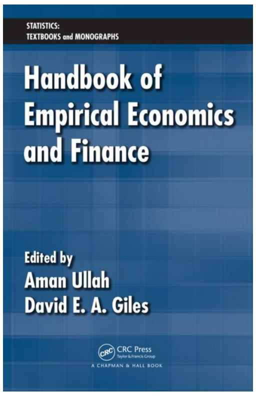

Handbook of Empirical Economics and Finance

Ullah Aman, E. A. Giles David

Spatial Panels

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b10440-16
H. Baltagi Badi
Published online on: 20 Dec 2010

15

Spatial Panels

Badi H. Baltagi

CONTENTS

15.1 Introduction

Economists are interested in spill-over effects and externalities. Spatial models allow simple econometric methods for modeling these spill-over effects. For example, you spend more money on police in one neighborhood, you may increase the crime in an adjacent neighborhood. This externality is dependent on contiguity of the neighborhoods, their common borders, or the distance between these neighborhoods. The same idea can be applied for the analysis of welfare or trade. If California is generous in providing welfare to its residents, this may attract welfare recipients from adjacent states. Gravity models of trade use distance, common border, common language, culture and history, common colonizer, common currency, to see if these things enhance trade. These may be interpreted as distances that are economic, historic, or cultural in nature. In sum, these metrics can be used in a spatial economic model to explain crime or trade or dependency on welfare.

Spatial models deal with correlation across spatial units usually in a cross-section setting; see Anselin (1988, 2001) and Anselin and Bera (1998) for a nice introduction to this literature. Panel data models allow the researcher to control for heterogeneity across these units; see Baltagi (2008a). Spatial panel models can control for *both* heterogeneity and spatial correlation; see for example Baltagi, Song, and Koh (2003) for a joint test of spatial correlation and heterogeneity using panel data. Recent spatial panel data applications in economics include household level survey data from villages observed over time to study nutrition (see Case 1991); per capita expenditures on police to study their effect on reducing crime across counties (see Kelejian and Robinson 1992); the productivity of public capital like roads and highways in the private sector across U.S. states (see Holtz-Eakin 1994); hedonic housing equations using residential sales (see Bell and Bockstael 2000); unemployment clustering with respect to different social and economic metrics (see Conley and Topa 2002); spatial price competition in the wholesale gasoline markets (see Pinkse, Slade, and Brett 2002); and foreign direct investment (see Baltagi, Egger and Pfaffermayr 2007).

Usually one does not worry about cross-section correlation in randomly drawn samples at the individual level. However, when one starts looking at a cross-section of countries, regions, states, counties, etc., these aggregate units are likely to exhibit cross-sectional correlation that have to be dealt with. There is an extensive literature using spatial statistics that deals with this type of correlation. Spatial dependence models may use a metric of economic distance which provides cross-sectional data with a structure similar to that provided by the time index in time series. With the increasing availability of micro as well as macro level panel data, spatial panel data models are becoming increasingly attractive in empirical economic research. The recent literature on spatial panel data models with error components adopts two alternative spatial autoregressive error processes. One specification assumes that only the remainder error term is spatially correlated but the individual effects are not (Anselin 1988; Baltagi, Song, and Koh 2003; Anselin, Le Gallo, and Jayet 2008; we refer to this as the Anselin model). The other specification assumes that both the individual and remainder error components follow the same spatial error process (see Kapoor, Kelejian, and Prucha 2007; we refer to this as the KKP model). Maximum likelihood (ML) estimation, even in its simplest form entails substantial computational problems when the number of cross-sectional units N is large. Kelejian and Prucha (1999) suggested a generalized moments (GM) estimation method which is computationally feasible even when N is large. Kapoor, Kelejian, and Prucha (2007) generalized this GM procedure from cross-section to panel data and derived its large sample properties when T is fixed and $N \to \infty$. Baltagi, Egger, and Pfaffermayr (2008a) introduced a generalized spatial panel data model which nests these two alternative processes in a more general model. They derive LM tests of the generalized model against its restricted alternatives and study their size and power performance against LR tests. In a companion paper, Baltagi, Egger, and Pfaffermayr (2008b) compare the performance of ML estimates of these models under misspecification and suggest a pretest estimator based on the LM tests derived by Baltagi, Egger, and Pfaffermayr (2008a). They show that misspecified MLE can cause substantial loss in MSE where as the pretest estimator performs well, ranking a close second to the true MLE. Monte Carlo experiments are performed to shed some light on the performance of say the Anselin MLE when the true specification is that of KKP, and vice versa. Also, to see how robust is the MLE of the general spatial panel model to *overspecification*, i.e., if the true model is KKP or Anselin. Conversely, how the Anselin and KKP maximum likelihood estimates are affected by *underspecification* of the general model. Since the researcher does not know the true model, the Monte Carlo experiments show that the pretest estimator is a viable second best alternative to the true MLE in practice.

The outline of this chapter is as follows: Section 15.2 introduces the spatial error component regression model and the associated methods of estimation in these models including maximum likelihood and generalized method of moments. Section 15.3 introduces an encompassing spatial error component model and the associated tests for the restricted models. Section 15.4 discusses prediction in the context of spatial panel models, while Section 15.5 studies the performance of various panel unit root tests when spatial correlation across the panel is present. Section 15.6 gives some recent developments in this area and further thoughts for future research.

15.2 Spatial Error Component Regression Model

One can model the spatial correlation as well as the heterogeneity across countries using a spatial error component regression model:

$$y_{ti} = X'_{ti}\beta + u_{ti}, \qquad i = 1, \ldots, N; t = 1, \ldots, T, \tag{15.1}$$

where y_{ti} is the observation on the ith country for the tth time period, X_{ti} denotes the $(k \times 1)$ vector of observations on the nonstochastic regressors and u_{ti} is the regression disturbance. In vector form, the disturbance vector is assumed to have random country effects as well as spatially autocorrelated remainder disturbances, see Anselin (1988):

$$u_t = \mu + \epsilon_t \tag{15.2}$$

with

$$\epsilon_t = \rho W \epsilon_t + \nu_t \tag{15.3}$$

where $\mu' = (\mu_1, \ldots, \mu_N)$ denote the vector of random country effects which are assumed to be IIN$(0, \sigma^2_\mu)$. ρ is the scalar spatial autoregressive coefficient with $|\rho| < 1$. W is a known $(N \times N)$ spatial weight matrix whose diagonal elements are zero.[1] W also satisfies the condition that $(I_N - \rho W)$ is nonsingular.

[1] 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, diagonal elements are null: $w_{ii} = 0$ and the weights are usually standardized such that the elements of each row sum to 1. Alternatively, W could be based on physical distances such as port to port or capital to capital, or commuting distances; see Anselin (1988) for more details on the properties of this W matrix.

$\nu'_t = (\nu_{t1}, \ldots, \nu_{tN})$, where ν_{ti} is assumed to be $\text{IIN}(0, \sigma^2_\nu)$ and also independent of μ_i. One can rewrite ϵ_t as

$$\epsilon_t = (I_N - \rho W)^{-1}\nu_t = B^{-1}\nu_t \tag{15.4}$$

where $B = I_N - \rho W$ and I_N is an identity matrix of dimension N. The model can be rewritten in matrix notation as

$$y = X\beta + u \tag{15.5}$$

where y is now of dimension $(NT \times 1)$, X is $(NT \times k)$, β is $(k \times 1)$ and u is $(NT \times 1)$. X is assumed to be of full column rank and its elements are assumed to be bounded in absolute value. The error can be written in vector form as

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})\nu \tag{15.6}$$

where $\nu' = (\nu'_1, \ldots, \nu'_T)$. Under these assumptions, the variance–covariance matrix for u is given by

$$\Omega = \sigma^2_\mu(J_T \otimes I_N) + \sigma^2_\nu(I_T \otimes (B'B)^{-1}), \text{ and } J_T \text{ is } \alpha(T \times T) \text{ matrix of ones.} \tag{15.7}$$

This matrix can be rewritten as

$$\Omega = \sigma^2_\nu\Big[\bar{J}_T \otimes (T\phi I_N + (B'B)^{-1}) + E_T \otimes (B'B)^{-1}\Big] = \sigma^2_\nu\Sigma \tag{15.8}$$

where $\phi = \sigma^2_\mu/\sigma^2_\nu$, $\bar{J}_T = J_T/T$ and $E_T = I_T - \bar{J}_T$. Using results in Wansbeek and Kapteyn (1983), Σ^{-1} is given by

$$\Sigma^{-1} = \bar{J}_T \otimes (T\phi I_N + (B'B)^{-1})^{-1} + E_T \otimes B'B. \tag{15.9}$$

Also, $|\Sigma| = |T\phi I_N + (B'B)^{-1}| \cdot |(B'B)^{-1}|^{T-1}$. Under the assumption of normality, the log-likelihood function for this model was derived by Anselin (1988, p. 154) as

$$\begin{aligned} L &= -\frac{NT}{2}\ln 2\pi\sigma^2_\nu - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma^2_\nu}u'\Sigma^{-1}u \\ &= -\frac{NT}{2}\ln 2\pi\sigma^2_\nu - \frac{1}{2}\ln[|T\phi I_N + (B'B)^{-1}|] + \frac{(T-1)}{2}\ln|B'B| \\ &\quad -\frac{1}{2\sigma^2_\nu}u'\Sigma^{-1}u \end{aligned} \tag{15.10}$$

with $u = y - X\beta$. For a derivation of the first-order conditions of MLE as well as the LM test for $\rho = 0$ for this model; see Anselin (1988). As an extension to this work, Baltagi, Song, and Koh (2003) derived the joint LM test for spatial error correlation as well as random country effects. Additionally, they derived conditional LM tests, which test for random country effects given the presence of spatial error correlation. Also, spatial error correlation given the presence of random country effects. These conditional LM tests are an alternative to the one directional LM tests that test for random country effects ignoring the presence of spatial error correlation or the one directional LM tests for spatial error correlation ignoring the presence of random country effects. Extensive Monte Carlo experiments are conducted to study the performance of these LM tests as well as the corresponding Likelihood Ratio tests. Baltagi, Song, Jung and Koh (2007) generalize the Baltagi, Song, and Koh (2003) paper by allowing for serial correlation over time for each spatial unit and spatial dependence across these units at a particular point in time. In addition, the model allows for heterogeneity across the spatial units through random effects. Testing for any one of these symptoms ignoring the other two is shown to lead to misleading results. Baltagi, Song, and Kwon (2009) extend these LM statistics to a panel data regression model with heteroskedastic as well as spatially correlated disturbances. A joint LM test for homoskedasticity and no spatial correlation is derived. In addition, a conditional LM test for no spatial correlation given heteroskedasticity, as well as a conditional LM test for homoskedasticity given spatial correlation, are also derived. These LM tests are compared with marginal LM tests that ignore heteroskedasticity in testing for spatial correlation, or spatial correlation in testing for homoskedasticity. Monte Carlo results show that these LM tests as well as their LR counterparts, perform well even for small N and T. However, misleading inference can occur when using marginal rather than joint or conditional LM tests when spatial correlation or heteroskedasticity is present.

Baltagi and Liu (2008) derive a joint LM test which simultaneously tests for the absence of spatial lag dependence and random individual effects in a panel data regression model. This is an extension of the above model to allow for spatial lag dependence in the dependent variable, i.e.,

$$y_t = \lambda W y_t + X_t\beta + u_t, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

where $y_t' = (y_{t1}, \ldots, y_{tN})$ is a vector of observations on the dependent variables for N regions or households at time $t = 1, \ldots, T$. λ is a scalar spatial autoregressive coefficient and W is a known $N \times N$ spatial weight matrix whose diagonal elements are zero. W also satisfies the condition that $(I_N - \lambda W)$ is nonsingular for all $|\lambda| < 1$. X_t is an $N \times k$ matrix of observations on k explanatory variables at time t. $u_t' = (u_{t1}, \ldots, u_{tN})$ is a vector of disturbances following an error component model as described in Equation 15.2. It turns out that this LM statistic is the sum of two standard LM statistics. The first one tests for the absence of spatial lag dependence ignoring the random individual effects, and the second one tests for the absence of random individual effects ignoring the spatial lag dependence. Baltagi and Liu (2008) derive two conditional LM tests. The first one tests for the absence of random individual effects allowing for the possible presence of spatial lag dependence. The second one tests for the absence of spatial lag dependence allowing for the possible presence of random individual effects.

As an alternative to the MLE, generalized method of moments have been proposed for spatial cross-section models by Conley (1999) and Kelejian and Prucha (1999) and an application of the latter method to housing data is given in Bell and Bockstael (2000). Frees (1995) derives a distribution-free test for spatial correlation in panels. This is based on Spearman-rank correlation across pairs of cross-section disturbances. Driscoll and Kraay (1998) show through Monte Carlo simulations that the presence of even modest spatial dependence can impart large bias to OLS standard errors when N is large. They present conditions under which a simple modification of the standard nonparametric time series covariance matrix estimator yields estimates of the standard errors that are robust to general forms of spatial and temporal dependence as $T \to \infty$. However, if T is small, they conclude that the problem of consistent nonparametric covariance matrix estimation is much less tractable. Parametric corrections for spatial correlation are possible only if one places strong restrictions on their form, i.e., knowing W. For typical micropanels with N much larger than T, estimating this correlation is impossible without imposing restrictions, since the number of spatial correlations increases at the rate N^2, while the number of observations grow at rate N. Even for macropanels where $N = 100$ countries observed over $T = 20$ to 30 years, N is still larger than T and prior restrictions on the form of spatial correlation are still needed.

ML estimation, even in its simplest form entails substantial computational problems when the number of cross-sectional units N is large. Kelejian and Prucha (1999) suggested a generalized moments (GM) estimation method which is computationally feasible even when N is large. Kapoor, Kelejian, and Prucha (2007) generalized this GM procedure from cross-section to panel data and derived its large sample properties when T is fixed and $N \to \infty$.

The basic regression model is the same as above; however, the disturbance term u follows the first order spatial autoregressive process

$$u = \rho(I_T \otimes W)u + \epsilon \tag{15.11}$$

with

$$\epsilon = (\iota_T \otimes I_N)\mu + \nu \tag{15.12}$$

where μ, ν and W were defined earlier. This is different from the Anselin (1988) specification described above since it also allows the individual country effects μ to be spatially correlated.

Defining $\bar{u} = (I_T \otimes W)u$, $\bar{\bar{u}} = (I_T \otimes W)\bar{u}$ and $\bar{\epsilon} = (I_T \otimes W)\epsilon$, Kapoor, Kelejian, and Prucha (2007) suggest a GM estimator based on the following six moment conditions

$$
\begin{aligned}
&E[\epsilon' Q\epsilon/N(T-1)] = \sigma_\nu^2 \\
&E[\bar{\epsilon}' Q\bar{\epsilon}/N(T-1)] = \sigma_\nu^2 \,\mathrm{tr}(W'W)/N \\
&E[\bar{\epsilon}' Q\epsilon/N(T-1)] = 0 \\
&E(\epsilon' P\epsilon/N) = T\sigma_\mu^2 + \sigma_\nu^2 = \sigma_1^2 \\
&E(\bar{\epsilon}' P\bar{\epsilon}/N) = \sigma_1^2 \,\mathrm{tr}(W'W)/N \\
&E(\bar{\epsilon}' P\epsilon/N) = 0
\end{aligned}
\tag{15.13}
$$

where, $\epsilon = u - \rho\bar{u}$ and $\bar{\epsilon} = \bar{u} - \rho\bar{\bar{u}}$, substituting these expressions in the six moment conditions we obtain a system of six equations involving the second moments of u, $\bar{u}$ and $\bar{\bar{u}}$. Under the random effects specification considered, the OLS estimator of β is consistent. Using $\widehat{\beta}_{OLS}$ one gets a consistent estimator of the disturbances $\widehat{u} = y - X\widehat{\beta}_{OLS}$. The GM estimator of σ_1^2, σ_ν^2 and ρ is the solution of the sample counterpart of these six equations.

Kapoor, Kelejian, and Prucha (2007) suggest three GM estimators. The first involves only the first three moments which do not involve σ_1^2 and yield estimates of ρ and σ_ν^2. The fourth moment condition is then used to solve for σ_1^2 given estimates of ρ and σ_ν^2. Kapoor, Kelejian, and Prucha (2007) give the conditions needed for the consistency of this estimator as $N \to \infty$. The second GM estimator is based upon weighing the moment equations by the inverse of a properly normalized variance–covariance matrix of the sample moments evaluated at the true parameter values. A simple version of this weighting matrix is derived under normality of the disturbances. The third GM estimator is motivated by computational considerations and replaces a component of the weighting matrix for the second GM estimator by an identity matrix. They perform Monte Carlo experiments comparing MLE and these three GM estimation methods. They find that, on average, the RMSE of ML and their weighted GM estimators are quite similar. However, the first unweighted GM estimator has a RMSE that is 17%–14% larger than that of the weighted GM estimators. For an application of this GM estimator to Foreign Direct Investment (FDI), see Baltagi, Egger, and Pfaffermayr (2007) and to the spatial competition in excise taxation among U.S. states, see Egger, Pfaffermayr, and Winner (2005). Fingleton (2008) extends the GM estimator of Kapoor, Kelejian, and Prucha (2007) to the spatial moving average panel data model. The generalized moments estimator has the advantage that is computationally less demanding than MLE, especially as N gets large.

15.3 A Generalized Spatial Error Component Model

More recently, Baltagi, Egger, and Pfaffermayr (2008a) suggest a generalized spatial panel model which encompasses the Anselin (1988) and Kapoor, Kelejian, and Prucha (2007) models and allows for spatial correlation in the individual and remainder error components that may have different spatial autoregressive parameters. They derive the maximum likelihood estimator (MLE) for this more general spatial panel model when the individual effects are assumed to be random. This in turn allows the researcher to test whether this generalized model reduces to (1) the Anselin model, (2) the Kapoor, Kelejian, and Prucha model, or (3) a simple random effects model that ignores the spatial correlation in the residuals. Baltagi, Egger, and Pfaffermayr (2008a) derive the corresponding LM and LR tests for these three hypotheses and compare their size and power performance using Monte Carlo experiments.

In fact, Baltagi, Egger, and Pfaffermayr (2008a) consider the following generalized spatial error components model:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\beta + \mathbf{u} \\ \mathbf{u} &= \mathbf{Z}_\mu \mathbf{u}_1 + \mathbf{u}_2 \\ \mathbf{u}_1 &= \rho_1 \mathbf{W}_N \mathbf{u}_1 + \mu \\ \mathbf{u}_2 &= \rho_2 \mathbf{W} \mathbf{u}_2 + \nu. \end{aligned} \tag{15.14}$$

This is a balanced panel, which consists of $n = NT$ observations, where N is the number of unique cross-sectional units, while T is the number of time periods. The $(n \times 1)$ vector $\mathbf{y}$ denotes the dependent variable, $\mathbf{X}$ is an $(n \times K)$ matrix of nonstochastic exogenous variables. β is the corresponding $(K \times 1)$ parameter vector. $\mathbf{Z}_\mu = \iota_T \otimes \mathbf{I}_N$ denotes the design matrix for the $(N \times 1)$ vector of random individual effects $\mathbf{u}_1$. ι_T is a $(T \times 1)$ vector of ones and $\mathbf{I}_N$ is an identity matrix of dimension N. The vector of individual effects μ is assumed to be i.i.d. $N(0, \sigma_\mu^2 \mathbf{I}_N)$, while the $(n \times 1)$ vector of remainder disturbances ν is assumed to be i.i.d. $N(0, \sigma_\nu^2 \mathbf{I}_n)$. Furthermore, the elements of μ and ν are assumed to be independent of each other. Both $\mathbf{u}_1$ and $\mathbf{u}_2$ are spatially correlated involving the same spatial weight matrix $\mathbf{W}_N$ for each time period, but with different spatial autocorrelation parameters ρ_1 and ρ_2, respectively. $\mathbf{W}_N$ exhibits zero diagonal elements, the remaining entries are usually assumed to decline with distance. The eigenvalues of $\mathbf{W}_N$ are bounded and smaller than 1 in absolute value. The latter assumption holds for the row normalized $\mathbf{W}_N$. It also holds for the maximum-row normalized spatial weights matrices. This assumption also implies that all row and column sums of $\mathbf{W}_N$ are uniformly bounded in absolute value. In addition, we assume that $|\rho_r| < 1$ for $r = 1, 2$. The data are ordered such that $i = 1, \dots, N$ is the fast index and $t = 1, \dots, T$ is the slow one. The spatial weights matrix for the panel is then given by $\mathbf{W} = \mathbf{I}_T \otimes \mathbf{W}_N$, which is block diagonal and of dimension $(n \times n)$.

This model encompasses both the KKP model, which assumes that $\rho_1 = \rho_2$, and the Anselin model, which maintains that $\rho_1 = 0$. The familiar random effects (RE) panel data model without any spatial correlation is represented by $\rho_1 = \rho_2 = 0$ (see Baltagi 2008a).

In order to derive the $(n \times n)$ variance–covariance of the generalized model, we define $\mathbf{A} = (\mathbf{I}_N - \rho_1 \mathbf{W}_N)$ and $\mathbf{B} = (\mathbf{I}_N - \rho_2 \mathbf{W}_N)$. This allows us to write

$$\mathbf{u}_1 = \mathbf{A}^{-1}\mu \sim N\big(0, \sigma_\mu^2 (\mathbf{A}'\mathbf{A})^{-1}\big) \tag{15.15}$$

$$\mathbf{u}_2 = (\mathbf{I}_T \otimes \mathbf{B}^{-1})\nu \sim N(0, \sigma_\nu^2\big(\mathbf{I}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}\big)). \tag{15.16}$$

and

$$\begin{aligned}\Omega_u &= E(\mathbf{u}\mathbf{u}') = E[(\mathbf{Z}_\mu \mathbf{u}_1 + \mathbf{u}_2)(\mathbf{Z}_\mu \mathbf{u}_1 + \mathbf{u}_2)'] \\ &= \bar{\mathbf{J}}_T \otimes [T\sigma_\mu^2 (\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^2 (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}) = \sigma_\nu^2 \Sigma_u. \end{aligned} \tag{15.17}$$

This uses the fact that $E[\mathbf{u}_1\mathbf{u}_2'] = \mathbf{0}$ since μ and ν are independent by assumption. Note that $\mathbf{Z}_\mu \mathbf{Z}_\mu' = \mathbf{J}_T \otimes \mathbf{I}_N$, where $\mathbf{J}_T$ again denotes a $(T \times T)$ matrix of ones. We define $\mathbf{E}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T$, where $\bar{\mathbf{J}}_T = \mathbf{J}_T / T$ is the averaging matrix over T. The inverse of Ω_u can then be obtained from the inverse of smaller dimension $(N \times N)$ matrices as follows:

$$\Omega_u^{-1} = \bar{\mathbf{J}}_T \otimes [T\sigma_\mu^2 (\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1}]^{-1} + \frac{1}{\sigma_\nu^2}(\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})) \tag{15.18}$$

$$= \frac{1}{\sigma_\nu^2}\left[\bar{\mathbf{J}}_T \otimes [\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1}]^{-1} + (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B}))\right] = \frac{1}{\sigma_\nu^2}\Sigma_u^{-1}$$

Furthermore, $\det[\Omega_u] = \det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] \det[\sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1}]^{T-1}$. Assuming normality of the disturbances the log-likelihood function of the unrestricted model is given by

$$\begin{aligned} L\big(\beta, \sigma_\nu^2, \sigma_\mu^2, \rho_1, \rho_2\big) = &-\frac{NT}{2}\ln 2\pi - \frac{1}{2}\ln\det\big[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}\big] \\ &-\frac{T-1}{2}\ln\det\big(\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}\big) - \frac{1}{2}\mathbf{u}'\Omega_u^{-1}\mathbf{u}, \end{aligned} \tag{15.19}$$

where $\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$. For the special case of $\rho_1 = 0$, this implies that $\mathbf{A} = \mathbf{I}_N$ and the restricted log-likelihood function reduces to the one considered by Anselin (1988, p. 154):

$$\begin{aligned} L_{\mathrm{A}}\big(\beta, \sigma_\nu^2, \sigma_\mu^2, \rho_2\big) = &-\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{1}{2}\ln\det[T\sigma_\mu^2\mathbf{I}_N + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]^{-1} \\ &+\frac{T-1}{2}\ln\det(\mathbf{B}'\mathbf{B}) - \frac{1}{2}\mathbf{u}'\Omega_{u,\mathrm{A}}^{-1}\mathbf{u} \end{aligned} \tag{15.20}$$

$$\Omega_{u,\mathrm{A}}^{-1} = \frac{1}{\sigma_\nu^2}\left[\bar{\mathbf{J}}_T \otimes \left(\frac{T\sigma_\mu^2}{\sigma_\nu^2}\mathbf{I}_N + (\mathbf{B}'\mathbf{B})^{-1}\right)^{-1}\right] + \frac{1}{\sigma_\nu^2}\big[\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})\big].$$

For the alternative case with $\rho_1 = \rho_2 = \rho \neq 0$, $\mathbf{A} = \mathbf{B}$ and we obtain the log-likelihood representation of the KKP estimator:

$$L_{\mathrm{KKP}}(\beta, \sigma_\nu^2, \sigma_\mu^2, \rho) = -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{N}{2}\ln\left(\frac{\sigma_1^2}{\sigma_\nu^2}\right) + \frac{T}{2}\ln\det(\mathbf{B}'\mathbf{B}) - \frac{1}{2}\mathbf{u}'\Omega_{u,\mathrm{KKP}}^{-1}\mathbf{u}$$
$$\Omega_{u,\mathrm{KKP}}^{-1} = \frac{1}{T\sigma_\mu^2 + \sigma_\nu^2}[\bar{\mathbf{J}}_T \otimes (\mathbf{B}'\mathbf{B})] + \frac{1}{\sigma_\nu^2}[\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})]. \qquad (15.21)$$

Finally, with $\rho_1 = \rho_2 = 0$, the log-likelihood reduces to the one representing the familiar RE model without any spatial autocorrelation:

$$L_{\mathrm{RE}}(\beta, \sigma_\nu^2, \sigma_\mu^2) = -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{N}{2}\ln\frac{\sigma_1^2}{\sigma_\nu^2} - \frac{1}{2}\mathbf{u}'\Omega_{u,\mathrm{RE}}^{-1}\mathbf{u}$$
$$\Omega_{u,\mathrm{RE}}^{-1} = \frac{1}{T\sigma_\mu^2 + \sigma_\nu^2}(\bar{\mathbf{J}}_T \otimes \mathbf{I}_N) + \frac{1}{\sigma_\nu^2}(\mathbf{E}_T \otimes \mathbf{I}_N). \qquad (15.22)$$

The pretest estimator is based on a sequence of LM tests derived by Baltagi, Egger, and Pfaffermayr (2008b). Specifically, the following hypotheses were considered:

$$H_0^A : \rho_1 = \rho_2 = 0 \text{ vs. } H_1^A : \text{at least one of the } \rho_1 \text{ or } \rho_2 \neq 0$$
$$H_0^B : \rho_1 = \rho_2 \text{ vs. } H_1^B : \rho_1 \neq \rho_2$$
$$H_0^C : \rho_1 = 0 \text{ vs. } H_1^C : \rho_1 \neq 0 \qquad (15.23)$$

First, we test H_0^A; $\rho_1 = \rho_2 = 0$, to see whether there is no spatial correlation in the error term. If H_0^A is not rejected, the pretest estimator reverts to the random effects MLE. In case H_0^A is rejected, we test H_0^B; $\rho_1 = \rho_2$. If H_0^B is not rejected, the pretest estimator reverts to the KKP MLE. Otherwise, $\rho_1 \neq 0$ or $\rho_2 \neq 0$ and $\rho_1 \neq \rho_2$. Next, we test H_0^C; $\rho_1 = 0$. In case H_0^C is not rejected, the pretest estimator reverts to the Anselin MLE. If H_0^C is rejected, the pretest estimator reverts to the MLE of the general model considered by Baltagi, Egger, and Pfaffermayr (2008b). In other words,

$$\begin{aligned}\widehat{\beta}_{\text{pretest}} &= \widehat{\beta}_{\text{RE,MLE}} \text{ if } H_0^A \text{ is not rejected}\\ &= \widehat{\beta}_{\text{KKP,MLE}} \text{ if } H_0^A \text{ is rejected, and } H_0^B \text{ is not rejected}\\ &= \widehat{\beta}_{\text{Anselin,MLE}} \text{ if } H_0^A \text{ and } H_0^B \text{ are rejected, and } H_0^C \text{ is not rejected}\\ &= \widehat{\beta}_{\text{General,MLE}} \text{ if } H_0^A \text{ and } H_0^B \text{ and } H_0^C \text{ are rejected.}\end{aligned} \qquad (15.24)$$

It has to be emphasized that the pretest estimator becomes the MLE of the general model when all three hypotheses are rejected. Also, it is the MLE of the RE model when H_0^A is not rejected. Hence changing the sequence of tests for H_0^B and H_0^C will not affect the number of times the pretest estimator reverts to the MLE of the RE or General model. This affects only the number of times the pretest estimator reverts to the Anselin or KKP ML estimators. In using the same data set to select the estimator to use based on a series of tests makes the statistical properties of the resulting pretest estimator difficult to derive.[2]

LM tests for these hypotheses were derived by Baltagi, Egger, and Pfaffermayr (2008a) under the assumption of normality. For H_0^A the LM-test statistic is given by

$$LM_A = \frac{1}{2b_A\widetilde{\sigma}_1^4}G_A^2 + \frac{1}{2b_A(T-1)\widetilde{\sigma}_\nu^4}M_A^2, \tag{15.25}$$

where $\widetilde{\sigma}_1^2 = T\widetilde{\sigma}_\mu^2 + \widetilde{\sigma}_\nu^2$, $b_A = tr[(\mathbf{W}_N' + \mathbf{W}_N)^2]$, $G_A = \widetilde{\mathbf{u}}'\{\bar{\mathbf{J}}_T \otimes (\mathbf{W}_N' + \mathbf{W}_N)\}\widetilde{\mathbf{u}}$, and $M_A = \widetilde{\mathbf{u}}'\{\mathbf{E}_T \otimes (\mathbf{W}_N' + \mathbf{W}_N)\}\widetilde{\mathbf{u}}$. Here, $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\beta}_{\text{mle,re}}$ denotes the vector of restricted ML residuals under H_0^A. Baltagi, Egger, and Pfaffermayr (2008a) show that under H_0^A, the LM_A statistic is asymptotically distributed as χ_2^2.

For H_0^B, the LM-test statistic is given by

$$LM_B = \frac{1}{2b_B\overline{\sigma}_1^4}G_B^2 + \frac{1}{2b_B\overline{\sigma}_\nu^4(T-1)}M_B^2, \tag{15.26}$$

with $G_B = \overline{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \mathbf{F})\overline{\mathbf{u}} - \overline{\sigma}_1^2 tr[\mathbf{D}]$, $M_B = \overline{\mathbf{u}}'(\mathbf{E}_T \otimes \mathbf{F})\overline{\mathbf{u}} - \overline{\sigma}_\nu^2(T-1)tr[\mathbf{D}]$, $\mathbf{D} = (\mathbf{W}_N'\overline{\mathbf{A}} + \overline{\mathbf{A}}'\mathbf{W}_N)(\overline{\mathbf{A}}'\overline{\mathbf{A}})^{-1}$ and $\mathbf{F} = \mathbf{W}_N'\overline{\mathbf{A}} + \overline{\mathbf{A}}'\mathbf{W}_N$. Also, $b_B = tr[\mathbf{D}^2] - (tr[\mathbf{D}])^2/N$, $\overline{\sigma}_1^2 = \frac{\overline{\mathbf{u}}'\{\bar{\mathbf{J}}_T \otimes (\overline{\mathbf{A}}'\overline{\mathbf{A}})\}\overline{\mathbf{u}}}{N}$ and $\overline{\sigma}_\nu^2 = \frac{\overline{\mathbf{u}}'\{\mathbf{E}_T \otimes (\overline{\mathbf{A}}'\overline{\mathbf{A}})\}\overline{\mathbf{u}}}{N(T-1)}$. Here, $\overline{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\beta}_{\text{mle,KKP}}$ denotes the vector of restricted ML residuals under H_0^B. The LM_B statistic is asymptotically distributed as χ_1^2 under H_0^B.

Finally, to test H_0^C, we let $\mathbf{C}_1 = [T\widehat{\sigma}_\mu^2\mathbf{I}_N + \widehat{\sigma}_\nu^2(\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1}]^{-1}$, and $\mathbf{C}_2 = (\mathbf{W}_N' + \mathbf{W}_N)$. The corresponding LM test for H_0^C, which has no simple closed form representation is given by

$$LM_C = \widehat{d}_{\rho_1}^2 J_{33}^{-1}, \tag{15.27}$$

where

$$\widehat{d}_{\rho_1} = \left.\frac{\partial L}{\partial \rho_1}\right|_{H_0^B} = -\frac{1}{2}T\widehat{\sigma}_\mu^2 tr[\mathbf{C}_1\mathbf{C}_2] + \frac{1}{2}\widehat{\sigma}_\mu^2\widehat{\mathbf{u}}'\{\mathbf{J}_T \otimes \mathbf{C}_1\mathbf{C}_2\mathbf{C}_1\}\widehat{\mathbf{u}},$$

$\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\beta_{\text{mle, Anselin}}$ denotes the vector of restricted ML residuals under H_0^C, i.e., the Anselin model, and J_{33}^{-1} is the (3,3) element of the inverse of the information matrix described in Baltagi, Egger, and Pfaffermayr (2008a).

Given that the researcher does not know the true model, Baltagi, Egger, and Pfaffermayr (2008b) recommend the pretest estimator which performed well in Monte Carlo experiments no matter what the true underlying model. In fact this pretest estimator was a close second in MSE performance to the true MLE. Additionally, the Monte Carlo experiments shed some light on the performance of the Anselin MLE when the true model is KKP, and vice versa. Ignoring spatial correlation in panel data and performing RE MLE leads to considerable loss in MSE efficiency. When the true model is a general spatial panel model with $\rho_1 \neq \rho_2 \neq 0$, both KKP and Anselin MLE impose

[2] Pretest estimators in econometrics are surveyed in Giles and Giles (1993).

wrong restrictions on the ρ parameters, which in turn, introduce bias and lead to bad MSE performance of the resulting MLEs. Fortunately, this does not translate fully into bad MSE performance for the regression coefficients. The pretest estimator of the regression coefficients always performs better than the misspecified MLE and is recommended in practice.

15.4 Forecasts Using Panel Data with Spatial Error Correlation

The literature on forecasting is rich with time series applications, but this is not the case for spatial panel data applications. Exceptions are Baltagi and Li (2004, 2006) with applications to forecasting sales of cigarette and liquor per capita for U.S. states over time. In order to explain how spatial autocorrelation may arise in the demand for cigarettes, we note that cigarette prices vary among states primarily due to variation in state taxes on cigarettes. Border effect purchases not included in the cigarette demand equation can cause spatial autocorrelation among the disturbances. In forecasting sales of cigarettes, the spatial autocorrelation due to neighboring states and the individual heterogeneity across states is taken explicitly into account. Baltagi and Li (2004) derive the best linear unbiased predictor for the random error component model with spatial correlation using a simple demand equation for cigarettes based on a panel of 46 states over the period 1963–1992. They compare the performance of several predictors of the states demand for cigarettes for 1 year and 5 years ahead. The estimators whose predictions are compared include OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random effects GLS estimator ignoring spatial correlation and random effects estimator accounting for the spatial correlation. Based on the RMSE criteria, the fixed effects and the random effects spatial estimators gave the best out of sample forecast performance.

Best linear unbiased prediction (BLUP) in panel data using an error component model have been surveyed in Baltagi (2008b). However, these panel forecasting applications do not deal with spatial dependence across the panel units. Following Baltagi and Li (2004), Baltagi, Bresson, and Pirotte (2010) compare various forecasts using panel data with spatial error correlation. This is done using a Monte Carlo setup rather than empirical applications. The true data generating process is assumed to be a simple error component regression model with spatial remainder disturbances of the autoregressive or moving average type. The best linear unbiased predictor is compared with other forecasts ignoring spatial correlation, or ignoring heterogeneity due to the individual effects. The paper checks the performance of these forecasts under misspecification of the spatial error process, different spatial weight matrices, and various sample sizes.

Goldberger (1962) has shown that, for a given Ω, the best linear unbiased predictor (BLUP) for the ith individual at a future period $T+\tau$ is given by:

$$\widehat{y}_{i,T+\tau} = X_{i,T+\tau}\widehat{\beta}_{GLS} + \omega'\Omega^{-1}\widehat{u}_{\text{GLS}} \tag{15.28}$$

where $\omega = E[u_{i,T+\tau}u]$ is the covariance between the future disturbance $u_{i,T+\tau}$ and the sample disturbances u. $\widehat{\beta}_{\text{GLS}}$ is the GLS estimator of β based on Ω and $\widehat{u}_{\text{GLS}}$ denotes the corresponding GLS residual vector. For the error component without spatial autocorrelation ($\rho = 0$), this BLUP reduces to

$$\widehat{y}_{i,T+\tau} = X_{i,T+\tau}\widehat{\beta}_{\text{GLS}} + \frac{\sigma_\mu^2}{\sigma_1^2}\left(\iota_T' \otimes l_i'\right)\widehat{u}_{\text{GLS}} \tag{15.29}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$ and l_i is the ith column of I_N. The typical element of the last term of Equation 15.29 is $(T\theta)\overline{u}_{i.,\text{GLS}}$, where $\overline{u}_{i.,\text{GLS}} = \sum_{t=1}^{T}\widehat{u}_{ti,\text{GLS}}/T$ and $\theta = \sigma_\mu^2/\sigma_v^2$; see Baltagi (2008b). Therefore, the BLUP of $y_{i,T+\tau}$ for the RE model modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith individual. In order to make this forecast operational, $\widehat{\beta}_{\text{GLS}}$ is replaced by its feasible GLS estimate and the variance components are replaced by their feasible estimates.

Baltagi and Li (2004, 2006) derived the BLUP correction term when both error components and spatial autocorrelation are present and ϵ_t follows a SAR process. So, the predictor for the SAR is given by:

$$\begin{aligned}\widehat{y}_{i,T+\tau} &= X_{i,T+\tau}\widehat{\beta}_{\text{MLE}} + \theta\left(\iota_T' \otimes l_i'C_1^{-1}\right)\widehat{u}_{\text{MLE}} \\ &= X_{i,T+\tau}\widehat{\beta}_{\text{MLE}} + T\theta\sum_{j=1}^{N} c_{1,j}\overline{u}_{j.,\text{MLE}}\end{aligned} \tag{15.30}$$

where c_{1j} is the jth element of the ith row of C_1^{-1} with $C_1 = [T\theta I_N + (B'B)^{-1}]$ and $\overline{u}_{j.,\text{MLE}} = \sum_{t=1}^{T}\widehat{u}_{tj,\text{MLE}}/T$. In other words, the BLUP of $y_{i,T+\tau}$ adds to $X_{i,T+\tau}\widehat{\beta}_{\text{MLE}}$ a weighted average of the MLE residuals for the N individuals averaged over time. The weights depend upon the spatial matrix W_N and the spatial autoregressive coefficient ρ. To make these predictors operational, we replace θ and ρ by their estimates from the RE-spatial MLE with SAR. When there are no random individual effects, so that $\sigma_\mu^2 = 0$, then $\theta = 0$ and the BLUP prediction terms drop out completely from Equation 15.30. In these cases, Ω reduces to $\sigma_v^2[I_T \otimes (B'B)^{-1}]$ for SAR, and the corresponding MLE for these models yield the pooled spatial MLE with SAR remainder disturbances. This result can be extended to the spatial moving average model (SMA); see Baltagi, Bresson, and Pirotte (2010).

For the Kapoor, Kelejian, and Prucha (2007) model, the BLUP of $y_{i,T+\tau}$ for the SAR-RE also modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith individual. More specifically, the predictor is given by

$$\begin{aligned}\widehat{y}_{i,T+\tau} &= X_{i,T+\tau}\widehat{\beta}_{\text{FGLS}} + \left(\frac{\sigma_\mu^2}{\sigma_1^2}\right) b_i\left(\iota_T' \otimes B_N\right)\widehat{u}_{\text{FGLS}} \\ &= X_{i,T+\tau}\widehat{\beta}_{\text{FGLS}} + \left(\frac{\sigma_\mu^2}{\sigma_1^2})(\iota_T' \otimes l_i'\right)\widehat{u}_{\text{FGLS}}\end{aligned} \tag{15.31}$$

where b_i is the ith row of the matrix B_N^{-1}. This holds because $b_i(\iota_T' \otimes B_N) = (1 \otimes b_i)(\iota_T' \otimes B_N) = (\iota_T' \otimes l_i')$, where l_i' is the ith row of I_N as defined above. $B_N^{-1} B_N = I_N$ and therefore $b_i B_N = l_i'$. This proof applies to both the Kapoor, Kelejian, and Prucha (2007) SAR-RE specification and the Fingleton (2008) SMA-RE specification. Therefore, the BLUP of $y_{i,T+\tau}$ for the SAR-RE and the SMA-RE, like the usual RE model with no spatial effects, modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith individual. While the predictor formula is the same, the MLEs for these specifications yield different estimates which in turn yield different residuals and hence different forecasts.

The results of the Monte Carlo study by Baltagi, Bresson, and Pirotte (2010) find that when the true DGP is RE with a SAR or SMA remainder disturbances, estimators that ignore heterogeneity/spatial correlation perform badly in RMSE forecasts. Accounting for heterogeneity improves the forecast performance by a big margin and accounting for spatial correlation improves the forecast but by a smaller margin. Ignoring both leads to the worst forecasting performance. Heterogeneous estimators based on averaging perform worse than homogeneous estimators in forecasting performance. This performance improves with a larger sample size and seems robust to the type of spatial error structure imposed on the remainder disturbances. These Monte Carlo experiments confirm earlier empirical studies that report similar findings.

15.5 Panel Unit Root Tests and Spatial Dependence

Baltagi, Bresson, and Pirotte (2007) studied the performance of panel unit root tests when spatial effects are present that account for cross-section correlation. Monte Carlo simulations show that there can be considerable size distortions in panel unit root tests when the true specification exhibits spatial error correlation.

Panel data unit root tests have been proposed as alternative more powerful tests than those based on individual time series unit roots tests; see Baltagi (2008a) and Breitung and Pesaran (2008) for some recent reviews of this literature. One of the advantages of panel unit root tests is that their asymptotic distribution is standard normal. This is in contrast to individual time series unit roots which have nonstandard asymptotic distributions. But these tests are not without their critics. The first generation panel unit root tests assumed cross-section independence. These tests include the one proposed by Levin, Lin, and Chu (2002), hereafter denoted by LLC, where the null hypothesis is that each individual time series contains a unit root against the alternative that each time series is stationary. As Maddala (1999) pointed out, the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Im, Pesaran, and Shin (2003), hereafter denoted by IPS, allow for heterogeneous panels and propose panel unit root tests which are based on the average of the individual ADF unit root tests computed from each time series. The null hypothesis is that each individual time series contains a unit root while the alternative allows for some but not all of the individual series to have unit roots. One major criticism of both the LLC and IPS tests is that they require cross-sectional independence. This is a restrictive assumption given the cross-section correlation and spillovers across countries, states, and regions.

Maddala and Wu (1999) and Choi (2001) proposed combining the p-values from the individual unit root ADF tests applied to each time series. Once again, these tests follow a standard normal limiting distribution. They have the advantage that N, the number of cross sections, can be finite or infinite; the time series can be of different length; and the alternative allows some groups to have unit roots while others may not.

Recent studies that try to account for cross-sectional dependence in panel unit root testing include the following: Chang (2002) who explored the nonlinear IV methodology to solve the inferential difficulties in the panel unit root testing which arise from the intrinsic heterogeneities and dependencies of panel models. Chang (2002) suggests an average of individual nonlinear IV t-ratio statistics of the autoregressive coefficient obtained from using an integrable transformation of the lagged level as instrument. These methods assume cross-sectional correlation in the innovation terms driving the autoregressive processes. Choi (2002), on the other hand, generalizes the three unit root tests (inverse chi-square, inverse normal and logit) to the case where the cross-sectional correlation is modeled by error component models. The tests are formulated by combining p-values from the ADF test applied to each individual time series whose stochastic trend components and cross-sectional correlations are eliminated using GLS-demeaning and GLS-detrending. Choi (2002) shows that the combination tests have a standard normal limiting distributions under the sequential asymptotics $T \to \infty$ and $N \to \infty$.

To avoid the restrictive nature of cross-section demeaning procedure, Bai and Ng (2004), and Phillips and Sul (2003), among others, propose dynamic factor models by allowing the common factors to have differential effects on cross-section units. Phillips and Sul's model is a one-factor model where the factor is independently distributed across time. They propose a moment-based method to eliminate the common factor which is different from principal components. More specifically, in the context of a residual one-factor model, Phillips and Sul (2003) provide an orthogonalization procedure which in effect asymptotically eliminates the common factors before preceding to the application of standard unit root tests. Pesaran (2007) suggests a simple way of getting rid of cross-sectional dependence that does not require the estimation of factor loading. His method is based on augmenting the usual ADF regression with the lagged cross-sectional mean and its first-difference to capture the cross-sectional dependence that arises through a single factor model.

Baltagi, Bresson, and Pirotte (2007) run Monte Carlo simulations to compare the empirical size of panel unit root tests with and without spatial error dependence. The structure of the dependence is based on some commonly used spatial error processes: the spatial autoregressive (SAR) and the spatial moving average (SMA) error process and the spatial error components model (SEC). For each experiment, they perform nine panel unit root test statistics: the Levin, Lin, and Chu test (2002), the Breitung (2000) test, the Im, Pesaran, and Shin test (2003), the Maddala and Wu test (1999), the Choi tests (2001, 2002) with and without cross-sectional correlation, the Chang IV test (2002), the Phillips and Sul test (2003), and the Pesaran test (2007). The experiments include a case of no spatial correlation as well as four types of spatial correlation (SAR, SMA, SEC1, and SEC3), with two values of the parameters indicating weak versus strong spatial dependence. They also consider 10 weight matrices, differing in their degree of sparseness, four pairs of (N, T) and two models including individual effects and individual deterministic trends. Even with this modest design, the total number of experiments considered is 1600. They find that ignoring spatial dependence when present can seriously bias the size of panel unit root tests.

15.6 Extensions

Elhorst (2003) considers the ML estimation of a fixed and random effects panel data model extended either to include spatial error autocorrelation or a spatially lagged dependent variable. This is also extended to the case of random coefficients model. In another paper, Elhorst (2005) considers the estimation of a fixed effects dynamic panel data model extended either to include spatial error autocorrelation or a spatially lagged dependent variable. The latter models are first differenced to eliminate the fixed effects and then the unconditional likelihood function is derived taking into account the density function of the first-differenced observations on each spatial unit. Lee and Yu (2010) consider the estimation of a SAR panel model with fixed effects and SAR disturbances. If T is finite but N is large, they show that direct ML estimation of all the parameters including the fixed effects will yield consistent estimators except for the variance of disturbances. Using a transformation that eliminates the individual fixed effects, they provide consistent estimates for all the parameters including the variance of disturbances. The transformation approach is shown to be a conditional likelihood approach if the disturbances are normally distributed. Next, they extend their results to the SAR model with both individual and time-fixed effects. In this case, the transformation approach yields consistent estimators of all the parameters when either N or T are large. For the direct approach, consistency of the variance parameter requires both N and T to be large and consistency of other parameters requires N to be large. Monte Carlo results are provided illustrating the finite sample properties of the various estimators with N and/or T being small or moderately large.

Yu, de Jong, and Lee (2007, 2008) study the asymptotic properties of quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both the number of individuals N and the number of time periods T are large. They cover both the stationary and nonstationary cases. When the roots in the DGP are not all unitary, the estimators' rates of convergence will be the same as the stationary case, and the estimators can be asymptotically normal. In fact, for the distribution of the common parameters, when T is asymptotically large relative to N, the estimators are $\sqrt{NT}$ consistent and asymptotically normal, with the limiting distribution centered around 0. When N is asymptotically proportional to T, the estimators are $\sqrt{NT}$ consistent and asymptotically normal, but the limiting distribution is not centered around 0. When N is large relative to T, the estimators are consistent with rate T, and have a degenerate limiting distribution. Compared to the stationary case, the estimators' rate of convergence will be the same, but the asymptotic variance matrix will be driven by the nonstationary component and it is singular. Consequently, a linear combination of the spatial and dynamic effects can converge at a higher rate. They also propose a bias correction which performs well when T grows faster than $N^{1/3}$.

Pesaran and Tosetti (2008) study large panel data sets where even after conditioning on common observed effects the cross-section units might remain dependently distributed. This could be due to unobserved common factors and/or spatial effects. They introduce the concepts of time-specific weak and strong cross-section dependence and show that the commonly used spatial models are examples of weak cross-section dependence. Pesaran's (2006) common correlated effects (CCE) estimator of panel data model with a multifactor error structure continues to provide consistent estimates of the slope coefficient, even in the presence of spatial error processes.

This chapter highlights some of the recent research in spatial panels. Due to space limitations, several applications and related extensions have not been discussed. Hopefully, this will entice the reader to read more papers on this subject and spur some needed research in this area.

15.7 Acknowledgment

A preliminary version of this chapter was presented as a keynote speech at the 13th African Econometric Society meeting held at the University of Pretoria, South Africa, July 9–11, 2008. Also as the keynote address for the 10th Econometrics and Statistics Symposium held at Ataturk University, Turkey, May 27–29, 2009, and in a session in honor of Cheng Hsiao at the 15th International Conference on Panel Data at the University of Bonn, Germany, July 3–5, 2009. I would like to thank my coauthors Georges Bresson, Alain Pirotte, Dong Li, Seuck Heun Song, Peter Egger, Michael Pfaffermayer, Byoung Cheol Jung, Jae Hyeok Kwon, and Won Koh for allowing me to draw freely on our work.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. 2001. Spatial econometrics. In B. Baltagi, (ed.). *A Companion to Theoretical Econometrics*. pp. 310–330. Oxford, U.K.: Blackwell.

Anselin, L., and A. K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah, D.E.A. Giles (eds). *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.

Anselin, L., J. Le Gallo, and H. Jayet. 2008. Spatial panel econometrics. In L. Mátyás and P. Sevestre (eds.). *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Chap. 19. Berlin: Springer, pp. 625–660.

Bai, J., and S. Ng. 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72:1127–1177.

Baltagi, B. H. 2008a. *Econometric Analysis of Panel Data.* Chichester, U.K.: Wiley.

Baltagi, B. H. 2008b. Forecasting with panel data. *Journal of Forecasting* 27:153–173.

Baltagi, B. H., G. Bresson, and A. Pirotte. 2007. Panel unit root tests and spatial dependence. *Journal of Applied Econometrics* 22:339–360.

Baltagi, B. H., G. Bresson, and A. Pirotte. 2010. Forecasting with spatial panel data. *Computational Statistics and Data Analysis* (forthcoming).

Baltagi, B. H., and D. Li. 2004. Prediction in the panel data model with spatial correlation. In L. Anselin, R. J. G. M. Florax, and S. J. Rey (eds.). *Advances in Spatial Econometrics: Methodology, Tools and Applications* Chap. 13. Berlin: Springer, pp. 283–295.

Baltagi, B. H., and D. Li. 2006. Prediction in the panel data model with spatial correlation: The case of liquor. *Spatial Economic Analysis* 1:175–185.

Baltagi, B. H., and L. Liu. 2008. Testing for random effects and spatial lag dependence in panel data models. *Statistics and Probability Letters* 17:3304–3306.

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

Baltagi, B. H., P. Egger, and M. Pfaffermayr. 2007. Estimating models of complex FDI: Are there third country effects? *Journal of Econometrics* 140:260–281.

Baltagi, B. H., P. Egger, and M. Pfaffermayr. 2008a. A generalized spatial panel data model with random effects. Working paper, Syracuse University, Department of Economics and Center for Policy Research, Syracuse, NY.

Baltagi, B. H., P. Egger, and M. Pfaffermayr. 2008b. A Monte Carlo study for pure and pretest estimators of a panel data model with spatially autocorrelated disturbances, *Annales d'Économie et de Statistique* 87/88:11–38.

Baltagi, B. H., S. H. Song, and J. H. Kwon. 2009. Testing for heteroskedasticity and spatial correlation in a random effects panel data model. *Computational Statistics and Data Analysis* 53:2897–2922.

Baltagi, B. H., S. H. Song, B. C. Jung, and W. Koh. 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140:5–51.

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

Breitung, J. 2000. The local power of some unit root tests for panel data. *Advances in Econometrics* 15:161–177.

Breitung, J., and M. H. Pesaran. 2008. Unit roots and cointegration in panels, In L. Mátyás and P. Sevestre (eds.). *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Chap. 9, Berlin: Springer, pp. 279–322.

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

Chang, Y. 2002. Nonlinear IV unit root tests in panels with cross sectional dependency. *Journal of Econometrics* 110:261–292.

Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20:249–272.

Choi, I. 2002. Instrumental variables estimation of a nearly nonstationary, heterogeneous error component model. *Journal of Econometrics* 109:1–32.

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

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

Driscoll, J. C., and A. C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80:549–560.

Egger, P., M. Pfaffermayr, and H. Winner. 2005. An unbalanced spatial panel data approach to US state tax competition. *Economics Letters* 88:329–335.

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

Elhorst, J. P. 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37:85–106.

Fingleton, B. 2008. A generalized method of moments estimator for a spatial panel model with an endogeneous spatial lag and spatial moving average errors. *Spatial Economic Analysis* 3(1):27–44.

Frees, E. W. 1995. Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69:393–414.

Giles, J. A., and D. E. A. Giles. 1993. Pre-test estimation and testing in econometrics: Recent developments. *Journal of Economic Surveys* 7:145–197.

Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57:369–375.

Holtz-Eakin, D. 1994. Public-sector capital and the productivity puzzle. *Review of Economics and Statistics* 76:12–21.

Im, K. S., M. H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115:53–74.

Kapoor, M., H. H. Kelejian, and I. R. Prucha. 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140:97–130.

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

Kelejian, H. H., and D. P. Robinson. 1992. Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics* 22:317–331.

Lee, L. F., and J. Yu. 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154:165–185.

Levin, A., C. F. Lin, and C. Chu. 2002. Unit root test in panel data: Asymptotic and finite sample properties. *Journal of Econometrics* 108:1–24.

Maddala, G. S. 1999. On the use of panel data methods with cross country data. *Annales D'Économie et de Statistique* 55–56:429–448.

Maddala, G. S., and S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61:631–652.
Pesaran, M. H. 2006. Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica* 74:967–1012.
Pesaran, M. H. 2007. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* 27:265–312.
Pesaran, M. H., and E. Tosetti. 2008. Large panels with common factors and spatial correlations. Working paper, Faculty of Economics, Cambridge University.
Phillips, P. C. B., and D. Sul. 2003. Dynamic Panel Estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6:217–259.
Pinkse, J., M. E. Slade, and C. Brett. 2002. Spatial price competition: A semiparametric approach. *Econometrica* 70:1111–1153.
Wansbeek, T. J., and A. Kapteyn. 1983. A note on spectral decomposition and maximum likelihood estimation of ANOVA models with balanced data. *Statistics and Probability Letters* 1:213–215.
Yu, J., R. de Jong, and L. F. Lee. 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: A nonstationary case. Working paper, Ohio State University, Department of Economics.
Yu, J., R. de Jong, and L. F. Lee. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146:118–134.