B_{NT} = [I_T \otimes (I_N - \theta W_N)].$$

3.2 Spatial Heterogeneity

We limit our attention in the treatment of spatial heterogeneity to models with unobserved heterogeneity, specified in the usual manner as either fixed effects or random effects. Both have been extended with spatial lag and spatial error specifications.

3.2.1 Fixed Effects Models. The classic fixed effects model (e.g., Baltagi, 2001, pp. 12–15, and Arellano, 2003, pp. 11–18) includes an individual specific "dummy variable" to capture unobserved heterogeneity. For each observation i, t this yields, keeping the same notation as before:

$$y_{i,t} = \alpha_i + x_{it}\beta + \epsilon_{it}$$

for $i = 1, \ldots, N$, $t = 1, \ldots, T$, and with an additional constraint of the form $\sum_i \alpha_i = 0$, such that the individual effects α_i are separately identifiable from the constant term in β.

As is well known, consistent estimation of the individual fixed effects is not possible when $N \to \infty$, due to the incidental parameter problem. Since spatial models rely on the asymptotics in the cross-sectional dimension to obtain consistency and asymptotic normality of estimators, this would preclude the fixed effects model from being extended with a spatial lag or spatial error term (Anselin, 2001b).

Nevertheless, it has been argued that when the interest is primarily in obtaining consistent estimates for the β coefficients, the use of *demeaned* spatial regression models may be appropriate, for example, using the standard maximum likelihood estimation approach (Elhorst, 2003, p. 250–251).

There are several slight complications associated with this approach. One is that the demeaning operator takes on a different form from the usual expression in the literature, since the observations are stacked as cross-sections for different time periods. Also, the demeaned models no longer contain a constant term, which may be incompatible with assumptions made by standard spatial econometric software. More importantly, the variance covariance matrix of the demeaned error terms is no longer $\sigma_\epsilon^2 I$, but becomes $\sigma_\epsilon^2 Q$, where Q is the demeaning operator (this aspect is ignored in the likelihood functions presented in Elhorst, 2003, p. 250).

To illustrate these points, consider a fixed effects spatial lag model in stacked form, using the same setup as in equation 18.5, with the addition of the fixed effects:

$$y = \rho(I_T \otimes W_N)y + (\iota_T \otimes \alpha) + X\beta + \epsilon, \tag{18.23}$$

where α is a $N \times 1$ vector of individual fixed effects, with the constraint that $\alpha'\iota_N = 0$, and, as before, $\mathrm{E}[\epsilon\epsilon'] = \sigma_\epsilon^2 I_{NT}$. Note the difference with the classic formulation in the Kronecker product for the fixed effects, due to the stacking of cross-sections, rather than individual time series.

The demeaned form of equation 18.23 is obtained by substracting the average for each cross-sectional unit computed over the time dimension, which wipes out the individual fixed effects (as well as the constant term). Formally, this can be expressed as:

$$Q_{NT}y = \rho(I_T \otimes W_N)Q_{NT}y + Q_{NT}X\beta + Q_{NT}\epsilon, \tag{18.24}$$

where Q_{NT} is the demeaning operator (and $Q_{NT}X$ and β no longer contain a constant term). The demeaning operator is a $NT \times NT$ matrix that takes the form:

$$Q_{NT} = I_{NT} - (\iota_T\iota_T'/T \otimes I_N)$$

with, as before, ι as a vector of ones and the subscripts denoting the dimension of vectors and matrices. Again, note the difference with the standard textbook notation, due to stacking by cross-section. The matrix Q_{NT} is idempotent, and, as a result, the expression for the variance of the error in 18.24 becomes:

$$\mathrm{E}[\epsilon\epsilon'] = \sigma_\epsilon^2 Q_{NT}.$$

Since Q_{NT} is idempotent, this variance is singular, which limits the practicality of this approach. The use of a conditional likelihood perspective, standard in the panel data literature (e.g., Arellano, 2003, pp. 24–25) does not apply in a straightforward manner to a spatial lag specification, since the average of the dependent variable $\bar{y}_i = (1/T)\sum_t y_{it}$ is spatially correlated among the i. A more careful elaboration of the relative merits of alternative approaches remains largely a topic of ongoing research.

3.2.2 Random Effects Models. In the random effects approach to modeling unobserved heterogeneity, interest has centered on incorporating spatial error correlation into the regression error term, in addition to the standard cross-sectional random component. Note that the latter induces serial correlation over time (of the equi-correlated type). The addition of a spatial lag term in these models is straightforward, by specifying the proper error variance covariance structure. We therefore focus attention on the one-way error component specification (e.g, Baltagi, 2001, pp. 15–20, Arellano, 2003, Ch. 3).[16]

In contrast to the fixed effects case, asymptotics along the cross-sectional dimension (with $N \to \infty$) present no problem for random effects models. The standard specification of the error term in this model, is, for each i, t:

$$\epsilon_{it} = \mu_i + \nu_{it},$$

where $\mu_i \backsim \text{IID}(0, \sigma^2_\mu)$ is the cross-sectional random component, and $\nu_{it} \backsim \text{IID}(0, \sigma^2_\nu)$ is an idiosyncratic error term, with μ_i and ν_{it} independent from each other. In each cross-section, for $t = 1, \ldots, T$, the $N \times 1$ error vector ϵ_t becomes:

$$\epsilon_t = \mu + \nu_t, \tag{18.25}$$

where μ is a $N \times 1$ vector of cross-sectional random components. The common approach to introduce spatial error autocorrelation in equation 18.25 is to specify a SAR process for the error component ν_t, for $t = 1, \ldots, T$ (Anselin, 1988a, p. 153, and, more recently, Baltagi et al., 2003b):

$$\nu_t = \theta W_N \nu_t + u_t, \tag{18.26}$$

with θ as the spatial autoregressive parameter (constant over time), W_N as the spatial weights matrix, and u_t as an $i.i.d$ idiosyncratic error term with variance σ^2_u.

Using the notation $B_N = I_N - \theta W_N$, we obtain the familiar result:

$$\nu_t = (I_N - \theta W_N)^{-1} u_t = B_N^{-1} u_t.$$

In stacked form, the $NT \times 1$ error term then becomes:

$$\epsilon = (\iota_T \otimes I_N)\mu + (I_T \otimes B_N^{-1})u, \tag{18.27}$$

where $u \backsim \text{IID}(0, \sigma^2_u I_{NT})$ is a $NT \times 1$ vector of idiosyncratic errors, and the variance covariance matrix for ϵ is:

$$\Sigma_{NT} = \text{E}[\epsilon\epsilon'] = \sigma^2_\mu(\iota_T \iota'_T \otimes I_N) + \sigma^2_u[I_T \otimes (B'_N B_N)^{-1}]. \tag{18.28}$$

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).

An alternative specification for spatial correlation in this model applies the SAR process first and the error components specification to its remainder error (Kapoor et al., 2003). Consider a SAR process for the $NT \times 1$ error vector ϵ:

$$\epsilon = \theta(I_T \otimes W_N)\epsilon + \nu,$$

or, using similar notation as the spatially correlated component in equation 18.27:

$$\epsilon = (I_T \otimes B_N^{-1})\nu.$$

Now, the innovation vector ν is specified as a one way error component model (Kapoor et al., 2003):

$$\nu = (\iota_T \otimes I_N)\mu + u,$$

with μ as the $N \times 1$ vector of cross-sectional random components, and $u \backsim$ IID$(0, \sigma_u^2 I_{NT})$. As a result, the error variance covariance matrix takes on a form that is different from 18.28:

$$\Sigma_{NT} = \mathrm{E}[\epsilon\epsilon'] = (I_T \otimes B_N^{-1})[\sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + \sigma_u^2 I_{NT}](I_T \otimes B_N^{-1'}). \quad (18.29)$$

Again, this model combines both time-wise as well as cross-sectional correlation.

Statistical inference for error components models with spatial SAR processes can be carried out as a special case of models with non-spherical error covariance. This is addressed in sections 18.4 and 18.5.

3.3 Spatio-Temporal Models

The incorporation of dependence in both time and space dimensions in an econometric specification adds an additional order of difficulty to the identification of the $NT \times (NT-1)/2$ elements of the variance covariance matrix. An important concept in this regard is the notion of *separability*. Separability requires that a $NT \times NT$ space-time covariance matrix Σ_{NT} can be decomposed into a component due to space and a component due to time (see, e.g., Mardia and Goodall, 1993), or:

$$\Sigma_{NT} = \Sigma_T \otimes \Sigma_N,$$

where Σ_T is a $T \times T$ variance covariance matrix for the time-wise dependence and Σ_N is a $N \times N$ variance covariance matrix for the spatial dependence.[17] This ensures that the space-time dependence declines in a multiplicative fashion over the two dimensions. It also addresses a central difficulty in space-time modeling, i.e., the lack of a common "distance" metric that works both in the cross-sectional and the time dimension. The approach taken in spatial panel econometrics is to define "neighbors" in space by means of a spatial weights matrix and "neighbors" in time by means of the customary time lags. However, the speed of the dynamic space-time process may not be compatible with these choices, leading to further misspecification.

3.3.1 Model Taxonomy. Ignoring for now any space-time dependence in the error terms, we can distinguish four basic forms to introduce correlation in both space and time in panel data models (following Anselin, 2001b, p. 317–318). As before, we focus on models where $N \gg T$ and do not consider specifications where the time dimension is an important aspect of the model.[18] Also, we limit the scope of the discussion to the pooled dynamic model, since the incorporation of spatial effects in heterogeneous dynamic panel models has seen little attention to date. To facilitate exposition, we express these models for a $N \times 1$ cross-section at time $t = 1, \ldots, T$.

Pure space recursive models, in which the dependence pertains only to neighboring locations in a previous period:

$$y_t = \gamma W_N y_{t-1} + X_t \beta + \epsilon_t, \tag{18.30}$$

with γ as the space-time autoregressive parameter, and $W_N y_{t-1}$ as a $N \times 1$ vector of observations on the spatially lagged dependent variable at $t-1$. Note that this can be readily extended with time and spatial lags of the explanatory variables, X_{t-1} or $W_N X_t$. However, since $W_N y_{t-1}$ already includes $W_N X_{t-1}$, adding a term of this form would create identification problems. This is sometimes overlooked in other taxonomies of dynamic space-time models (e.g., in the work of Elhorst, 2001, p. 121, where space-time lags for both the dependent and the exploratory variables are included in the specification).

Consider the space-time *multiplier* more closely. Start by substituting the equation for y_{t-1} in 18.30, which yields:

$$y_t = \gamma W_N [\gamma W_N y_{t-2} + X_{t-1}\beta + \epsilon_{t-1}] + X_t \beta + \epsilon_t,$$

or,

$$y_t = \gamma^2 W_N^2 y_{t-2} + X_t \beta + \gamma W_N X_{t-1} \beta + \epsilon_t + \gamma W_N \epsilon_{t-1}.$$

Successive substitution reveals a space-time multiplier that follows from a series of consecutively higher orders of both spatial and time lags applied to the X (and error terms). Also, since the spatial dependence takes one period to manifest itself, this specification becomes quite suitable to study spatial diffusion phenomena (see the early discussion in Upton and Fingleton, 1985, and Dubin, 1995).

Time-space recursive models, in which the dependence relates to both the location itself as well as its neighbors in the previous period:

$$y_t = \phi y_{t-1} + \gamma W_N y_{t-1} + X_t \beta + \epsilon_t, \tag{18.31}$$

with ϕ as the serial (time) autoregressive parameter, operating on the cross-section of dependent variables at $t-1$. Spatially lagged contemporaneous explanatory variables ($W_N X_t$) may be included as well, but time lagged explanatory variables will result in identification problems. This model has particular appeal in space-time forecasting (e.g., Giacomini and Granger, 2004).

Again, the nature of the space-time multiplier can be assessed by substituting the explicit form for the spatially and time lagged terms:

$$\begin{aligned} y_t &= \phi[\phi y_{t-2} + \gamma W_N y_{t-2} + X_{t-1}\beta + \epsilon_{t-1}] \\ &\quad +\gamma W_N[\phi y_{t-2} + \gamma W_N y_{t-2} + X_{t-1}\beta + \epsilon_{t-1}] \\ &\quad +X_t\beta + \epsilon_t, \end{aligned}$$

or,

$$\begin{aligned} y_t &= (\phi^2 + 2\phi\gamma W_N + \gamma^2 W_N^2)y_{t-2} \\ &\quad +X_t\beta + (\phi + \gamma W_N)X_{t-1}\beta \\ &\quad +\epsilon_t + (\phi + \gamma W_N)\epsilon_{t-1}, \end{aligned}$$

revealing a much more complex form for the effect of space-time lagged explanatory variables (and errors), including the location itself as well as its neighbors.

Time-space simultaneous models, which include a time lag for the location itself together with a contemporaneous spatial lag:

$$y_t = \phi y_{t-1} + \rho W_N y_t + X_t\beta + \epsilon_t,$$

with ρ as the (contemporaneous) spatial autoregressive parameter.

The mulitplier in this model is complex, due to the combined effect of the cross-sectional *spatial* multiplier (in each period) and the space-time multiplier that follows from the time lag in the dependent variable. First, consider the pure cross-sectional multiplier:

$$y_t = (I_N - \rho W_N)^{-1}[\phi y_{t-1} + X_t\beta + \epsilon_t].$$

Next, substitute the corresponding expression for y_{t-1}:

$$\begin{aligned} y_t &= (I_N - \rho W_N)^{-1}[\phi[(I_N - \rho W_N)^{-1}(\phi y_{t-2} \\ &\quad +X_{t-1}\beta + \epsilon_{t-1})] + X_t\beta + \epsilon_t], \end{aligned}$$

which yields:

$$\begin{aligned} y_t &= \phi^2(I_N - \rho W_N)^{-2}y_{t-2} \\ &\quad +(I_N - \rho W_N)^{-1}X_t\beta + \phi(I_N - \rho W_N)^{-2}X_{t-1}\beta \\ &\quad +(I_N - \rho W_N)^{-1}\epsilon_t + \phi(I_N - \rho W_N)^{-2}\epsilon_{t-1}. \end{aligned}$$

From this it follows that the inclusion of any spatially lagged X in the original specification will lead to identification problems.

Time-space dynamic models, where all three forms of lags for the dependent variable are included:

$$y_t = \phi y_{t-1} + \rho W_N y_t + \gamma W_N y_{t-1} + X_t\beta + \epsilon_t.$$

While this model is sometimes suggested as a *general* space-time specification, it results in complex nonlinear constraints on the parameters, and, in practice, often suffers from identification problems. For example, focusing only on the time lagged terms and substituting their expression for $t-1$ (and rearranging terms) yields:

$$\begin{aligned} y_t \;=\;& \phi[\phi y_{t-2} + \rho W_N y_{t-1} + \gamma W_N y_{t-2} + X_{t-1}\beta + \epsilon_{t-1}] \\ & + \gamma W_N[\phi y_{t-2} + \rho W_N y_{t-1} + \gamma W_N y_{t-2} + X_{t-1}\beta + \epsilon_{t-1}] \\ & + \rho W_N y_t + X_t \beta + \epsilon_t, \end{aligned}$$

or, grouping by time period:

$$\begin{aligned} y_t \;=\;& \rho W_N y_t + X_t\beta + \epsilon_t \\ & + \phi\rho W_N y_{t-1} + \gamma\rho W_N^2 y_{t-1} + \phi X_{t-1}\beta + \gamma W_N X_{t-1}\beta \\ & + \phi\epsilon_{t-1} + \gamma W_N \epsilon_{t-1} \\ & + \phi^2 y_{t-2} + \gamma\phi W_N y_{t-2} + \gamma^2 W_N^2 y_{t-2}. \end{aligned}$$

The same types of space-time dependence processes can also be specified for the error terms in panel data models (e.g., Fazekas et al., 1994). However, combinations of both spatially lagged dependent variables and spatially lagged error terms may lead to identification problems unless the parameters of the explanatory variables are non-zero. An alternative form of error space-time dependence takes the error components approach, to which we turn briefly.

3.3.2 Error Components with Space-Time Dependence. The starting point for including explicit serial dependence (in addition to the equicorrelated form) in random effects models is the spatially autocorrelated form considered in equations 18.25–18.26. However, instead of the indiosyncratic error u_t in 18.26, a serially correlated term ζ_t is introduced (Baltagi et al., 2003a):

$$\nu_t = \theta W_N \nu_t + \zeta_t \tag{18.32}$$

with

$$\zeta_t = \phi\zeta_{t-1} + u_t, \tag{18.33}$$

where, as before, u_t is used to denote the idiosyncratic error, and $t = 1, \ldots, T$. The counterpart of the $N \times 1$ cross-sectional error vector ϵ_t in equation 18.27 becomes:

$$\epsilon_t = (I_N - \theta W_N)^{-1}\zeta_t = B_N^{-1}\zeta_t,$$

with ζ replacing the original error u. In stacked form, this becomes:

$$\epsilon = (\iota_T \otimes I_N)\mu + (I_T \otimes B_N^{-1})\zeta,$$

with μ of dimension $N \times 1$ and both ϵ and ζ of dimension $NT \times 1$. The serial correlation in ζ will yield serial covariances of the familiar AR(1) form, with:

$$\mathrm{E}[\zeta_{i,t}\zeta_{i,t-k}] = \sigma_u^2 \left(\frac{\phi^k}{1-\phi^2} \right),$$

for $k = 0, \ldots, T-1$, and $i = 1, \ldots, N$, where σ_u^2 is the variance of the error term u. Grouping these serial covariances into a $T \times T$ variance covariance matrix Ω_T yields the overall variance covariance matrix for ϵ as (Baltagi et al., 2003a):

$$\Sigma_{NT} = \mathrm{E}[\epsilon\epsilon'] = \sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + [\Omega_T \otimes (B_N' B_N)^{-1}].$$

4. Estimation of Spatial Panel Models

The estimation of panel data models that include spatially lagged dependent variables and/or spatially correlated error terms follows as a direct extension of the theory developed for the single cross-section. In the first case, the endogeneity of the spatial lag must be dealt with, in the second, the non-spherical nature of the error variance covariance matrix must be accounted for. Two main approaches have been suggested in the literature, one based on the maximum likelihood principle, the other on method of moments techniques. We consider each in turn.

We limit our attention to models with a parameterized form for the spatial dependence, specified as a spatial autoregressive process.[19] Note that some recent results in the panel econometrics literature have also addressed estimation in models with general, unspecified cross-sectional correlation (see, e.g., Driscoll and Kraay, 1998, Coakley et al., 2002, Coakley et al., 2005, Pesaran, 2005, Kapetanios and Pesaran, 2005).

4.1 Maximum Likelihood Estimation

The theoretical framework for maximum likelihood estimation of spatial models in the single cross-section setup is by now well developed (see, among others, Ord, 1975, Mardia and Marshall, 1984, Anselin, 1988a, Cressie, 1993, Anselin and Bera, 1998). While the regularity conditions are non-standard, and require a consideration of triangular arrays (Kelejian and Prucha, 1999), the results for error terms with a Gaussian distribution are fairly well established.

In practice, estimation consists of applying a non-linear optimization to the log-likelihood function, which (in most circumstances) yields a consistent estimator from the numerical solution to the first order conditions. Asymptotic inference is based on asymptotic normality, with the asymptotic variance matrix derived from the information matrix. This requires the second order partial derivatives of the log-likelihood, for which analytical solutions exist in many of the models considered (for technical details, see the review in Anselin and Bera, 1998).

A main obstacle in the practical implementation of ML estimation in a single cross-section is the need to compute a Jacobian determinant for an N-dimensional matrix (the dimension of the cross-section). In panel data models, this Jacobian is of dimension $N \times T$, but it can often be simplified to a product of T N-dimensional determinants. The classic solution to this problem is to decompose the Jacobian in terms of the eigenvalues of the spatial weights matrix. For example, in the spatial lag model, the Jacobian would be $|I_N - \rho W_N| = \prod_i (1 - \rho\omega_i)$, with ω_i as the eigenvalues of W_N (Ord, 1975).[20] For large cross-sections, the computation of the eigenvalues becomes numerically unstable, precluding this method from being applicable. Alternative solutions avoid the computation of the Jacobian determinant, but instead approximate it by a polynomial function or by means of simulation methods (Barry and Pace, 1999). Other methods are based on Cholesky or LU decomposition methods that exploit the sparsity of the spatial weights (Pace and Barry, 1997), or use a characteristic polynomial approach (Smirnov and Anselin, 2001).

We now briefly review a number of useful log-likelihood expressions that result when incorporating spatial lag or spatial error terms in panel data settings. Numerical procedures to carry out estimation and inference can be implemented along the same lines as for the single cross-section, and will not be further elaborated.

4.1.1 Spatial Lag Models. As a point of departure, consider the pooled spatial lag model given in equation 18.5. Assuming a Gaussian distribution for the error term, with $\epsilon \backsim N(0, \sigma_\epsilon^2 I_{NT})$, the log-likelihood (ignoring the constants) follows as:

$$L = \ln |I_T \otimes (I_N - \rho W_N)| - \frac{NT}{2} \ln \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \epsilon'\epsilon,$$

with $\epsilon = y - \rho(I_T \otimes W_N)y - X\beta$, and $|I_T \otimes (I_N - \rho W_N)|$ as the Jacobian determinant of the spatial transformation. Given the block diagonal structure of the Jacobian, the log-likelihood further simplifies to:

$$L = T \ln |I_N - \rho W_N| - \frac{NT}{2} \ln \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \epsilon'\epsilon, \tag{18.34}$$

which boils down to a repetition of the standard cross-sectional model in T cross-sections.

Generalizing this model slightly, we now assume $\epsilon \backsim N(0, \Sigma)$ to allow for more complex error covariance structures (including spatial correlation). The log-likelihood remains essentially the same, except for the new error covariance term:

$$L = T \ln |I_N - \rho W_N| - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \epsilon' \Sigma^{-1} \epsilon. \tag{18.35}$$

Two special cases result by introducing some structure into the variance covariance matrix Σ. First, consider the classic one-way error components model from equation 18.25, which, in stacked form, becomes (again, using cross-sections for T time periods and assuming a Gaussian distribution for ϵ):

$$\epsilon = (\iota_T \otimes I_N)\mu + u.$$

The error covariance matrix follows as:

$$\Sigma_{NT} = \mathrm{E}[\epsilon\epsilon'] = \sigma_\mu^2(\iota_T \iota_T' \otimes I_N) + \sigma_u^2 I_{NT}. \tag{18.36}$$

Using standard results, the inverse and determinant of this $NT \times NT$ matrix can be expressed in terms of matrix determinants and inverses of orders N and T only. Inserting 18.36 into 18.35 yields the log-likelihood for the spatial lag model with error components as:

$$\begin{aligned} L = \quad & T \ln |I_N - \rho W_N| - \tfrac{1}{2} \ln |\sigma_\mu^2(\iota_T \iota_T' \otimes I_N) + \sigma_u^2 I_{NT}| \\ & -\tfrac{1}{2}\epsilon' \left[\sigma_\mu^2(\iota_T \iota_T' \otimes I_N) + \sigma_u^2 I_{NT}\right]^{-1} \epsilon. \end{aligned}$$

A second specification of interest is the SUR model that includes a spatial lag term, equation 18.19. Its log-likelihood can be obtained in a similar fashion. Using the same notation and stacking of observation matrices and parameters as in equations 18.18–18.19, the log Jacobian follows as $\ln |I_{NT} - (R_T \otimes W_N)|$. The block diagonal structure of the matrix can be exploited to simplify this expression to $\sum_t \ln |I_N - \rho_t W_N|$ (with the sum over $t = 1, \ldots, T$). Using 18.17 for the error variance covariance matrix in the SUR model, the log-likelihood follows as:

$$L = \sum_t \ln |I_N - \rho_t W_N| - \frac{N}{2} \ln |\Sigma_T| - \frac{1}{2}\epsilon' \left(\Sigma_T^{-1} \otimes I_N\right) \epsilon,$$

with $\epsilon = [I_{NT} - (R_T \otimes W_N)]y - X\beta$ (for further details, see Anselin, 1988a, pp. 145–146).

4.1.2 Spatial Error Models. The log-likelihood functions for the various spatial error models considered in this chapter follow directly as special cases of the standard result for maximum likelihood estimation with a non-spherical error covariance (Magnus, 1978). With $\epsilon \backsim N(0, \Sigma)$ as the error vector, the familiar expression for the log-likelihood is (ignoring the constant terms):

$$L = -\frac{1}{2} \ln |\Sigma| - \frac{1}{2} \epsilon' \Sigma^{-1} \epsilon.$$

In the pooled model with SAR error terms, equation 18.9, the relevant determinant and inverse matrix are:

$$|I_T \otimes (B_N' B_N)^{-1}| = |B_N|^{-2T}$$

with B_N as in 18.9, and:

$$\Sigma_{NT}^{-1} = \frac{1}{\sigma_u^2}[I_T \otimes (B_N' B_N)].$$

The corresponding log-likelihood function is then:

$$\begin{aligned} L &= -\frac{NT}{2} \ln \sigma_u^2 + T \ln |B_N| \\ &\quad -\frac{1}{2\sigma_u^2} \epsilon' [I_T \otimes (B_N' B_N)] \epsilon, \end{aligned}$$

with $\epsilon = y - X\beta$. The estimates for the regression coefficient β are the result of a spatial FGLS, using a consistent estimator for θ:

$$\hat{\beta} = [X'(I_T \otimes B_N' B_N) X]^{-1} X'(I_T \otimes B_N' B_N) y. \tag{18.37}$$

Exploiting the block diagonal nature of $B_N' B_N$, this is equivalent to a regression of the stacked spatially filtered dependent variables, $(I_N - \theta W_N) y_t$ on the spatially filtered explanatory variables $(I_N - \theta W_N) X_t$, as a direct generalization of the single cross-section case.

Two special cases are of particular interest. One is the random effects model with spatial error correlation. Its error variance covariance matrix, equation 18.28, can be simplified in order to facilitate the computation of the determinant and inverse term needed in the log-likelihood. Set $\eta = \sigma_\mu^2 / \sigma_u^2$, such that $\Sigma_{NT} = \sigma_u^2 \Psi_{NT}$, with:

$$\Psi_{NT} = \iota_T \iota_T' \otimes \eta I_N + [I_T \otimes (B_N' B_N)^{-1}],$$

using the same notation and observation stacking as for 18.28. This particular expression allows the determinant and inverse to be obtained as (see Anselin, 1988a, p. 154, for details):

$$|\Psi_{NT}| = |(B_N' B_N)^{-1} + (T\eta) I_N| |B_N|^{-2(T-1)}$$

and,

$$\Psi_{NT}^{-1} = \frac{\iota_T \iota_T'}{T} \otimes [(B_N' B_N)^{-1} + (T\eta) I_N]^{-1} + (I_T - \frac{\iota_T \iota_T'}{T}) \otimes (B_N' B_N).$$

The log-likelihood thus becomes:

$$\begin{aligned} L &= -\frac{NT}{2}\ln\sigma_u^2 + (T-1)\ln|B_N| \\ &\quad -\frac{1}{2}\ln|(B_N'B_N)^{-1} + (T\eta)I_N| \\ &\quad -\frac{1}{2\sigma_u^2}\epsilon'[\frac{\iota_T\iota_T'}{T}\otimes[(B_N'B_N)^{-1} + (T\eta)I_N]^{-1}]\epsilon \\ &\quad -\frac{1}{2\sigma_u^2}\epsilon'[(I_T - \frac{\iota_T\iota_T'}{T})\otimes(B_N'B_N)]\epsilon, \end{aligned}$$

with $\epsilon = y - X\beta$.

A second special case is the the spatial SUR model with spatial SAR error autocorrelation. Its error variance covariance matrix is given by equation 18.22. The required determinant and inverse for the log-likelihood are (see Anselin, 1988a, p. 143):

$$|B_{NT}^{-1}(\Sigma_T\otimes I_N)B_{NT}^{-1'}| = |\Sigma_T|^N|B_{NT}|^{-2},$$

and,

$$[B_{NT}^{-1}(\Sigma_T\otimes I_N)B_{NT}^{-1'}]^{-1} = B_{NT}'[\Sigma_T^{-1}\otimes I_N]B_{NT}.$$

Furthermore, due to the block-diagonal structure of B_{NT}:

$$\ln|B_{NT}| = \sum_t \ln|I_N - \theta_t W_N|.$$

The log-likelihood for this model then follows as:

$$\begin{aligned} L &= -\frac{N}{2}\ln|\Sigma_T| + \sum_t \ln|I_N - \theta_t W_N| \\ &\quad -\frac{1}{2}\epsilon' B_{NT}'(\Sigma_T^{-1}\otimes I_N)B_{NT}\epsilon, \end{aligned}$$

with $B_{NT}\epsilon$ corresponding to the residuals from spatially filtered dependent and explanatory variables, $[I_{NT} - (\Theta_T\otimes W_N)](y - X\beta)$, a generalization of the pooled model case.

4.2 Instrumental Variables and GMM

As an alternative to reliance on an often unrealistic assumption of normality and to avoid some of the computational problems associated with the Jacobian term in ML estimation, instrumental variables and GMM methods have been suggested for single cross-section spatial regression models (e.g., Anselin, 1988a, Anselin, 1990, Kelejian and Robinson, 1993, Kelejian and Prucha, 1998, Kelejian and Prucha, 1999, Conley, 1999). These can be extended to the panel data setting. We will consider the spatial lag and spatial error models in turn.

4.2.1 Spatial Lag Models. The endogeneity of the spatially lagged dependent variable suggests a straightforward instrumental variables strategy in which the spatially lagged (exogenous) explanatory variables WX are used as instruments (Kelejian and Robinson, 1993, Kelejian and Prucha, 1998, and also Lee, 2003 for the choice of optimal instruments). This applies directly to the spatial lag in the pooled model, where the instruments would be $(I_T \otimes W_N)X$ (with X as a stacked $NT \times (K-1)$ matrix, excluding the constant term).

A special case is the spatial SUR model with a spatial lag term, equation 18.19. Following the same approach as taken in the single cross-section, consider the spatially lagged dependent variable and the explanatory variables in each equation grouped into a matrix $Z_t = [W_N y_t \; X_t]$, with parameter vector $\gamma_t = [\rho_t \; \beta_t']'$. The individual Z_t terms can be stacked into a $NT \times T(K+1)$ matrix Z, using the same setup as in equation 18.18, with a matching stacked coefficient vector γ. For each equation, construct a matrix of instruments, $H_t = [X_t \;\; W_N X_t]$, stacked in block-diagonal form into H. With a consistent estimate for the error variance covariance matrix, $\hat{\Sigma}_T \otimes I_N$, the model parameters can be estimated by means of the IV estimator with a general non-spherical error variance (Anselin, 1988a, p. 146):

$$\hat{\gamma} = \left[Z'H[H'(\hat{\Sigma}_T \otimes I_N)H]^{-1}H'Z \right]^{-1} Z'H[H'(\hat{\Sigma}_T \otimes I_N)H]^{-1}H'y \tag{18.38}$$

with an estimate for the coefficient variance as:

$$\text{Var}[\hat{\gamma}] = \left[Z'H[H'(\hat{\Sigma}_T \otimes I_N)H]^{-1}H'Z \right]^{-1}.$$

This suggests an iterative spatial three stages least squares estimator (S3SLS): first estimate each regression using spatial 2SLS (S2SLS); use the S2SLS residuals to obtain a consistent estimate of $\hat{\Sigma}$; and finally use $\hat{\Sigma}$ in equation 18.38. Consistency and asymptotic normality of the spatial generalized IV estimator can be based on the arguments developed for the cross-sectional S2SLS case (Kelejian and Robinson, 1993, Kelejian and Prucha, 1998).

4.2.2 Spatial Error Models. The spatially weighted least squares result (equation 18.37) for the regression parameters in the pooled model with SAR errors also holds in a more general setting, without assuming normality. As long as a consistent estimator for the *nuisance* parameter θ can be obtained, the FGLS estimator will also be consistent for β.

In the single cross-section, a consistent estimator can be constructed from a set of moment conditions on the error terms, as demonstrated in the Kelejian-Prucha generalized moments (KPGM) estimator (Kelejian and Prucha, 1999). These conditions can be readily extended to the pooled model, by replacing the single equation spatial weights by their pooled counterparts $(I_T \otimes W_N)$. The point of departure is the stacked vector of SAR errors:

$$\epsilon = \theta(I_T \otimes W_N)\epsilon + u,$$

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

The three KPGM moment conditions (Kelejian and Prucha, 1999, p. 514) pertain to the idiosyncratic error vector u. Extending them to the pooled setting yields:

$$\begin{aligned} \text{E}[\frac{1}{NT}u'u] &= \sigma_u^2 \\ \text{E}[\frac{1}{NT}u'(I_T \otimes W_N')(I_T \otimes W)u] &= \frac{1}{N}\sigma_u^2 \text{tr}(W_N' W_N) \\ \text{E}[\frac{1}{NT}u'(I_T \otimes W_N)u] &= 0, \end{aligned}$$

where tr is the matrix trace operator and use is made of $\text{tr}(I_T \otimes W_N' W_N) = T\text{tr}W_N' W_N$, and $\text{tr}(I_T \otimes W_N) = 0$.

The estimator is made operational by substituting $u = \epsilon - \theta(I_t \otimes W_N)\epsilon$, and replacing ϵ by the regression residuals. The result is a system of three equations in θ, θ^2 and σ_u^2, which can be solved by nonlinear least squares (for technical details, see Kelejian and Prucha, 1999). Under some fairly general regularity conditions, substituting the consistent estimator for θ into the spatial FGLS (equation 18.37) will yield a consistent estimator for β. Recently, this approach has been extended to the error components model with spatial error dependence (18.29), yielding a system of six moment equations (for details, see Kapoor et al., 2003).

5. Testing for Spatial Dependence

Testing for spatial effects in spatial panel models centers on the null hypotheses $\text{H}_0 : \rho = 0$ and/or $\text{H}_0 : \theta = 0$ in the various models that include spatial lag terms or spatial error autocorrelation. Arguably, the preferred approach is based on Lagrange Multiplier (LM) or Rao Score (RS) tests, since these only require estimation of the model under the null, avoiding the complexities associated with ML estimation (for a recent review, see Anselin, 2001a). The test statistics developed for the single cross-section case can be readily extended to the pooled model. In addition, specialized diagnostics have been developed to test for spatial effects in spatial SUR (Anselin, 1988b), and for error components models (Anselin, 1988a, Baltagi et al., 2003a, Baltagi et al., 2003b). More recently, a strategy has been suggested to test for general unspecified cross-sectional dependence (Pesaran, 2004).

We focus our attention on the LM tests and first briefly review the generic case. This is followed by an illustration of applications of the LM principle to tests against error correlation in the spatial SUR and error components models.

5.1 Lagrange Multiplier Tests for Spatial Lag and Spatial Error Dependence in Pooled Models

The results for the pooled models follow as straightforward generalizations of the single cross-section case, with proper adjustments for the spatial weights matrix and weights matrix traces. Consider the pooled regression model 18.2 as the point of departure, with $e = y - X\hat{\beta}$ as a $NT \times 1$ vector of regression residuals.

The single cross-section Lagrange Multiplier test statistic for spatial error correlation, LM_E (Burridge, 1980), which is asymptotically distributed as $\chi^2(1)$, is readily extended to the pooled model with spatial weights matrix $(I_T \otimes W_N)$ as:

$$LM_E = \frac{[e'(I_T \otimes W_N)e/(e'e/NT)]^2}{\text{tr}[(I_T \otimes W_N^2) + (I_T \otimes W_N'W_N)]}$$

or, using simplified trace terms:

$$LM_E = \frac{[e'(I_T \otimes W_N)e/(e'e/NT)]^2}{T\text{tr}(W_N^2 + W_N'W_N)}.$$

Similarly, the single cross-section LM test statistic for a spatial lag alternative, LM_L (Anselin, 1988a), becomes:

$$LM_L = \frac{[e'(I_T \otimes W_N)y/(e'e/NT)]^2}{[(W\hat{y})'M(W\hat{y})/\hat{\sigma}^2] + T\text{tr}(W_N^2 + W_N'W_N)}$$

with $W\hat{y} = (I_T \otimes W_N)X\hat{\beta}$ as the spatially lagged predicted values in the regression, and $M = I_{NT} - X(X'X)^{-1}X'$. This statistic is also asymptotically distributed as $\chi^2(1)$.

This simple approach can be generalized to account for more realistic error variance structures, such as heteroskedasticity across the time periods, in the same manner that heteroskedasticity is included in test statistics for the single cross-section (see, e.g., Kelejian and Robinson, 1998). Alternatively, each of the test statistics can be robustified against the alternative of the other form, using the standard approach (see Anselin et al., 1996).

5.2 Testing for Spatial Error Correlation in Panel Data Models

5.2.1 Spatial SUR Model. In the spatial SUR model (equations 18.15–18.17), the LM test statistics are based on the residuals from a standard ML or

FGLS estimation. In contrast to the pooled model, the null hypothesis pertains to T parameter constraints, $\mathrm{H}_0 : \theta_1 = \ldots = \theta_T = 0$ for the spatial error alternative.

To construct the statistic, consider a $N \times T$ matrix E with the $N \times 1$ individual equation residual vectors as columns. The LM_E test statistic then follows as (Anselin, 1988b):

$$LM_E = \iota_T'(\hat{\Sigma}_T^{-1} * E'W_N E)J^{-1}(\hat{\Sigma}_T^{-1} * E'W_N E)'\iota_T$$

with $*$ as the Hadamard product, and

$$J = [\mathrm{tr}(W_N^2)]I_T + [\mathrm{tr}(W_N'W_N)](\hat{\Sigma}_T^{-1} * \hat{\Sigma}_T)$$

The LM_E statistic is distributed asymptotically as $\chi^2(T)$.

5.2.2 Error Components Models. In the error components model with spatial autoregressive errors (equations 18.25–18.26), the null hypothesis is $\mathrm{H}_0 : \theta = 0$. A LM test statistic can be constructed from the residuals obtained by estimating the standard error components model by FGLS or ML. With e as the $NT \times 1$ residual vector, and, to simplify notation, with $\hat{\kappa} = (\hat{\sigma}_\mu^2/\hat{\sigma}_u^2)/[1+T(\hat{\sigma}_\mu^2/\hat{\sigma}_u^2)]$, the test statistic follows as (Anselin, 1988a, p. 155):

$$LM_E = \frac{\left[(1/\hat{\sigma}_u^2)e'[[I_T + \hat{\kappa}(T\hat{\kappa} - 2)\iota_T\iota_T'] \otimes W_N]e\right]^2}{p},$$

with $p = (T^2\hat{\kappa}^2 - 2\hat{\kappa} + T)(\mathrm{tr}W_N^2 + \mathrm{tr}W_N'W_N)$. It is distributed asymptotically as $\chi^2(1)$.

When the point of departure is not the error components model, but the pooled specification 18.2, 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 resulting 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 $\mathrm{H}_0 : \theta = 0$ (the spatial parameter) or on $\mathrm{H}_0 : \sigma_\mu^2 = 0$ (the error component), based on the residuals of the pooled model. A joint test is on $\mathrm{H}_0 : \theta = \sigma_\mu^2 = 0$, and conditional tests are for $\mathrm{H}_0 : \theta = 0$ (assuming $\sigma_\mu^2 \geq 0$), or $\mathrm{H}_0 : \sigma_\mu^2 = 0$ (assuming θ 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 with parameter ϕ, as in equations 18.32–18.33 (for detailed derivations, see Baltagi et al., 2003a).

6. Conclusions

The econometrics of panel data models with spatial effects constitutes an active area of research, as evidenced by a growing number of recent papers on the topic. The focus to date has been primarily on theoretical and methodological aspects. Arguably, the dissemination of these methods to empirical practice has been hampered by the lack of ready to use software. None of the standard econometric packages include built-in facilities to carry out single cross-section spatial econometrics, let alone spatial panel econometrics.

For single cross-section spatial econometrics, there are now several software resources available, ranging from freestanding packages such as GeoDa (Anselin et al., 2004b), to collections of routines in Matlab and R (Bivand, 2002). However, for panel spatial econometrics, the situation is bleak. A promising development in this regard is the effort under the auspices of the U.S. Center for Spatially Integrated Social Science (CSISS) to develop software for spatial econometrics in the open source Python language. The PySpace collection of modules that is currently under active development includes the basic tests and estimation methods for the pooled panel model as well as the spatial SUR model (Anselin and Le Gallo, 2004).

While much progress has been made, many areas remain where very little insight has been gained into the complexities that result from explicitly introduction spatial dependence and spatial heterogeneity into panel data models. Directions with particular promise for future research would be the extension to models with discrete dependent variables. Also of particular interest to applied researchers would be greater insight into the trade offs involved in using strategies for general cross-sectional dependence relative to the use of parameterized spatial processes.

It is hoped that the review provided in the current chapter may provide a stimulus and resource to theoretical and applied researchers alike to aid in pursuing these directions in the future.

Notes

1. The treatment of spatial effects in panel data models with discrete dependent variables is still in its infancy.
2. Note that we couch the discussion using "time" as the second dimension for the sake of simplicity. In general, it is also possible to have the second dimension reflect another cross-sectional characteristic, such as an industry sector, and. along the same lines, extension to higher order panel structures are possible as well.
3. A more extensive technical review can be found in Anselin and Bera, 1998.
4. In the literature of spatial statistics, spatially varying coefficients are treated in (Bayesian) hierarchical models (Gelfand et al., 2003, Gamerman et al., 2003).
5. In what follows, we will use the symbol W for the spatial weights and assume row-standardization.
6. An extensive discussion of spatial weights is outside the scope of this chapter. For a detailed assessment of technical issues, see the recent review papers by Anselin and Bera, 1998, and Anselin, 2002.

7. Since the spatial weights enter into a model premultiplied by a scalar parameter, changes in the interaction structure over time can be accounted for by allowing this parameter to vary. Alternatively, but less tractable, would be to let the weights vary and keep the parameter constant. Obviously, letting both parameter and weights vary over time would lead to problems with identification and interpretation (for example, see Druska and Horrace, 2004).

8. In spatial statistics, the preferred perspective is that of a conditional process, which is geared to spatial prediction (see Cressie, 1993, Stein, 1999, and for a discussion of the implications in a spatial econometric context, Anselin, 2002). Rather than specifying the joint distribution of all the y_i in the system, each y_i is modeled conditional upon the y_j for the neighbors. For detailed discussion, see the previous references.

9. This can be relaxed in more flexible space-time models, see, for example, section 18.3.3.1.

10. For row-standardized weights, $\rho = 1$ violates a standard regularity condition for spatial models that requires that the inverse $(I - \rho W_N)^{-1}$ exists (Kelejian and Prucha, 1999).

11. For the sake of simplicity, we use a homoskedastic model with constant variance across all time periods. This restriction can be readily relaxed. Similarly, the assumption of isotropy (only distance matters, not direction) may be relaxed by including separate functions to account for directional effects.

12. In the notation that follows, we use the subscripts T, N and NT to refer to the dimension of the matrix, and the subscript t to refer to the cross-section at time t.

13. Note that this simplification only holds in the strictly homogeneous case with $\sigma_t^2 = \sigma^2 \; \forall t$.

14. For example, with W_N specified as first order contiguity, the spatial covariance in equation 18.10 only includes first and second order neighbors.

15. We are assuming the same number of explanatory variables (K) in each equation, but this can be readily generalized to allow the number of explanatory variables to vary by time period.

16. Explicit space-time dependence is treated in section 18.3.3.

17. The notion of separable stationary spatio-temporal processes originates in the geostatistical literature, but can be readily applied to the current framework. Extension to non-separable structures have been suggested in the recent literature (e.g., Cressie and Huang, 1999).

18. In the statistical literature, specifications of space-time dependence are often conceptualized as hierarchical or multilevel models. This is beyond the scope of our current review (see, for example, Waller et al., 1997a, Waller et al., 1997b, Wikle et al., 1998, Banerjee et al., 2004, and the extensive set of references therein).

19. Models with other forms for the error dependence have seen limited attention in a panel data context and are not considered here.

20. In practice, the log Jacobian is used, with $\ln |I_N - \rho W_N| = \sum_i \ln(1 - \rho\omega_i)$.

References

Andrews, Donald W.K. (2005). Cross-section regression with common shocks. *Econometrica*, 73:1551–1585.

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

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

Anselin, Luc (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics*, 20:141–163.

Anselin, Luc (2001a). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, 97:113–139.

Anselin, Luc (2001b). Spatial econometrics. In Baltagi, Badi, editor, *A Companion to Theoretical Econometrics*, pages 310–330. Blackwell, Oxford.

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

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

Anselin, Luc and Bera, Anil (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, Amman and Giles, David E.A., editors, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York.

Anselin, Luc, Bera, Anil, Florax, Raymond J.G.M., and Yoon, Mann (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26:77–104.

Anselin, Luc and Florax, Raymond J.G.M. (1995). *New Directions in Spatial Econometrics*. Springer-Verlag, Berlin.

Anselin, Luc, Florax, Raymond J.G.M., and Rey, Sergio J. (2004a). Econometrics for spatial models, recent advances. In Anselin, Luc, Florax, Raymond J.G.M., and Rey, Sergio J., editors, *Advances in Spatial Econometrics. Methodology, Tools and Applications*, pages 1–25. Springer-Verlag, Berlin.

Anselin, Luc and Le Gallo, Julie (2004). *Panel Data Spatial Econometrics with PySpace*. Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

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

Anselin, Luc, Syabri, Ibnu, and Kho, Youngihn (2004b). GeoDa, an introduction to spatial data analysis. *Geographical Analysis*. forthcoming.

Arellano, Manuel (2003). *Panel Data Econometrics*. Oxford University Press, Oxford, United Kingdom.

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

Baltagi, Badi H., Song, Seuck Heun, Jung, Byoung Cheol, and Koh, Won (2003a). Testing for serial correlation, spatial autocorrelation and random effects using panel data. Working paper, Texas A&M University, College Station, TX.

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

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

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

Bivand, Roger (2002). Spatial econometrics functions in R: Classes and methods. *Journal of Geographical Systems*, 4:405–421.

Brock, William A. and Durlauf, Steven N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68(2):235–260.

Brueckner, Jan K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, 26(2):175–188.

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

Case, Anne C. (1991). Spatial patterns in household demand. *Econometrica*, 59:953–965.

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

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

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

Chen, Xiaoheng and Conley, Timothy G. (2001). A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 105:59–83.

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

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

Coakley, Jerry, Fuentes, Ana-Maria, and Smith, Ron (2005). Unobserved heterogeneity in panel time series models. *Computational Statistics and Data Analysis*. in press.

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

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

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

Cressie, N. and Huang, H.-C (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94:1330–1340.

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

Driscoll, John C. and Kraay, Aart C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80:549–560.

Druska, Viliam and Horrace, William C. (2004). Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics*, 86(1):185–198.

Dubin, Robin (1988). Estimation of regression coefficients in the presence of spatially autocorrelated errors. *Review of Economics and Statistics*, 70:466–474.

Dubin, Robin (1995). Estimating logit models with spatial dependence. In Anselin, Luc and Florax, Raymond J.G.M., editors, *New Directions in Spatial Econometrics*, pages 229–242. Springer-Verlag, Berlin.

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

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

Fazekas, Istvan, Florax, Raymond, and Folmer, Hendrik (1994). On maximum likelihood estimators of parameters of spatio-temporal econometric models. Technical Report No. 109/1994, Kossuth University, Debrecen, Hungary.

Florax, Raymond J.G.M. and Van Der Vlist, Arno J. (2003). Spatial econometric data analysis: Moving beyond traditional models. *International Regional Science Review*, 26(3):223–243.

Fotheringham, A. Stewart, Brunsdon, Chris, and Charlton, Martin (2002). *Geographically Weighted Regression*. John Wiley, Chichester.

Gamerman, Dani, Moreira, Ajax R.B., and Rue, Havard (2003). Space-varying regression models: Specifications and simulation. *Computational Statistics & Data Analysis*, 42(3):513–533.

Gelfand, Alan E., Kim, Hyon-Jung, Sirmans, C.F., and Banerjee, Sudipto (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98:387–396.

Giacomini, Raffaella and Granger, Clive W.J. (2004). Aggregation of space-time processes. *Journal of Econometrics*, 118:7–26.

Glaeser, Edward L., Sacerdote, Bruce I., and Scheinkman, Jose A. (2002). The social multiplier. Technical Report 9153, NBER, Cambridge, MA 02138.

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

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

Kapetanios, G. and Pesaran, M. Hashem (2005). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset results. In Phillips, Garry and Tzavalis, Elias, editors, *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge University Press, Cambridge, UK. forthcoming.

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

Kelejian, Harry H. and Prucha, Ingmar (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, Harry H. and Prucha, Ingmar (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.

Kelejian, Harry H. and Robinson, Dennis 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, Harry H. and Robinson, Dennis P. (1995). Spatial correlation: A suggested alternative to the autoregressive model. In Anselin, Luc and Florax, Raymond J.G.M., editors, *New Directions in Spatial Econometrics*, pages 75–95. Springer-Verlag, Berlin.

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

Lee, Lung-Fei (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, 18(2):252–277.

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

Magnus, Jan (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.

Manski, Charles F. (1993). Identification of endogenous social effects: The reflexion problem. *Review of Economic Studies*, 60:531–542.

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

Mardia, K.V. and Goodall, C. (1993). Spatio-temporal analyses of multivariate environmental monitoring data. In Patil, G.P. and Rao, C.R., editors, *Multivariate Environmental Statistics*, pages 347–386. Elsevier, Amsterdam.

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

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

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

Paelinck, Jean and Klaassen, Leo (1979). *Spatial Econometrics*. Saxon House, Farnborough.

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

Pesaran, M. Hashem (2005). Estimation and inference in large heterogenous panels with cross section dependence. Technical report, University of Cambridge, Cambridge, United Kingdom.

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

Smirnov, Oleg and Anselin, Luc (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics and Data Analysis*, 35:301–319.

Stein, Michael L. (1999). *Interpolation of Spatial Data, Some Theory for Kriging*. Springer-Verlag, New York.

Topa, Giorgio (2001). Social interactions, local spillover and unemployment. *Review of Economic Studies*, 68(2):261–295.

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

Waller, Lance, Carlin, Bradley, and Xia, Hong (1997a). Structuring correlation within hierarchical spatio-temporal models for disease rates. In Gregoire, 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, Lance, Carlin, Bradley, Xia, Hong, and Gelfand, Alan (1997b). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92:607–617.

Wikle, Christopher K., Berliner, L. Mark, and Cressie, Noel (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.