INTERNATIONAL REGIONAL SCIENCE REVIEW 26, 3: 244–268 (July 2003)

SPECIFICATION AND ESTIMATION OF SPATIAL PANEL DATA MODELS

J. PAUL ELHORST
Faculty of Economics, University of Groningen, The Netherlands,
J.P.Elhorst@eco.rug.nl

This article provides a survey of the specification and estimation of spatial panel data models. These models include spatial error autocorrelation, or the specification is extended with a spatially lagged dependent variable. In particular, the author focuses on the specification and estimation of four panel data models commonly used in applied research: the fixed effects model, the random effects model, the fixed coefficients model, and the random coefficients model. The survey discusses the asymptotic properties of the estimators and provides guidance with respect to the estimation procedures, which should be useful for practitioners.

Keywords*:* *panel data; spatial dependence; maximum likelihood technique; spatial heterogeneity; spatial econometrics*

In recent years, there has been a growing interest in the specification and estimation of econometric relationships based on panel data. This interest can be explained by the fact that panel data offer researchers extended modeling possibilities as compared to purely cross-sectional data or time-series data. Panel data are generally more informative, and they contain more variation and less collinearity among the variables. The use of panel data results in a greater availability of degrees of freedom and hence increases efficiency in the estimation. Panel data also allow for the specification of more complicated behavioral hypotheses, including effects that cannot be addressed using pure cross-sectional or time-series data (Hsiao 1986; Baltagi 2001).

Two problems may arise when panel data incorporate a locational component. The first problem is that spatial dependence may exist between the observations at each point in time. The fact that distance affects economic behavior is the main reason for an observation associated with a specific location to be dependent on observations at other locations. Regional science theory points out that economic agents may change their decisions depending on (1) market conditions in the region of location as compared to other regions and (2) the distance between regions. When specifying the spatial dependence between observations, the model may incorporate a spatial autoregressive process in the error term, or the model may contain a

DOI: 10.1177/0160017603253791

spatially autoregressive dependent variable. The first model is known as the *spatial error* model and the second as the *spatial lag* model (for the introduction of these terms, see Anselin and Hudak 1992).

The second problem potentially arising when panel data have a locational component is that parameters are not homogeneous over space but instead vary over different geographical locations. Coefficients can of course also vary over time, but this complication is not discussed in this article. Parameter heterogeneity across spatial units has become a topical issue in the literature. Pesaran and Smith (1995) and especially Fotheringham, Charlton, and Brunsdon (1997)[1] advocate that we abandon the fundamental assumption of homogeneous parameters underlying pooled models and refrain from relying on estimated average responses from individual regressions. Similar to cross-sectional regressions, the main problem of traditional panel data techniques is that they will only capture "average" or representative behavior. A panel data regression with constant slopes results in average effects across spatial units, even when allowing for a variable intercept, and it does not show the differences in behavior among individual spatial units (Quah 1996a, 1996b). A second reason for an estimated relationship to exhibit spatial variation is that the model may constitute a gross misspecification of reality because one or more relevant variables have erroneously been omitted from the model or are captured through an incorrect functional form.

It should be noted that the upsurge in the use of heterogeneous panel data estimators has also been criticized. First, it has been argued that heterogeneous panel data estimators produce less plausible estimates as compared to their pooled homogeneous counterparts (Baltagi and Griffin 1997; Baltagi, Griffin, and Xiong 2000). Second, the estimation of individual time-series models—in effect, one for each spatial unit—ignores the comovements across spatial units (Quah 1996b). Finally, the use of heterogeneous panel data models is only feasible when the number of observations on each spatial unit is large enough because these models effectively consist of separate regressions for each spatial unit. Most panel data sets do not meet this requirement.

This article surveys panel data models in which spatial dependence and spatial heterogeneity are incorporated. We deal with four panel data models commonly used in applied research (fixed effects, random effects, fixed coefficients, and random coefficients models) and extend the traditional models to include spatial error autocorrelation or a spatially lagged dependent variable. As a result, this article examines eight different models. Spatial heterogeneity is expressed in the coefficients of the explanatory variables. In the fixed and random effects models, the intercept is allowed to vary over spatial units, and in the fixed and random coefficients models, both the intercept and the slope coefficients are allowed to vary over spatial units.

We focus specifically on details of the estimation method and discuss complications that may arise from limitations of the available software as well. The treatment of these topics is motivated by the current lack of complete and detailed coverage of spatial panel data models in the literature. Anselin (1988) discussed some of these models in his seminal textbook on spatial econometrics. He covered the random effects model with spatially correlated errors at length (pp. 150-56, 164-66; see also Baltagi 2001, 195-97), but the fixed coefficients spatial lag model (p. 156) and the random coefficients spatial error model (pp. 129-31) are discussed only briefly. Anselin also closely examined two models that are similar to the fixed coefficients spatial error and spatial lag models (pp. 137-50, 157-63). However, whereas Anselin allows the coefficients to vary across time, we allow them to vary over space, following earlier work by Pesaran and Smith (1995) and Fotheringham, Charlton, and Brunsdon (1997).

The spatial econometric literature has shown that ordinary least squares (OLS) estimation is inappropriate for models incorporating spatial effects. In the case of spatial error autocorrelation, the OLS estimator of the response parameters remains unbiased, but it loses the efficiency property. In the case when the specification contains a spatially lagged dependent variable, the OLS estimator of the response parameters not only loses the property of being unbiased but also is inconsistent. The latter is, however, a minimal requirement for a useful estimator, and it is therefore commonly suggested to overcome these problems by using maximum likelihood techniques (Anselin 1988; Anselin and Hudak 1992). We adopt this principle of using maximum likelihood techniques, unless this approach is too complicated or not applicable. As the derivation of maximum likelihood functions is one of the main purposes of this article, we assume that the density function of the errors is always fully defined. Lee (2001a, 2001b) has recently investigated asymptotic properties of (quasi) maximum likelihood estimators for spatial cross-section models, assuming that the number of spatial cross sections approaches infinity. We will return to this issue below.

Recently, two studies proposed nonparametric covariance estimation techniques (specifically, general methods of moments [GMM]), yielding standard error estimates of the response parameters that are robust to spatial dependence among the error terms in spatial cross-section models (Conley 1999; Kelejian and Prucha 1999). Driscoll and Kraay (1998) and Bell and Bockstael (2000) applied GMM estimators in the context of spatial panel data sets. Although our focus is on parametric approaches to modeling, we also discuss the conditions under which GMM is useful as an alternative to the scalar indexed dependence among the error terms.

Spatial panel data models with spatial error autocorrelation have received more attention in the regional science literature than panel data models, including a spatially lagged dependent variable. There are three reasons for considering both types of models. First, the spatial extensions to the traditional panel data models are quite different from each other. The difference is aggravated when the intercepts and/or the slopes are assumed to vary with location and when the coefficients are treated as randomly distributed over space. Second, neither of the spatial extensions to the traditional panel data models is straightforward. Third, the extension of traditional panel data models with a spatially lagged dependent variable is likely to be relevant for further work on spatial unit roots, spatial cointegration, and the problem that regressing nonstationary data containing spatial unit roots can produce spurious regressions (Fingleton 1999; see also Mur and Trívez 2003 [this issue]).

The organization of this article is as follows. First, we introduce the four most commonly used panel data models, assuming relationships that are linear in the coefficients (eventually after a suitable transformation) and in which the regressand is a continuous variable. Subsequently, we extend these traditional panel data models to include spatial error autocorrelation or a spatially lagged dependent variable, and we explain how these models can be estimated. We do not cover dynamic effects and hence rule out the situation in which a serially lagged dependent variable appears on the right-hand side of the regression equation. This extension introduces additional econometric complications, and these are beyond the scope of the current article. The presentation of each spatial panel data model concludes with a discussion of the feasibility of the model, the asymptotic justification, and potential extensions and/or alternative estimation methods.

A Taxonomy of Panel Data Models

The analysis starts from a simple linear model between a dependent variable Y and a set of K independent variables X:

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_K X_{Kit} + \varepsilon_{it} = \beta' X_{it} + \varepsilon_{it}, \tag{1}$$

where i (= 1, . . ., N) refers to a spatial unit, t (= 1, . . ., T) refers to a given time period, $\beta_1, \ldots, \beta_K$ are fixed but unknown parameters, and ε_{it} are independently and identically distributed (i.i.d.) error terms for all i and t, with zero mean and variance σ^2. To avoid the breakdown of the asymptotic properties of the maximum likelihood estimator (MLE), we assume that the regression model contains at least one spatially varying regressor that is not irrelevant, implying that its coefficient β is unequal to zero (Lee 2001b).

The main objection to this model is that it does not account for spatial heterogeneity. Spatial units are likely to differ in their background variables, which are usually space-specific, time-invariant variables that do affect the dependent variable but are difficult to measure or hard to obtain. Failing to account for these variables, however, increases the risk of obtaining biased estimation results. One remedy is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit considered:

$$Y_{it} = \beta' X_{it} + \mu_i + \varepsilon_{it}, \tag{2a}$$

or in stacked form:

$$Y_t = X_t \beta + \mu + \varepsilon_t, \tag{2b}$$

where $Y_t = (Y_{1t}, \ldots, Y_{Nt})'$, $X_t = (X'_{1t}, \cdots, X'_{Nt})'$, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, and $\mu = (\mu_1, \ldots, \mu_N)'$. Conditional on the specification of the variable intercept, the regression equation can be estimated as a fixed or random effects model. In the fixed effects model, a dummy variable is introduced for each spatial unit as a measure of the variable intercept. In the random effects model, the variable intercept is treated as a random variable that is i.i.d. distributed with zero mean and variance σ^2_μ. Furthermore, it is assumed that the random variables μ_i and ε_{it} are independent of each other.

Although the variable intercept model accommodates spatial heterogeneity to a certain extent, the problem remains as to whether the data in such a model are pooled correctly. When spatial heterogeneity is not completely captured by the variable intercept, a natural generalization is to let the slope parameters of the regressors vary as well. The slope parameters can also be considered fixed or randomly distributed between spatial units. If the parameters are fixed but different across spatial units, each spatial unit is treated separately. For instance, for $Y_i = X_i\beta_i + \varepsilon_i$ being the ith equation in a set of N equations, with the observations stacked by spatial unit over time, the only way of relating the N separate regressions is to assume correlation between the error terms in different equations, a phenomenon that is known as *contemporaneous error correlation*. Such a specification is reasonable when the error terms for different spatial units, at a given point in time, are likely to reflect some common immeasurable or omitted factor. In full-sample notation, the set of N equations can be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_N \end{bmatrix}, \tag{3}$$

where $E(\varepsilon_i) = 0$, $E(\varepsilon_i \varepsilon_j') = \sigma^2_{ij} I_T$, and $i, j = (1, \ldots, N)$. This model is also known as the seemingly unrelated regressions (SUR) model.

If the parameters are treated as outcomes of random experiments between spatial units, the data can be pooled into one model to estimate the unknown parameters. This is known as the Swamy random coefficients model (Swamy 1970):

$$Y_i = X_i\beta + \varepsilon_i,\ E(\varepsilon_i) = 0,\ E(\varepsilon_i\varepsilon_i') = \sigma_i^2 I_T, \tag{4a}$$

$$\beta_i = \beta + v_i,\ E(v_i) = 0,\ E(v_i v_i') = V, \tag{4b}$$

where the β_i applying to a particular spatial unit are the outcome of a random process with a common mean coefficient vector β and covariance matrix V, which is a symmetric $(K \times K)$ matrix. In addition, it is assumed that $E(\varepsilon_i \varepsilon_j') = 0$ and $E(v_i v_j') = 0$ for $j \neq i$ and that the random variables ε_{it} and v_i are independent of each other.

Before extending these four traditional panel data models with spatial error autocorrelation or a spatially lagged dependent variable, we introduce the following notation. Let W denote a $(N \times N)$ spatial weight matrix describing the spatial arrangement of the spatial units and w_{ij} the (i,j)th element of W, where i and $j = (1, \ldots, N)$. It is assumed that W is a matrix of known constants, that all diagonal elements of the weights matrix are zero, and that the characteristic roots of W, denoted ω_i, are known. The first assumption excludes the possibility that the spatial weight matrix is parametric. The second assumption implies that no spatial unit can be viewed as its own neighbor, and the third assumption presupposes that the characteristic roots of W can be computed accurately using the computing technology typically available to empirical researchers. The latter is also needed to ensure that the log-likelihood function of the models we distinguish can be computed.

Lee (2001a, 2001b) pointed out that the familiar asymptotic properties of the maximum likelihood estimator, such as $\sqrt{N}$ consistency, depend on the characteristic features of the spatial weight matrix. According to Kelejian and Prucha (1999), the row and column sums of W must be bounded uniformly in absolute value as $N \to \infty$. When the spatial weight matrix is a binary contiguity matrix, this condition is satisfied. Lee (2001a) showed that this condition can be made less strict: the row and column sums should not diverge to infinity at a rate equal to or faster than the rate of the sample size N in the cross-section domain. When the spatial weight matrix is an inverse distance matrix, this condition is satisfied. This can be seen as follows. Consider an infinite number of spatial units that are linearly arranged (to simulate one particular row of the spatial weight matrix). The distance of each spatial unit to its first left- and right-hand neighbor is 1; to its second left- and right-hand neighbor, the distance is 2; and so on. When the off-diagonal elements of W are of the form $1/d_{ij}$, where d_{ij} is the distance between spatial units i and j, the row sum of W equals $\Sigma_{i=1}^{N}\ 2/d_{ij}$, representing a series that is not finite. By contrast, the ratio $1/N\ \ \Sigma_{i=1}^{N}\ 2/d_{ij} \to 0$ as N goes to infinity. Finally, it is possible that these conditions are not sufficient unless panel data are available (Kelejian and Prucha 2002). Such a complicated situation occurs when all the nondiagonals of the spatial weight matrix are equal, $w_{ij} = 1/(N-1)$ for $i \neq j$ (Lee 2001a).

THE FIXED EFFECTS SPATIAL ERROR AND SPATIAL LAG MODEL

The traditional fixed effects model extended to include spatial error autocorrelation can be specified as

$$Y_t = X_t\beta + \mu + \phi_t,\ \phi_t = \delta W\phi_t + \varepsilon_t,\ E(\varepsilon_t) = 0,\ E(\varepsilon_t\varepsilon_t^{'}) = \sigma^2 I_N, \tag{5}$$

and the traditional model extended with a spatially lagged dependent variable reads as

$$Y_t = \delta WY_t + X_t\beta + \mu + \varepsilon_t,\ E(\varepsilon_t) = 0,\ E(\varepsilon_t\varepsilon_t^{'}) = \sigma^2 I_N. \tag{6}$$

In the spatial error specification, the properties of the error structure have been changed, whereas in the spatial lag specification, the number of explanatory variables has increased by one. In the spatial error specification, δ is usually called the spatial autocorrelation coefficient, and in the spatial lag specification, it is referred to as the spatial autoregressive coefficient.

The standard estimation method for the fixed effects model is to eliminate the intercepts β_1 and μ_i from the regression equation by demeaning the Y and X variables,[2] then estimate the resulting demeaned equation by OLS and subsequently recover the intercepts β_1 and μ_i (Baltagi 2001, 12-15). It should be noted that only $(\beta_1 + \mu_i)$ are estimable and not β_1 and μ_i separately, unless a restriction such as $\Sigma_i \mu_i = 0$ is imposed.

Instead of estimating the demeaned equation by OLS, it can also be estimated by maximum likelihood (ML). The only difference is that ML estimators do not make corrections for degrees of freedom. The log-likelihood function corresponding to the demeaned equation incorporating spatial error autocorrelation is

$$-\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^{N} \ln(1-\delta\omega_i) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t' e_t,\; e_t = (I-\delta W)[Y_t - \bar{Y} - (X_t - \bar{X})\beta], \tag{7}$$

and with a spatially lagged dependent variable,

$$-\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^{N} \ln(1-\delta\omega_i) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t' e_t,\; e_t = (I-\delta W)(Y_t - \bar{Y}) - (X_t - \bar{X})\beta, \tag{8}$$

where $\bar{Y} = (\bar{Y}_{1.}, \ldots, \bar{Y}_{N.})'$ and $\bar{X} = (\bar{X}_{1.}', \cdots, \bar{X}_{N.}')'$. An iterative two-stage procedure can be used to maximize the log-likelihood function of the first model, and a simple two-stage procedure is available for the second model (Anselin 1988, 181-82). Anselin and Hudak (1992) give instructions on how to implement these procedures in commercial econometric software. One may also use Spacestat or the MATLAB routines of spatial error model (SEM) and spatial lag model (SAR), which are freely downloadable from LeSage's Web site at www.spatial-econometrics.com. Although these routines are written for spatial cross sections, they can easily be generalized to spatial panel models.

A distinct problem of the fixed effects model is related to the so-called incidental parameter problem. Only the slope coefficients can be estimated consistently, in the case of short panels, where T is fixed and $N \to \infty$. The coefficients of the spatial fixed effects cannot be estimated consistently because the number of observations available for the estimation of μ_i is limited to T observations (Anselin 2001). Fortunately, the inconsistency of μ_i is not transmitted to the estimator of the slope coefficients in the demeaned equation since this estimator is not a function of the estimated μ_i. This implies that the large sample properties of the fixed effects model when $N \to \infty$ do apply for the demeaned equation (Lee 2001a, 2001b). It should be stressed that the incidental parameters problem is independent of the extension to spatial error autocorrelation or to the inclusion of a spatially lagged dependent variable since it also occurs without these two extensions. The incidental parameters problem does not matter when β are the coefficients of interest and μ_i are not, which is the case in many empirical applications. The problem disappears in panels where N is fixed and $T \to \infty$.

If the fixed effects model also contains fixed effects for time periods, there are two feasible ways to proceed. First, one may simply add fixed effects for time periods to the set of explanatory variables. This is possible when T is small. Care should be taken concerning the dummy variable trap. For λ_t (t = 1, . . ., T), denoting a dummy referring to the tth time period, either the restriction $\Sigma_t \lambda_t = 0$ should be imposed or one time dummy should be dropped. Second, one can eliminate the intercepts β_1, μ_i, and λ_t from the regression equation by double demeaning of the Y and X variables[3] and proceed as described above. It automatically follows that for short panels, where T is fixed and $N \to \infty$, the fixed effects for time periods can be estimated consistently. This is not the case for the spatial fixed effects. For long panels, where $T \to \infty$ and N is fixed, the spatial fixed effects can be estimated consistently, but the time period fixed effects cannot. Finally, when N and T are of comparable size, the spatial and time period fixed effects can be estimated consistently only when N and T are sufficiently large.

Another potential problem is that for large N, the usual spatial econometric procedures are problematic because the eigenvalues of spatial weight matrices of dimensions over 400 cannot be estimated with sufficient reliability (Kelejian and Prucha 1999). One solution is to use the GMM estimator in the case of the fixed effects spatial error model (Bell and Bockstael 2000). Another solution, based on maximum likelihood estimation, is not to express the Jacobian term in the individual eigenvalues but in the coefficients of a characteristic polynomial (Smirnov and Anselin 2001) or to approximate the Jacobian term in its original form, $\ln|I - \delta W|$, using a Monte Carlo approach (Barry and Pace 1999). The latter procedure is incorporated in the MATLAB routines SEM and SAR mentioned above.

The fixed effects spatial error or spatial lag model can be tested against the spatial error or spatial lag model without fixed effects using the F test spelled out in Baltagi (2001, 14). One can also estimate the fixed effects model without spatial effects and subsequently test this restricted model against the unrestricted models given in (7) and (8) using, for instance, a Lagrange multiplier (LM) test.

The Random Effects Variant

An alternative that avoids the loss of degrees of freedom incurred in the fixed effects model associated with relatively large N is to consider the random effects model. If μ_i is treated as a random variable, we have $E(\mu_i \mu_j') = \sigma_\mu^2$ if $i = j$ and zero otherwise. This model is straightforwardly extended to include spatial error autocorrelation or a spatially lagged dependent variable (see also Baltagi and Li forthcoming). We discuss both models sequentially.

In full-sample notation, the T sets of N observations in the spatial error case may be written as

$$\begin{bmatrix} Y_1 \\ \cdot \\ Y_T \end{bmatrix} = \begin{bmatrix} X_1 \\ \cdot \\ X_T \end{bmatrix} \beta + v, \qquad \textit{with} \quad v = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})\varepsilon, \tag{9}$$

where ι_T is a $(T \times 1)$ vector of unit elements, and $B = I_N - \delta W$. The covariance matrix of v is

$$\Omega = E(vv') = \sigma_\mu^2(\iota_T \iota_T' \otimes I_N) + \sigma^2(I_T \otimes (B'B)^{-1}). \tag{10}$$

Following Magnus (1982), this covariance matrix can be rewritten in such a way that

$$\Omega = E(vv') = \frac{1}{T}\iota_T \iota_T' \otimes \left(T\sigma_\mu^2 I_N + \sigma^2 (B'B)^{-1}\right) + \sigma^2\left(\left(I_T - \frac{1}{T}\iota_T \iota_T'\right) \otimes (B'B)^{-1}\right), \tag{11a}$$

$$\Omega^{-1} = \frac{1}{T}\iota_T \iota_T' \otimes \left(T\sigma_\mu^2 I_N + \sigma^2 (B'B)^{-1}\right)^{-1} + \frac{1}{\sigma^2}\left(I_T - \frac{1}{T}\iota_T \iota_T'\right) \otimes (B'B), \tag{11b}$$

$$|\,\Omega\,| = |\,T\sigma_\mu^2 I_N + \sigma^2 (B'B)^{-1}| \times |\sigma^2 (B'B)^{-1}|^{\,T-1} = \left(\sigma^2\right)^{NT} |\,T\frac{\sigma_\mu^2}{\sigma^2} I_N + (B'B)^{-1}| \times |B|^{-2(T-1)} \tag{11c}$$

If we define $\theta^2 = \sigma_\mu^2 / \sigma^2$, the log-likelihood function becomes

$$\begin{aligned} LogL = {} & \frac{NT}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log|\,T\theta^2 I_N + \left(B'B\right)^{-1}| + (T-1)\sum_{i=1}^{N}\log(1-\delta\omega_i) \\ & - \frac{1}{2\sigma^2}\tilde{e}'\left(\frac{1}{T}\iota_T\iota_T' \otimes \left(T\theta^2 I_N + \left(B'B\right)^{-1}\right)\right)^{-1}\tilde{e} + \frac{1}{2\sigma^2}\tilde{e}'\left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes \left(B'B\right)\tilde{e} \end{aligned} \tag{12}$$

where $\tilde{e} = (\tilde{e}_1, \cdots, \tilde{e}_T)'$ and $\tilde{e}_t = Y_t - X_t\beta$. Note that the matrix $\frac{1}{T}\iota_T\iota_T' \otimes I_N$ averages the observations for each spatial unit over time, and the matrix $\left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes I_N$ refers to the observations of each spatial unit in deviations from their individual mean. These matrices do not change under power transformations for real numbers. The determinant $|T\theta^2 I_N + (B'B)^{-1}|$ can be expressed as a function of the characteristic roots of W (see Griffith 1988, Table 3.1):

$$|\,T\theta^2 I_N + \left(B'B\right)^{-1}| = \prod_{i=1}^{N}\left[T\theta^2 + \frac{1}{(1-\delta\omega_i)^2}\right]. \tag{13}$$

Consequently, the log-likelihood function simplifies to

$$\begin{aligned} LogL = {} & -\frac{NT}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\log\,(1 + T\theta^2(1-\delta\omega_i)^2) \\ & + T\sum_{i=1}^{N}\log\,(1-\delta\omega_i) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t' e_t, \end{aligned} \tag{14}$$

where $e_t = Y_t^* - X_t^* \beta$,

$$Y_t^* = P\overline{Y} + B(Y_t - \overline{Y}) = BY_t + (P - B)\overline{Y} = (I_N - \delta W)Y_t - (P - (I_N - \delta W))\overline{Y},$$

$X_t^* = (I_N - \delta W)X_t - (P - (I_N - \delta W))\overline{X}$, and P is such that $P'P = (T\theta^2 I_N + (B'B)^{-1})^{-1}$. P can be the upper-triangular Cholesky decomposition of $(T\theta^2 I_N + (B'B)^{-1})^{-1}$ or $P = \Lambda^{-1/2}R$, where R is an $(N \times N)$ matrix of which the ith column is the characteristic vector r_i of $(T\theta^2 I_N + (B'B)^{-1})^{-1}$, which is the same as the characteristic vector of the spatial weight matrix W (see Griffith 1988, Table 3.1), $R = (r_1, \ldots, r_N)$, and Λ is a $(N \times N)$ diagonal matrix with the ith diagonal element being $c_i = T\theta^2 + 1/(1 - \delta\omega_i)^2$, which is the characteristic root of the matrix $(T\theta^2 I_N + (B'B)^{-1})$ corresponding to r_i. It is clear that for large N, the numerical determination of P can be problematic.

The parameters β and σ^2 can be solved from their first-order maximizing conditions:

$$\hat{\beta} = \left(x^{*\prime} x^*\right)^{-1}\left(x^{*\prime} y^*\right) \text{ and } \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} e_t' e_t}{NT}, \text{ where } x^* = \begin{bmatrix} X_1^* \\ \cdot \\ X_T^* \end{bmatrix} \text{ and } y^* = \begin{bmatrix} Y_1^* \\ \cdot \\ Y_T^* \end{bmatrix}. \quad (15)$$

Upon substituting $\hat{\beta}$ and $\hat{\sigma}^2$ in the log-likelihood function, the concentrated log-likelihood function of δ and θ^2 is obtained:

$$LogL = C - \frac{NT}{2}\log\left(\sum_{i=1}^{T} e_t' e_t\right) - \frac{1}{2}\sum_{i=1}^{N} \log(1 + \theta^2(1 - \delta\omega_i)^2) + T\sum_{i=1}^{N} \log(1 - \delta\omega_i), \quad (16)$$

where C is a constant ($C = -NT/2 \times \log(2\pi) - NT/2 + NT/2 \times \log(NT)$). One can iterate between β and σ^2, on one hand, and δ and θ^2, on the other, until convergence. The estimator of β, given δ and θ^2, is a generalized least squares (GLS) estimator and can be obtained by OLS regression of the transformed variable Y on the transformed variables X. The variance σ^2, given δ and θ^2, can be obtained from the transformed residuals. Conversely, the estimators θ and σ^2, given β and σ^2, must be attained by numerical methods because the equations cannot be solved analytically. This problem can, however, be easily programmed using, for instance, the optimization toolbox of MATLAB. It should be noted that the estimator of θ^2 will not necessarily be positive. One way of ensuring a positive value is to write θ^2 as $\theta \times \theta$ and to maximize the concentrated log-likelihood function of δ and θ^2 with respect to δ and θ.

In full-sample notation, the T sets of N observations in the spatial lag case may be written as

$$\begin{bmatrix} Y_1 \\ \cdot \\ Y_T \end{bmatrix} = \delta \begin{bmatrix} WY_1 \\ \cdot \\ WY_T \end{bmatrix} + \begin{bmatrix} X_1 \\ \cdot \\ X_T \end{bmatrix} \beta + v, \quad \text{with} \quad v = \left(\iota_T \otimes I_N\right)\mu + \left(I_T \otimes I_N\right)\varepsilon, \quad (17)$$

where ι_T is a $(T \times 1)$ vector of unit elements. The covariance matrix of v is

$$\Omega = E(vv') = \sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + \sigma^2(I_T \otimes I_N). \tag{18}$$

Following Magnus (1982), this covariance matrix can be rewritten as

$$\Omega = E\left(vv'\right) = \left(T\sigma_\mu^2 I_N + \sigma^2\right)\left(\frac{1}{T}\iota_T\iota_T' \otimes I_N\right) + \sigma^2\left(\left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes I_N\right), \tag{19a}$$

$$\Omega^{-1} = \frac{1}{T\sigma_\mu^2 I_N + \sigma^2}\left(\frac{1}{T}\iota_T\iota_T' \otimes I_N\right) + \frac{1}{\sigma^2}\left(\left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes I_N\right), \tag{19b}$$

$$\begin{aligned} |\,\Omega\,| &= |\left(T\sigma_\mu^2 + \sigma^2\right)I_N| \times |\sigma^2 I_N|^{\,T-1} \\ &= \left(T\sigma_\mu^2 + \sigma^2\right)^N \left(\sigma^2\right)^{N(T-1)} = \left(\frac{\sigma^2}{T\sigma_\mu^2 + \sigma^2}\right)^{-N}\left(\sigma^2\right)^{NT}. \end{aligned} \tag{19c}$$

If we define $\theta^2 = \sigma^2 / (T\sigma_\mu^2 + \sigma^2)$, we obtain

$$\begin{aligned} \Omega^{-1} &= \frac{1}{\sigma^2}\left[\theta^2 \frac{1}{T} e_T e_T' \otimes I_N + \left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes I_N\right], \\ \Omega^{-1/2} &= \frac{1}{\sigma}\left[\theta\frac{1}{T}\iota_T\iota_T' \otimes I_N + \left(I_T - \frac{1}{T}\iota_T\iota_T'\right) \otimes I_N\right] = \frac{1}{\sigma}\left[I_{NT} - (1-\theta)\frac{1}{T}\iota_T\iota' \otimes I_N\right], \end{aligned} \tag{20}$$

where the last matrix between square brackets transforms the observations of each spatial unit in deviations from their individual mean premultiplied by $(1 - \theta)$. Consequently, the log-likelihood function can be written as

$$LogL = -\frac{NT}{2}\log(2\pi\sigma^2) + \frac{N}{2}\log\theta^2 + T\sum_{i=1}^{N}\log(1-\delta\omega_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{T} e_T' e_T\ , \tag{21}$$

where $e_t = Y_t^* - X_t^*\beta$, $Y_t^* = BY_t - (1-\theta)\overline{Y} = (I_N - \delta W)Y_t - (1-\theta)\overline{Y}$, and $X_t^* = (I_N - \delta W)X_t - (1-\theta)\overline{X}$.

The parameter θ^2 measures the weight attached to the variation between spatial units. If this weight tends to zero, the random effects spatial lag model reduces to the fixed effects spatial lag model, constituting a model that only uses the variation within the spatial units over time in forming the parameter estimates of δ and β. If the weight tends to unity, the random effects spatial lag model reduces to the standard spatial lag model as given, for instance, in Anselin and Hudak (1992).

The parameters β and σ^2 can be solved from their first-order maximizing conditions:

$$\hat{\beta} = \left(x^{*\prime}x^*\right)^{-1}\left(x^{*\prime}y^*\right) \text{ and } \hat{\sigma}^2 = \frac{\sum_{i=1}^{T} e_t' e_t}{NT}, \text{ where } x^* = \begin{bmatrix} X_1^* \\ \cdot \\ X_T^* \end{bmatrix} \text{ and } y^* = \begin{bmatrix} Y_1^* \\ \cdot \\ Y_T^* \end{bmatrix}. \tag{22}$$

Upon substituting $\hat{\beta}$ and $\hat{\sigma}^2$ in the log-likelihood function, the concentrated log-likelihood function of δ and θ^2 is obtained as

$$LogL = C - \frac{NT}{2}\log\left(\sum_{t=1}^{T} e_t^{*\prime} e_t\right) + \frac{N}{2}\log\theta^2 + T\sum_{t=1}^{N}\log(1-\delta\omega_i), \tag{23}$$

where C is a constant equal to $-NT/2 \times \log(2\pi) - NT/2 + NT/2 \times \log(NT)$. One can iterate between β and σ^2, on one hand, and δ and θ^2, on the other, until convergence occurs. The estimator of β, given δ and θ^2, is a GLS estimator and can be obtained by OLS regression of the transformed variable Y on the transformed variables X. The variance σ^2, given δ and θ^2, can be obtained from the transformed residuals. Conversely, the estimators δ and θ^2, given β and σ^2, must be solved by numerical methods because they cannot be attained analytically. This problem is also straightforward to program using, for example, the optimization toolbox of MATLAB. It should be noted that finding a local maximum cannot be ruled out (Breusch 1987).

The iterative two-stage procedure needed to estimate the parameters of the random effects spatial error and spatial lag model bears similarities to the nonspatial random effects model (Breusch 1987). The difference is that the concentrated log-likelihood function must be maximized for two parameters (δ, θ^2) instead of only one (θ^2). The random effects spatial lag model appears to be simpler than the random effects spatial error model because the log-likelihood function of the latter model contains the additional term $-1/2 \log|T\theta^2 I_N + (B'B)^{-1}|$. Case (1991) combined the random effects spatial error and spatial lag model, although it appears that this additional term is missing from the log-likelihood function. Another complication in the random effects spatial error model is that the matrix P used to transform the variables Y and X may not be reliably estimated for large N.

The parameters of the random effects spatial error and spatial lag model can be consistently estimated when $N \to \infty$, $T \to \infty$, or $N,T \to \infty$, although a problem of the random effects model is that it may not be an appropriate specification when observations on irregular spatial units are used. The spatial units of observation should be representative of a larger population, and the number of units should potentially be able to go to infinity in a regular fashion. When the random effects model is implemented for a given set of irregular spatial units, such as all counties of a state or all regions in a country, the population is sampled exhaustively (Nerlove and Balestra 1996), and the individual spatial units have characteristics that actually set them apart from a larger population (Anselin 1988, 51). Moreover, the assumption of zero correlation between μ_i and the explanatory variables is particularly restrictive. Hence, the fixed effects model is compelling, even when N is large and T is small.

To test the random effects spatial error or spatial lag model against their counterpart without random effects, various well-known tests are available (Baltagi 2001, chap. 4). One can also estimate the random effects model without spatial effects and subsequently test this restricted model against the unrestricted model with the help of an LM test based on (14) or (21). For a test of the random effects spatial error or spatial lag model against its fixed effects counterpart, the Hausman specification test can be used (Baltagi 2001, 65-66). This does cause a practical problem, however. Commercial econometric packages usually do not apply maximum likelihood to estimate the random effects model without spatial effects. This complicates the use of the suggested LM test because maximum likelihood estimation of the standard random effects model, developed along the lines given in Breusch (1987), should be programmed first before the test can be carried out.

THE FIXED COEFFICIENTS SPATIAL ERROR AND SPATIAL LAG MODEL

The fixed coefficients or SUR model given in (3), with one equation for every spatial unit over time and with contemporaneous error correlation, does not have to be changed to cope with the spatial error case since the set of $\sigma_{ij}(i, j = 1, \ldots, N)$ already reflects the interactions between the spatial units. In the literature, this is regarded as an advantage because no a priori assumptions are required about the nature of interactions over space (White and Hewings 1982). The explanation is that the specification of a particular spatial weight matrix does not alter the estimates of the response parameters β, and the estimate of each σ_{ij} immediately adapts itself to the value of w_{ij} by which it is multiplied. As the SUR model is discussed in almost every econometric textbook and available in almost every commercial econometrics software package, it hardly requires any further explanation.

The standard method to attain the maximum likelihood estimates of the parameters in an SUR model is by iterating the feasible GLS procedure. In every iteration, the residuals of the separate regressions are used to update the elements of the covariance matrix $\sigma_{ij} = e_i' \, e_j / T$ until convergence. It should be observed that the estimates of β_i and σ_{ij} obtained by iterating the feasible GLS procedure are equivalent to those that would be obtained by maximizing the log-likelihood function of the model, assuming that there are no restrictions on the response parameters β across or within the equations.

The set of N equations, with one equation for every spatial unit over time, in a model with fixed coefficients and spatially lagged dependent variables can be expressed as

$$\begin{bmatrix} Y_1 & Y_2 & \ldots & Y_N \end{bmatrix} \begin{bmatrix} 1 & -\delta_{21} & \cdot & -\delta_{N1} \\ -\delta_{12} & 1 & \cdot & -\delta_{N2} \\ \cdot & \cdot & \cdot & \cdot \\ -\delta_{1N} & -\delta_{2N} & \cdot & 1 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_N \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_N \end{bmatrix}, \tag{24}$$

or, equivalently,

$$Y\Gamma = XB + \varepsilon,\ E(\varepsilon) = 0,\ E(\varepsilon\varepsilon') = \Sigma \otimes I_N, \text{ with } \Sigma = \sigma_{ii}\, I_T\ (i = 1, \ldots, N). \tag{25}$$

Each equation can also be written as

$$Y_i = \left[Y_1 \ldots Y_{i-1}\, Y_{i+1} \ldots Y_N X_i\right] \times \begin{bmatrix} \delta_{i1} \\ \cdot \\ \delta_{ii-1} \\ \delta_{ii+1} \\ \cdot \\ \delta_{iN} \\ \beta_i \end{bmatrix} + \varepsilon_i \equiv Z_i \eta_i + \varepsilon_i. \tag{26}$$

Note that the δs and the βs are assumed to differ across spatial units. Furthermore, the assumption of contemporaneous error correlation is dropped, and the assumption $E(\varepsilon_i \varepsilon_j') = \sigma_{ij} I_T$ is changed to $E(\varepsilon_i \varepsilon_i' = \sigma_{ii} I_T$ and $E(\varepsilon_i \varepsilon_j') = 0$ for $i \neq j$. Although the latter is not strictly necessary, we have made this change to be able to discriminate between the spatial error specification and the spatial lag specification.

The log-likelihood function and the first-order maximizing conditions of a linear simultaneous equations model are given in Hausman (1975, 1983). Due to dropping the assumption of contemporaneous error correlation, the full-information maximum likelihood (FIML) estimator of each single η_i is

$$\eta_i = \begin{bmatrix} \delta_i \\ \beta_i \end{bmatrix} = \left(\hat{Z}_i' Z_i\right)^{-1} \hat{Z}_i' Z_i, \tag{27a}$$

$$\text{where } \hat{Z}_i = [(XB\Gamma^{-1})_i\; X_i], \text{ while } \sigma_{ii} = \frac{(Y_i - Z_i\eta_i)'(Y_i - Z_i\eta_i)}{T}. \tag{27b}$$

The matrix $XB\Gamma^{-1}$ consists of N columns. In the case where Y_j ($j = 1, \ldots, N$) is an explanatory variable of Y_i ($i = 1, \ldots, N$), the jth column of $XB\Gamma^{-1}$ is part of the matrix of estimated values of Z_i. The matrix of estimated values of Z_i, $\hat{Z}_i$, consists of ($N - 1$ + K) columns: ($N - 1$) columns with respect to the spatially lagged dependent variables explaining Y_i, as well as K columns with respect to the independent variables explaining Y_i. Note that the estimated values of Z_i can also be seen as instrumental variables (Hausman 1975, 1983). As the estimated values of Z_i at the right-hand side of (27) depend on η, equation 27 defines no closed-form solution for η. One can attempt to solve for η by the Jacobi iteration method. We require a solution $\eta = f(\eta)$. The Jacobi iteration method iterates according to $\eta^{h+1} = f(\eta^h)$. This method is available in only a limited number of commercial econometric software packages.[4]

Because a fixed coefficients spatial lag model has different spatial autoregressive coefficients δ for different spatial units, it follows that the Jacobian term, $T \ln |\Gamma|$, cannot be expressed in function of the characteristic roots of the spatial weight matrix. This difference with the fixed and random effects spatial lag models complicates the numerical determination of the FIML estimator. As an alternative, one can use two-stage least squares (2SLS) since this estimator has the same asymptotic distribution as the FIML estimator. The benefit of the 2SLS estimator is that it is considerably easier to compute. The incurred cost is a loss in asymptotic efficiency because 2SLS does not take account of the restrictions on the coefficients within the matrices B and Γ.

Anselin (1988, 137-50) also derived the log-likelihood function for a fixed coefficient model that includes spatial error autocorrelation or a spatially lagged dependent variable, but his case considers response coefficients that are constant across space but vary over time. This model is called spatial SUR.

The efficiency gain in the fixed coefficients spatial error model is greater, the greater the correlation of the disturbances, the less correlation exists among variables across equations and the more correlation exists among variables within an equation (Fiebig 2001). When $\sigma_{ij} = 0$, $i \neq j$, joint estimation of the set of N equations is not required. Shiba and Tsurumi (1988) provided a complete set of LM and likelihood ratio (LR) tests for this null hypothesis. A hypothesis of particular interest is the homogeneity restriction of equal coefficient vectors β. This hypothesis can be investigated using F or LR tests (Greene 1997).

A disadvantage of a model with different parameters for different spatial units is the large number of parameters to be estimated: $(N \times K)$ different regression coefficients (β) and $(1/2N(N+1))$ different (σ) parameters of the (symmetric) covariance matrix in the spatial error model, as well as $(N \times K + N(N-1))$ different regression coefficients (β, δ) and N different (σ) parameters of the (diagonal) covariance matrix in the spatial lag model. These models are therefore only of use when T is large and N is small. Another practical problem is that the value of N in most commercial econometrics software is restricted. For instance, the upper bound on N in LIMDEP (version 7.0) is 20 for both the SUR model and the simultaneous linear equation model.

Driscoll and Kraay (1998) have pointed out that if N is too large relative to T, it will not be possible to estimate all parameters in a manner that yields a nonsingular estimate. In this case, it is necessary to place prior restrictions on the parameters to reduce the dimensionality of the problem. However, even if these restricted estimators are feasible, the quality of the asymptotic approximation used to justify their use is suspect, unless the ratio N/T is close to zero.

One way to reduce the number of parameters of the covariance matrix to N in the spatial error model, which at the same time reestablishes the use of the spatial weight matrix, is obtained by imposing the restrictions $\sigma_{ij} = \delta_i w_{ij}$ for $i \neq j$. These restrictions may be reasonable if one has prior information about the nature of interactions over space. Under these restrictions, the elements of the covariance matrix must be updated by

$$\sigma_{ii} = \frac{e_i' e_i}{T}, \quad \delta_i = \sum_{j=1,\, j \neq i}^{N} w_{ij}\, e_i' e_j \;/\; T \sum_{j=1,\, j \neq i}^{N} w_{ij} \tag{28}$$

in each iteration. Similarly, the number of regression coefficients in the spatial lag model can be reduced to $(N \times K + N)$. Under these circumstances, we obtain

$$\Gamma = \begin{bmatrix} 1 & -\delta_2 w_{21} & \cdot & -\delta_N w_{N1} \\ -\delta_1 w_{12} & 1 & \cdot & -\delta_N w_{N2} \\ \cdot & \cdot & \cdot & \cdot \\ -\delta_1 w_{1N} & -\delta_2 w_{2N} & \cdot & 1 \end{bmatrix} \text{ and } \hat{Z}_i = \sum_{j=1}^{N} w_{ij} \left[XB\Gamma^{-1} \right]^j, \tag{29}$$

where $[XB\Gamma^{-1}]^j$ denotes the jth column of the matrix $XB\Gamma^{-1}$. Although these restrictions simplify the estimation procedure, use of the Jacobi iteration method cannot be avoided.

In both cases, the number of parameters still depends on N, causing the appropriateness of the asymptotic approximation to be suspect. An alternative, more rigorous, way to reduce the number of parameters is to make a compromise between estimating a uniform equation that is valid for all spatial units and a separate equation for each single spatial unit. First, homogeneous spatial units are joined within groups, and then a separate equation is considered for each group. Schubert (1982) used this approach in building an interregional labor market model for Austria, and Murphy and Hofler (1984) used it for estimating a regional unemployment rate equation for the United States. Froot (1989) suggested this approach, in more formal terms, in the accounting and finance literature to deal with cross-sectional time-series data of firms. In addition, one can choose between spatial dependence among the observations within groups (as in Froot 1989) or spatial dependence between groups. The former may be applicable when neighboring spatial units are grouped, and the latter may be applicable when spatial units with comparable characteristics are put together. Let P denote the number of groups p (= 1, . . ., P) and N_p the number of spatial units in each group, so that $\Sigma_p N_p = N$. Then, the number of parameters for spatial dependence within groups reduces to $P \times K + \Sigma_p \tfrac{1}{2} N_p (N_p + 1)$ in the spatial error model and to $P \times K + \Sigma_p N_p (N_p - 1) + P$ in the spatial lag model. In the case of spatial dependence between groups, the number of parameters reduces to $P \times K + \tfrac{1}{2} P(P-1)$ in the spatial error model and to $P \times K + P(P - 1) + P$ in the spatial lag model.

Another possibility of dealing with spatial error autocorrelation is to employ groups and a nonparametric covariance estimation technique (such as GMM). The GMM technique avoids estimating the parameters of the covariance matrix (Driscoll and Kraay 1998). However, these parameter reduction techniques and the nonparametric covariance estimation technique (Driscoll and Kraay 1998, n. 5) rule out applications in which the parameters are allowed to vary across all spatial units, which constitutes the initial purpose of the fixed coefficients model.

RANDOM COEFFICIENTS MODEL

The number of parameters to be estimated can also be reduced by treating the coefficients in the regression equation as outcomes of random experiments between spatial units. Consequently, the number of response coefficients no longer grows with the number of spatial units. This approach also improves the efficiency of the estimators due to the availability of substantially more degrees of freedom. Unfortunately, the random effects approach does not reduce the number of parameters of the covariance matrix in the spatial error model or the number of parameters associated with the spatially lagged dependent variables in the spatial lag model. Therefore, a large value of N relative to T remains a problem.

The random coefficient model with *spatial error autocorrelation* can be specified as in equation 4, incorporating the following extension:

$$E(\varepsilon_i \varepsilon_j') = \sigma_{ij} I_T. \tag{30}$$

Note that we change the notation slightly by using σ_{ii} for $i = j$ instead of σ_i^2 as in equation 4a. Similar to the fixed coefficients model, no prior assumptions are required about the nature of the interactions over space. In this model, the random vector $Y \equiv (Y_1', \cdots, Y_N')'$ can be assumed to be distributed with mean $X\beta$, where $X \equiv (X_1', \cdots, X_N')'$, and covariance matrix

$$\Sigma = \begin{bmatrix} X_1 V X_1' + \sigma_{11} I_T & \sigma_{12} I_T & \cdot & \sigma_{1N} I_T \\ \sigma_{21} I_T & X_2 V X_2' + \sigma_{22} I_T & \cdot & \sigma_{2N} I_T \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{N1} I_T & \sigma_{N2} I_T & \cdot & X_N V X_N' + \sigma_{NN} I_T \end{bmatrix} \tag{31}$$

$$= D(I_N \otimes V)D' + (\Sigma_\sigma \otimes I_T),$$

where D is a $(NT \times NK)$ block-diagonal matrix, $D = \text{diag}[X_1, \ldots, X_N]$, and Σ_σ is a $(N \times N)$ matrix with $\Sigma_\sigma = \{\sigma_{ij}\}$. The ML and the GLS estimator of β are known to be equivalent (Lindstrom and Bates 1988) and equal to

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y, \tag{32}$$

although the major problem is that Σ contains unknown parameters $\Sigma = \Sigma(\Sigma_\sigma, V)$ that must also be estimated. There are two ways to proceed. A feasible GLS estimator of β can be constructed on the basis of a consistent estimate of Σ_σ and V. To obtain this estimator, the following steps must be carried out. First, estimate the model assuming that all response parameters are fixed and different for differing spatial units. We use the mnemonic *FC* to refer to these estimates. This model is actually the fixed coefficients model without restrictions on the covariance matrix, as given in equation 3. This step results in estimates for β_i^{FC} and σ_{ij}^{FC}. Second, estimate V by the following (see Swamy 1974):

$$V = \frac{1}{N-1} S - \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{ii}^{FC} \left(X_i' X_i\right)^{-1} + \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{i=1,\, j \neq i}^{N} \hat{\sigma}_{ij}^{FC} \left(X_i' X_i\right)^{-1} X_i' X_j \left(X_j' X_j\right)^{-1}, \tag{33}$$

where $S = \sum_{i=1}^{N} \left(\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i^{FC}\right) \left(\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i^{FC}\right)'$.

The estimator of V, although unbiased, may not be positive definite. To ensure the positive definiteness of the estimated matrix, one can also use the consistent estimator $V = 1/(N-1) \times S$ (for details, see Swamy 1970). Finally, estimate the common mean coefficient vector β by GLS according to equation 32. A distinct problem of the final step is that it requires a matrix inversion of order $(N \times T)$. As an alternative, the inverse of Σ can be computed with the following expression:

$$\Sigma^{-1} = (\Sigma_\sigma^{-1} \otimes I_T) - (\Sigma_\sigma^{-1} \otimes I_T)D[D'(\Sigma_\sigma^{-1} \otimes I_T)D + I_N \otimes V^{-1}]^{-1} D'(\Sigma_\sigma^{-1} \otimes I_T), \quad (34)$$

which requires the inversion of three matrices, one of order K (V), one of order N (Σ_σ), and one of order $(N \times K)$ for the matrix between square brackets. In the case where T is large and/or $K < < T$, this alternative computation is to be preferred, although the inversion of a matrix of order $(N \times K)$ may still create computational difficulties in some of the commercial econometric software packages.

Despite the mathematical equivalence, the feasible GLS estimator of β does not coincide with the ML estimator of β. This is the case because the feasible GLS estimator of β is based on a consistent but not on the ML estimate of Σ_σ and V. The statistical literature shows that ML estimation of β, Σ_σ, and V, although possible, is laborious. There are three reasons for this. First, Σ_σ and V cannot be solved algebraically from the first-order maximizing conditions of the log-likelihood function. Consequently, Σ_σ and V must be solved by numerical methods. Second, a common estimation problem is associated with the restrictions on the parameters of the covariance matrix. A variance estimate should be nonnegative, and a covariance matrix estimate should be nonnegative definite. Moreover, it must be feasible that an estimate takes on values at the boundary of the parameter space. Thus, a variance estimate may be zero, and a covariance matrix estimate may be a nonnegative definite matrix of any rank. In fact, such boundary cases provide useful exploratory information during the model-building process. It is desirable that numerical algorithms for ML estimators can successfully produce the defined estimates for all possible samples, including those where the maximum is attained at the boundary of the parameter space. However, this parameter space problem often causes difficulties with existing ML algorithms (Shin and Amemiya 1997, 190). Third, although some studies assert to have developed efficient and effective algorithms for the likelihood-based estimation of the parameters, they generally assume that $E(\varepsilon_i \varepsilon_i') = \sigma^2 I_T$ and $E(\varepsilon_i \varepsilon_j') = 0$ for $i, j = 1, \ldots, N$ and $j \neq i$ instead of $E(\varepsilon_i \varepsilon_j') = \sigma_{ij}^2 I_T$ (Jenrich and Schluchter 1986; Lindstrom and Bates 1988, 1014, left column; Longford 1993; Goldstein 1995; Shin and Amemiya 1997, 189). This naturally simplifies the parameter space problem, and it is therefore not clear whether these algorithms work for the more general case.

A full random parameter model with *spatial lags* of the dependent variables does in fact not exist. The main reason for this is that the assumption of a random element in the coefficients of lagged dependent variables raises intractable difficulties at the level of identification and estimation (Kelejian 1974; Balestra and Negassi 1992; Hsaio 1996). Instead, a mixed model can be used that contains fixed coefficients for the spatial dependent variables and random coefficients for the exogenous variables. This model reads as

$$Y_{it} = \delta_{1i} Y_{1t} + \cdots + \delta_{ii-1} Y_{i-1t} + \delta_{i+1t} Y_{i+1t} + \cdots + \delta_{Nt} Y_{Nt} + \beta_i' X_{it} + \varepsilon_{it}. \tag{35}$$

A problem that causes this model not to be used very often is the number of observations needed for its estimation. The minimum number of observations on each spatial unit amounts to $(N + K)$, as the number of regressors is $(N - 1 + K)$. Most panels do not meet this requirement, even if N is small. Provided that information is available about the nature of interactions over space, we therefore impose the restrictions $\delta_{ij} = \delta_i w_{ij}$, for $i \neq j$, to attain

$$Y_{it} = \delta_i \sum_{j=1}^{N} w_{ij} Y_{jt} + \beta_i' X_{it} + \varepsilon_{it} \equiv \delta_i Y_i(w) + \beta_i' X_{it} \equiv \eta_i' Z_{it} + \varepsilon_t. \tag{36}$$

In this case, the minimum number of observations needed on each spatial unit reduces to $(K + 1)$, which is independent of N.

Stacking the observations by time for each spatial unit, the full model can be expressed as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} Y_1(w) & 0 & \cdot & 0 \\ 0 & Y_2(w) & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & Y_N(w) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \delta_N \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ X_N \end{bmatrix} \beta + \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_N \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ v_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_N \end{bmatrix}. \tag{37a}$$

The covariance matrix of the composite disturbance term $\text{diag}[X_1, \ldots, X_N] \times v + \varepsilon$ is block diagonal, with the ith diagonal block given by

$$\Phi_i = X_i V X_i' + \sigma_i^2 I_T. \tag{37b}$$

Similar to the spatial error case, there are two ways to proceed. A feasible GLS estimator of δ and β may be constructed, extended to instrumental variables, and based on a consistent estimate of $\sigma_1^2, \cdots, \sigma_N^2$ and V. Alternatively, $\delta, \beta, \sigma_1^2, \cdots, \sigma_N^2$, and V may be estimated by ML.[5] We suggest the following feasible GLS analog instrumental variables estimator taken from Bowden and Turkington (1984, chap. 3).[6]

Let X_i denote the $(T \times K)$ matrix of the exogenous variables in the ith equation, Z_i is the $(T \times (1 + K))$ matrix of the spatially lagged dependent variable and the exogenous variables in the ith equation, and X is the $(T \times K_{ALL})$ matrix of all the explanatory variables in the full model, where K_{ALL} equals $(N(1 + K))$. Consequently, the inversion of the matrix $X'X$ of order $(K_{ALL} \times K_{ALL})$ may constitute a problem when N and/or K are large.

First, estimate the model assuming that all coefficients are fixed. We again use the mnemonic FC to refer to these estimates. The model is in effect the fixed coefficients model extended with spatially lagged dependent variables, as described in equations 24 to 26, but now we stick to the use of instrumental variable estimators. This results in the following estimates for η_i^{FC} and $\sigma_i^{2,FC}$:

$$\hat{\eta}_i^{FC} = \begin{bmatrix} \hat{\delta}_i^{FC} \\ \hat{\beta}_i^{FC} \end{bmatrix} = [Z_i'X(X'X)^{-1}X'Z_i]^{-1} Z_i'X(X'X)^{-1}XY_i, \tag{38a}$$

$$\hat{\sigma}_i^{2,FC} = \frac{(Y_i - Z_i\hat{\eta}_i^{FC})'(Y_i - Z_i\hat{\eta}_i^{FC})}{T-K}. \tag{38b}$$

Second, estimate V by the following (see Balestra and Negassi 1992; Hsiao and Tahmiscioglu 1997):

$$\hat{V} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\beta}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i^{FC}\right)\left(\hat{\beta}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i^{FC}\right)'. \tag{38c}$$

Let Z_i^p denote the predictive values from the multiequation regression of $Z_i = [Y_i(w)\ X_i]$ on X, with the observations for each spatial unit weighted by Φ_i^{-1}:

$$Z_i^p = X\left(X'\Phi_i^{-1}X\right)^{-1}X'\Phi_i^{-1}Z_i = [Y_i^p(w)\ X_i]. \tag{39}$$

The inverse of Φ_i can be computed by the following expression:

$$\Phi_i^{-1} = \frac{1}{\sigma_i^2}I_T - \frac{1}{\sigma_i^2}X_i\left[V^{-1} + \frac{1}{\sigma_i^2}X_i'X_i\right]^{-1}\frac{1}{\sigma_i^2}X_i', \tag{40}$$

as a result of which the formula for Z_i^p changes to

$$\begin{aligned} Z_i^p = X \times &\left[\frac{1}{\sigma_i^2}X'X - \frac{1}{\sigma_i^2}X'X_i\left[V^{-1} + \frac{1}{\sigma_i^2}X_i'X_i\right]^{-1}\frac{1}{\sigma_i^2}X_i'X_i\right]^{-1} \\ &\times\left[\frac{1}{\sigma_i^2}X'Z_i - \frac{1}{\sigma_i^2}X'X_i\left[V^{-1} + \frac{1}{\sigma_i^2}X_i'X_i\right]^{-1}\frac{1}{\sigma_i^2}X_i'Z_i\right]. \end{aligned} \tag{41}$$

Finally, estimate δ_i and β by

$$\begin{bmatrix} \delta_1 \\ \cdot \\ \delta_N \\ \beta \end{bmatrix} = \begin{bmatrix} Y_1^p(w)'\Phi_1^{-1}Y_1(w) & \cdot & 0 & Y_1^p(w)'\Phi_1^{-1}X_1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & Y_N^p(w)'\Phi_N^{-1}Y_N(w) & Y_N^p(w)'\Phi_N^{-1}X_N \\ X_1'\Phi_1^{-1}Y_1(w) & \cdot & X_N'\Phi_N^{-1}Y_N(w) & \sum_{i=1}^{N}X_i'\Phi_i^{-1}X_i \end{bmatrix}\begin{bmatrix} Y_1^p(w)'\Phi_1^{-1}Y_1(w) \\ \cdot \\ Y_N^p(w)'\Phi_N^{-1}Y_N(w) \\ \sum_{i=1}^{N}X_i'\Phi_i^{-1}Y_i(w) \end{bmatrix}, \tag{42}$$

where Φ_i^{-1} can be substituted for the expression given in equation 40.

Although the random coefficients spatial error and spatial lag models have only K response coefficients β and thus $((N-1)\times K)$ less parameters than their fixed counterparts, the problem that N may be too large relative to T remains. This implies that techniques to reduce the number of σ or δ parameters, as already presented for the fixed coefficients model, also apply to the random coefficients model. To test the homogeneity restriction of equal coefficient vectors β, a chi-square test may be used (see Greene 1993). The estimation of the parameters of a random coefficients model is obviously not a simple calculation, but it is feasible. In the manual of LIMDEP (version 6), programming instructions are given to estimate the Swamy random coefficients model that can be straightforwardly extended to spatial error autocorrelation, as well as to a mixed fixed and random coefficients model (Elhorst 1996). A practical problem is that the fixed coefficients, which must be estimated first, cannot be determined when T is smaller than K. In this case, one has to resort to studies asserting to have developed efficient and effective algorithms for the likelihood-based estimation of the parameters (see above).

Similar to the random effects model not necessarily being an appropriate specification when observations on irregular spatial units are used, the random coefficients model may not be either. In that case, the fixed coefficients model is compelling, even when N is large.

CONCLUSIONS

This article has given a systematic overview of panel data models extended to include spatial error autocorrelation or a spatially lagged dependent variable. At the outset, we stated that spatial panel data models are not very well documented in the literature, which may very well be caused by each model having its own specific problems. These problems can be summarized as follows for the four panel data models considered in this article.

Estimation of the spatial *fixed effects model* can be carried out with standard techniques developed by Anselin (1988, 181-82) and Anselin and Hudak (1992), but the regression equation must first be demeaned. This model is relatively simple. One methodological shortcoming is the incidental parameters problem. For short panels, where T is fixed and $N \to \infty$, the coefficients of the spatial fixed effects cannot be estimated consistently. This problem does not necessarily matter when β are the coefficients of interest while the spatial fixed effects are not. Moreover, the problem disappears in panels where N is fixed and $T \to \infty$.

Estimation of the spatial *random effects model* can be carried out by maximum likelihood, although it requires a specific approach. The iterative two-stage procedure needed to maximize the log-likelihood function of the random effects spatial lag model appears to be simpler than the procedure for the random effects spatial error model. The parameters of the random effects spatial error and spatial lag model can be consistently estimated when $N \to \infty$, $T \to \infty$ or $N,T \to \infty$, although the problem remains that the random effects model may not be an appropriate specification when observations on irregularly shaped spatial units are used. In addition, the assumption of zero correlation between the random effects and the explanatory variables is particularly restrictive. Hence, the fixed effects model is compelling, even when N is large and T is small.

A *fixed coefficients* spatial error model with varying coefficients for different spatial units is equivalent to a seemingly unrelated regressions model. Although the estimation of this model is standard, the number of equations allowed in commercial econometric software packages is often limited. A fixed coefficients spatial lag model with different coefficients for different spatial units is almost equivalent to a simultaneous linear equation model. Estimation of this model by maximum likelihood is complicated by the fact that the Jacobian term cannot be expressed in function of the characteristic roots of the spatial weight matrix. As a result, the Jacobi iteration method has to be used, but this method is available only in a limited number of commercial software packages. As an alternative, one can resort to the use of 2SLS, but this method does not take into account the restrictions on the coefficients within the coefficient matrices. A formidable problem of fixed coefficients models is the large number of parameters causing the estimators to be infeasible. Furthermore, even if the estimators are made feasible by introducing restrictions on the parameters, the quality of the asymptotic approximation used to justify the approach remains rather suspect, unless the ratio N/T tends to zero. The latter can eventually be achieved by joining spatial units within groups or by considering separate equations for each group.

Maximum likelihood estimation of the *random coefficients* model extended to spatial error autocorrelation or to spatially lagged dependent variables is possible, although it is laborious. It is simpler to use feasible GLS to estimate the random coefficients model with spatial error autocorrelation and to use feasible GLS in combination with instrumental variables to estimate the random coefficients model comprising spatially lagged dependent variables. These estimators may still be difficult to compute because they require matrix inversions of large orders, depending on the number of spatial units and the number of explanatory variables. In the random coefficients model containing spatially lagged dependent variables, a random element in the coefficients of the spatially lagged dependent variables should be avoided because it creates intractabilities with respect to both identification and estimation. Although the number of parameters in the random coefficients spatial error and spatial lag models are smaller than in their fixed coefficients counterparts, N may still be too large relative to T in typical spatial panel data sets. This may cause the estimators to be infeasible or asymptotically suspect. Finally, the random coefficients model may also not be an appropriate specification when observations on irregular spatial units are used.

Overall, the spatial panel data estimators discussed in this article justify reliance on asymptotics when $T \to \infty$ and N is fixed. Reliance on the asymptotics of $N \to \infty$ and T is fixed and fraught with difficulties for most of the spatial panel data models. The spatial random effects model is a favorable exception in this respect.

NOTES

1. These authors are advocates of geographically weighted regression.

2. The dependent and explanatory variables for every spatial unit are taken in deviations of their average over time. So, for instance, the dependent variable is defined as

$$Y_{it} - \overline{Y}_{i.}, \text{ where } \overline{Y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} Y_{it}.$$

3. The dependent variable reads as

$$Y_{it} - \overline{Y}_{i.} - \overline{Y}_{.t} + \overline{Y}_{..}, \text{ where } \overline{Y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} Y_{it},\ \overline{Y}_{.t} = \frac{1}{N}\sum_{i=1}^{N} Y_{it}, \text{ and } \overline{Y}_{..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}\ ,$$

and similar transformations apply to the X variables.

4. Examples are TSP and PC-Give (see Greene 1997).

5. We found one application of this model in the literature (Sampson, Morenoff, and Earles 1999), but this study does not describe the estimation procedure in detail.

6. Bowden and Turkington (1984) start from the regression equation $Y = X\beta + \mu$, where $E(uu') = \Omega$, and some of the X variables are endogenous. Let Z denote the set of instrumental variables. Then, the generalized least squares analog instrumental variables estimator is $\hat{\beta} = (X^{p'}\,\Omega^{-1}X^{p})^{-1}X^{p'}\,\Omega^{-1}Y$, where $X^{p} = Z(Z'\Omega^{-1}Z)^{-1}Z\Omega^{-1}X$.

REFERENCES

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht, the Netherlands: Kluwer.

———. 2001. Spatial econometrics. In *A companion to theoretical econometrics*, edited by B. H. Baltagi, 310-30. Malden: Blackwell.

Anselin, L., and S. Hudak. 1992. Spatial econometrics in practice: A review of software options. *Regional Science and Urban Economics* 22: 509-36.

Balestra, P., and S. Negassi. 1992. A random coefficient simultaneous equation system with an application to direct foreign investment by French firms. *Empirical Economics* 17: 205-20.

Baltagi, B. H. 2001. *Econometric analysis of panel data*. 2d ed. Chichester, UK: Wiley.

Baltagi, B. H., and J. M. Griffin. 1997. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 77: 303-27.

Baltagi, B. H., J. M. Griffin, and W. Xiong. 2000. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82: 117-26.

Baltagi, B. H., and D. Li. Forthcoming. Prediction in the panel data model with spatial correlation. In *Advances in spatial econometrics: Methodology, tools, and applications*, edited by L. Anselin, R. J. G. M. Florax, and S. J. Rey. Heidelberg: Springer.

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

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

Bowden, R. J., and D. A. Turkington. 1984. *Instrumental variables*. Cambridge, UK: Cambridge University Press.

Breusch, T. S. 1987. Maximum likelihood estimation of random effects models. *Journal of Econometrics* 36: 383-89.

Case, A.C. 1991. Spatial patterns in household demand. *Econometrica* 59: 953-65.

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

Driscoll, J. C., and A. C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80: 549-60.
Elhorst, J. P. 1996. A regional analysis of labour force participation rates across the member states of the European Union. *Regional Studies* 30: 455-65.
Fiebig, D. G. 2001. Seemingly unrelated regression. In *A companion to theoretical econometrics*, edited by B. H. Baltagi, 101-21. Malden: Blackwell.
Fingleton, B. 1999. Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *Journal of Regional Science* 39: 1-19.
Fotheringham, A. S., M. Charlton, and C. Brunsdon. 1997. Measuring spatial variations in relationships with geographically weighted regression. In *Recent developments in spatial analysis*, edited by M. M. Fischer and A. Getis, 60-82. Berlin: Springer.
Froot, K. A. 1989. Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial Analysis* 24: 333-55.
Goldstein, H. 1995. *Multilevel statistical models*. 2d ed. London: Arnold.
Greene, W. H. 1993. *Econometric analysis*. 2d ed. New York: Macmillan.
———. 1997. *Econometric analysis*. 3d ed. New York: Macmillan.
Griffith, D. A. 1988. *Advanced spatial statistics*. Dordrecht, the Netherlands: Kluwer.
Hausman, J. A. 1975. An instrumental variables approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica* 43: 727-38.
———. 1983. Specification and estimation of simultaneous equation models. In *Handbook of econometrics*, vol. 1, edited by Z. Griliches and M. D. Intriligator, 392-448. Amsterdam: Elsevier.
Hsiao, C. 1986. *Analysis of panel data*. Cambridge, UK: Cambridge University Press.
———. 1996. Random coefficients models. In *The econometrics of panel data*, 2d rev. ed., edited by L. Mátyás and P. Sevestre, 77-99. Dordrecht, the Netherlands: Kluwer.
Hsiao, C., and A. K. Tahmiscioglu. 1997. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92: 455-65.
Jenrich, R. I., and M. D. Schluchter. 1986. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42: 805-20.
Kelejian, H. H. 1974. Random parameters in simultaneous equations framework: Identification and estimation. *Econometrica* 42: 517-27.
Kelejian, H. H., and I. R. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509-33.
———. 2002. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics* 32: 691-707.
Lee, L.-F. 2001a. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models: I. Spatial autoregressive processes. Ohio State University.
———. 2001b. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models: II. Mixed regressive, spatial autoregressive processes. Ohio State University.
Lindstrom, M. J., and D. M. Bates. 1988. Newton-Raphson and EM algorithms for linear mixed-effects model for repeated-measures data. *Journal of the American Statistical Association* 83: 1014-22.
Longford, N. T. 1993. *Random coefficient models*. Oxford, UK: Clarendon.
Magnus, J. R. 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics* 19: 239-85.
Mur, J., and F. J. Trívez. 2003. Unit roots and deterministic trends in spatial econometric models. *International Regional Science Review* 26: 289-312.
Murphy, K. J., and R. A. Hofler. 1984. Determinants of geographic unemployment rates: A selectively pooled-simultaneous model. *Review of Economics and Statistics* 66: 216-23.
Nerlove, M., and P. Balestra. 1996. Formulation and estimation of econometric models for panel data. In *The econometrics of panel data*, 2d rev. ed., edited by L. Mátyás, and P. Sevestre, 3-22. Dordrecht, the Netherlands: Kluwer.

Pesaran, M. H., and R. Smith. 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68: 79-113.

Quah, D. T. 1996a. Empirics for economic growth and convergence. *European Economic Review* 40: 1353-75.

———. 1996b. Aggregate and regional disaggregate fluctuations. *Empirical Economics* 21: 137-62.

Sampson, R. J., J. D. Morenoff, and F. Earles. 1999. Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review* 64: 633-60.

Schubert, U. 1982. REMO: An interregional labor market model of Austria. *Environment and Planning A* 14: 1233-49.

Shiba, T., and H. Tsurumi. 1988. Bayesian and non-Bayesian tests of independence in seemingly unrelated regressions. *International Economic Review* 29: 377-95.

Shin, C., and Y. Amemiya. 1997. Algorithms for the likelihood-based estimation of the random coefficient model. *Statistics & Probability Letters* 32: 189-99.

Smirnov, O., and L. Anselin. 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis* 35: 301-19.

Swamy, P. A. V. B. 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38: 311-23.

———. 1974. Linear models with random coefficients. In *Frontiers in econometrics*, edited by P. Zarembka, 143-68. New York: Academic Press.

White, E. N., and G. J. D. Hewings. 1982. Space-time employment modeling: Some results using seemingly unrelated regression estimators. *Journal of Regional Science* 22: 283-302.