CHAPTER 11

Spatial Models and Tests for Cross-Sectional Dependence

11.1 INTRODUCTION – WEAK OR STRONG CROSS-CORRELATIONS

Cross-sectional units could be correlated due to agents taking actions that lead to interdependence among themselves or due to omitted factors that are cross-correlated. For example, the prediction that risk-averse agents will make insurance contracts allowing them to smooth idiosyncratic shocks implies dependence in consumption across individuals or a network. The presence of interactions among cross-sectional units can substantially complicate the model specification, identification, and statistical inference. For instance, an interdependent system raises the issue that an utility maximization agent's decision to establish the links with other agents could lead to multiple equilibria (e.g., Sheng 2020) and model misspecification. Moreover, even in a well-specified model, the inference could be grossly misleading if the error cross-sectional dependence is not properly taken into account.

Consider the $N \times 1$ vector of random variables $\boldsymbol{v}_t = (v_{1t}, \ldots, v_{Nt})'$ at time t with covariance matrix, Σ. If the largest eigenvalue of Σ is of the order $N, O(N)$, then v_{it} is *strongly cross-correlated*. If the largest eigenvalue of Σ is of the order $1, O(1)$, then v_{it} is *weakly cross-correlated*. (e.g., Bai and Silverstein 2004). Or equivalently (Chudik et al. 2011), if

$$\max_{i \in (1, \ldots, N)} \sum_{j=1}^{N} |\sigma_{ij}| = O(N), \tag{11.1.1}$$

v_{it} is strongly cross-correlated. If

$$\max_{i \in (1, \ldots, N)} \sum_{j=1}^{N} |\sigma_{ij}| = k \tag{11.1.2}$$

where k is a fixed constant as $N \to \infty$, v_{it} is weakly cross-correlated. Bailey et al. (2015) and Chudik et al. (2011) suggest summarizing the extent of cross-section dependence based on the behavior of cross-section averages of the variance of v_{it},

$$\bar{v}_t = \sum_{i=1}^{N} w_i v_{it}, \tag{11.1.3}$$

where the weight $\boldsymbol{w} = (w_1, \ldots, w_N)'$ satisfies the "granularity" conditions:

$$||\boldsymbol{w}|| = \sqrt{\boldsymbol{w}'\boldsymbol{w}} = O(N^{-1/2}), \tag{11.1.4}$$

and

$$\frac{w_i}{||\boldsymbol{w}||} = O(N^{-1/2}) \quad \text{uniformly in } i \text{ for } i = 1, \ldots, N. \tag{11.1.5}$$

Then v_{it} is weakly cross-sectionally dependent if

$$\lim_{N\to\infty} \text{Var}(\boldsymbol{w}'\boldsymbol{v}_t) = 0, \tag{11.1.6}$$

and is strongly cross-sectionally dependent if

$$\lim_{N\to\infty} \text{Var}(\boldsymbol{w}'\boldsymbol{v}_t) \geq k > 0. \tag{11.1.7}$$

Suppose

$$v_{it} = \boldsymbol{b}_i'\boldsymbol{f}_t + u_{it}, \tag{11.1.8}$$

where u_{it} is independently identically distributed over i and t with mean 0 and variance σ_u^2, $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are an $r \times 1$ vector of constants such that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}_t\boldsymbol{f}_t' \longrightarrow \Sigma_f$ and $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{b}_i\boldsymbol{b}_i' \longrightarrow \Sigma_b$, where Σ_f and Σ_b are positive definite matrices as assumed in Chapter 10, then $\boldsymbol{v}_t$ is strongly cross-correlated. On the other hand, if each row of the covariance matrix of $\boldsymbol{v}_t$, Σ, has only a maximum of h_N elements that are nonzero and $\frac{h_N}{N} \to 0$ as $N \to \infty$, then $\boldsymbol{v}_t$ is weakly cross-correlated.[1]

The difference between weakly and strongly cross-sectionally dependent data has important implications for panel statistical inference. For instance, consider a model of the form

$$\begin{aligned} y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}, \quad & i = 1, \ldots, N, \\ & t = 1, \ldots, T. \end{aligned} \tag{11.1.9}$$

If v_{it} is weakly cross-correlated, estimating $\boldsymbol{\beta}$ ignoring cross-sectional dependence could still be consistent and asymptotically normally distributed when $N \to \infty$, although they will not be efficient. However, the test statistics based on the formula ignoring cross-correlations could lead to severe size distortion (e.g., Breitung and Das 2008). On the other hand, if $\frac{h_N}{N} \to c \neq 0$ as $N \to \infty$ (i.e., v_{it} is strongly cross-correlated), estimators that ignore the presence of cross-sectional dependence could be inconsistent no matter how large N is (e.g., Hsiao and Tahmiscioglu 2008; Phillips and Sul 2007) if T is finite.[2] The covariance matrix of the pooled least squares estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{LS}$, is equal to

$$\text{Cov}\,(\hat{\boldsymbol{\beta}}_{LS}) = \left[\sum_{t=1}^{T} X_t'X_t\right]^{-1} \left[\sum_{t=1}^{T} X_t'\Sigma X_t\right]\left[\sum_{t=1}^{T} X_t'X_t\right]^{-1}, \tag{11.1.10}$$

where $X_t = (\boldsymbol{x}_{it}')$ denotes the $N \times K$ cross-sectionally stacked explanatory variables $\boldsymbol{x}_{it}$ for time period t. Since $\sum$ is a symmetrical positive definite matrix, Σ can be decomposed as

$$\Sigma = \vee\Lambda\vee', \tag{11.1.11}$$

where $\wedge$ is a diagonal matrix with the diagonal elements being the eigenvalues of $\sum$ and $\vee$ is an orthonormal matrix. If one or more eigenvalues of the Σ is of order N, $X_t'\Sigma X_t$ could

[1] This is equivalent to saying the eigenvalues of $\sum$ are bounded as $N \longrightarrow \infty$ (Pesaran and Tosetti 2010).

[2] This is equivalent to saying the eigenvalues of $\sum$ are $O(N)$, which Pesaran and Tosetti (2010) called the "strong dependence".

be of order N^2 under the conventional assumption that $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}$ converges to a constant vector. Hence, the

$$\text{Cov}\,(\hat{\boldsymbol{\beta}}_{LS}) = O\left(\frac{1}{T}\right). \tag{11.1.12}$$

In other words, the least squares estimator of $\boldsymbol{\beta}$ converges to a random variable rather than a constant when T is finite and N is large.

Contrary to the labeling of times t that has a natural ordering of occurrence of an outcome, cross-sectional labeling, i, is arbitrary. Modeling cross-sectional dependence is a lot more complicated. When N is fixed and T is large, one can ignore the labeling issue and estimate the covariance between i and j, σ_{ij}, by $\frac{1}{T}\sum_{t=1}^{T} v_{it}v_{jt}$ directly. When N is large, the estimation of $\frac{1}{2}N(N+1)\ \sigma_{ij}$ is computationally laborious. We will discuss the spatial approach first to model cross-sectional dependence and then discuss tests for cross-sectional dependence in this chapter. Section 11.2 discusses the basic formulation of spatial weight matrix and spatial coefficients. Section 11.3 covers spatial error models. Section 11.4 focuses on spatial regressive models. Section 11.5 covers some extensions of basic spatial models. Section 11.6 addresses mixed spatial and factor structure process. Section 11.7 concludes the chapter with tests of cross-sectional dependence.

11.2 THE BASIC FORMULATION ON SPATIAL WEIGHT MATRIX AND SPATIAL (DEPENDENCE) COEFFICIENTS

The basic spatial approach assumes that the correlations across cross-sectional units follow a certain spatial ordering; i.e., dependence among cross-sectional units is related to location and distance, in a geographic or more general economic or social network space (e.g., Anselin 1988; Anselin and Griffith 1988; Anselin, Le Gallo, and Jayet 2008). The neighbor relation is expressed by a so-called (known) spatial weights matrix, $W = (w_{ij})$, an N × N positive matrix (i.e., $w_{ij} \geq 0$) in which the rows and columns correspond to the cross-sectional units, that is specified to express the prior relative strength of the interaction between location i (in the row of the matrix) and location j (column), w_{ij}. By convention, the diagonal elements, $w_{ii} = 0$. The weights are often standardized so that the sum of each row, $\sum_{j=1}^{N} w_{ij} = 1$ through row normalization, for instance, let the ith row of W, $\boldsymbol{w}_i' = (d_{i1}, \ldots, d_{iN})/\sum_{j=1}^{N} d_{ij}$, where $d_{ij} \geq 0$ represents a function of the spatial distance of the ith and jth units in some (characteristic) space. A side effect of this standardization is that whereas the original weights may be symmetrical, the row-standardized form is no longer.

The relationship between a cross-correlated vector $\boldsymbol{v}_t$ and the cross-sectional independent shocks $\boldsymbol{u}_t$ is connected by either through a *spatial autoregressive* form,

$$\boldsymbol{v}_t = \theta W \boldsymbol{v}_t + \boldsymbol{u}_t, \tag{11.2.1}$$

or a *spatial moving average* form,

$$\boldsymbol{v}_t = \boldsymbol{u}_t + \delta W \boldsymbol{u}_t, \tag{11.2.2}$$

where $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})'$ is assumed to be independently distributed over i and t with $E\boldsymbol{u}_t\boldsymbol{u}_t' = \sigma_u^2 I_N$. The spatial weight matrix gives the relative strength of interaction between the units at locations i and j. The space dependence parameter, θ or δ, can be considered the scale parameter or a multiplier for the dependence or spillover effects. Because

$$(I_N - \theta W)^{-1} = I_N + \theta W + \theta^2 W^2 + \cdots, \tag{11.2.3}$$

or

$$(I_N + \delta W)^{-1} = I_N - \delta W + \delta^2 W^2 - \cdots, \tag{11.2.4}$$

δ and θ are assumed to have absolute values less than 1 to ensure a "distance" decaying effect among the cross-sectional units.[3]

The *spatial autoregressive* form (11.2.1) implies that the covariance matrix of the N cross-sectional units at time $t, \boldsymbol{v}_t = (v_{1t}, \ldots, v_{Nt})'$ takes the form,

$$E\boldsymbol{v}_t\boldsymbol{v}_t' = \sigma_u^2[I_N - \theta W]^{-1}[I_N - \theta W']^{-1} = V. \tag{11.2.5}$$

The *spatial moving average* form (11.2.2) implies that the covariance matrix of $\boldsymbol{v}_t$ takes the form

$$\begin{aligned} E\boldsymbol{v}_t\boldsymbol{v}_t' &= \sigma_u^2[I_N + \delta W][I_N + \delta W]' \\ &= \sigma_u^2[I_N + \delta(W + W') + \delta^2 WW'] = \tilde{V}. \end{aligned} \tag{11.2.6}$$

When W is *sparse*, i.e., many elements of W are prespecified to be zero, for instance, W could be a block diagonal matrix in which only observations in the same region are considered *neighbors*, and observations across regions are uncorrelated. W can also be a sparse matrix by some neighboring specification, e.g., if a district is a spatial unit, some specifications assume that a neighbor for this district is another one which has a common boundary. The spatial moving average form allows the cross-correlations to be "local" (Equation 11.2.6). On the other hand, the *spatial autoregressive* form suggests a much wider range of spatial covariance than specified by the nonzero elements of the weights matrix W, implying a "global" covariance structure (Equation 11.2.5); however, the interaction between unit i with other units declines as some distance measure between the ith unit and jth unit increases.

Generalizing the spatial approach, Conley (1999) suggests using the notion of "economic distance" to model proximity between two economic agents. The joint distribution of random variables at a set of points is assumed to be invariant to a shift in location and is only a function of the "economic distances" between them. For instance, the population of individuals is assumed to reside in a low-dimensional Euclidean space, say R^2, with each individual i located at a point s_i. The sample then consists of realization of agents' random variables at a collection of locations $\{s_i\}$ inside a sample region. If two agents' locations s_i and s_j are close, then y_{it} and y_{js} may be highly correlated. As the distance between s_i and s_j grows large, y_{it} and y_{js} approach independence. Under this assumption, the dependence among cross-sectional data can be estimated using methods analogous to time series procedures either parametrically or nonparametrically (e.g., Hall, Fisher, and Hoffman 1992; Newey and West 1987; Priestley 1982).

While the approach of defining cross-sectional dependence in terms of "economic distance" measure allows for more complicated dependence than models with time-specific (or group-specific) effects alone (e.g., Chapter 2.6), it still requires that the econometricians have information regarding this "economic distance." In certain urban, environmental, development, growth, and other areas of economics, this information may be available. For instance, in the investigation of peoples' willingness to pay for local public goods, the relevant economic distance may be the time and monetary cost of traveling between points to use these local public goods. Alternatively, if the amenity is air quality, then

[3] The combination of the row sum of W equal to 1 with θ or δ having absolute value less than 1 implies that the cross-sectional dependence of the spatial model is "weak" when N is large.

local weather conditions might constitute the major unobservable common to cross-sectional units in the neighborhood. Other examples include studies of risk sharing in rural developing economies where the primary shocks to individuals in such agrarian economies may be weather related. If so, measures of weather correlation on farms of two individuals could be the proxy for the economic distance between them. In many other situations, prior information like this may be difficult to come by.

Under the assumption that the parameters characterizing a model are independent of the sample size, the combination of θ or δ having absolute value of less than 1 together with the normalization condition $\sum_{j=1}^{N} w_{ij} = 1$ does not constitute innocuous normalization conditions. It implies that cross-sectional units with size sample N are the population under study. Or if N denotes a sample of size N in a population, then for θ (or δ) and w_{ij} to stay constant as N increases, the cross-sectional units have to be weakly dependent in the sense that only a finite number of w_{ij} for $j = 1, \dots, N$ are different from zero as N increases, i.e., only a finite number of units are interacting with a particular unit.

Under the assumption of weak cross-sectional dependence, the covariance estimator of $\boldsymbol{\beta}$ (2.2.10) for models with only individual specific effects or (2.6.13) for models with both individual- and time-specific effects remains consistent if T is fixed and $N \to \infty$, or if N is fixed and T tends to infinity, or both. However, there could be severe size distortion in hypothesis testing for ignoring cross-sectional dependence. The heteroscadasticity autocorrelation consistent covariance estimator proposed by Vogelsang (2012) (Equation 2.8.22) can be used to construct a robust covarian matrix of the least squares or covariance estimator even if the errors follow a spatial process.

11.3 SPATIAL ERROR MODELS

11.3.1 Static Model

Consider a linear regression model of the form,

$$\boldsymbol{y}_t = X_t \boldsymbol{\beta} + \boldsymbol{v}_t, \quad t = 1, \dots, T, \tag{11.3.1}$$

where $\boldsymbol{y}_t = (y_{1t}, \dots, y_{Nt})'$, $X_t = (\boldsymbol{x}_{1t}, \dots, \boldsymbol{x}_{Nt})'$, $\boldsymbol{v}_t = (v_{1t}, \dots, v_{Nt})'$, and $\boldsymbol{v}_t$ take the form (11.2.1) or (11.2.2). Suppose $\boldsymbol{x}_{it}$ is strictly exogenous with regard to $\boldsymbol{u}_t$, i.e., $E(\boldsymbol{u}_t | \boldsymbol{X}_s) = \boldsymbol{0}$. Under the assumption that $\boldsymbol{u}_t$ is independent normal, $N(\boldsymbol{0}, \sigma_u^2 I_N)$, the log-likelihood function of (11.3.1) takes the form

$$-\frac{1}{2} \log \mid \Omega \mid - \frac{1}{2} \boldsymbol{v}' \Omega^{-1} \boldsymbol{v}, \tag{11.3.2}$$

where $\boldsymbol{v} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_T)'$ and

$$\Omega = V \otimes I_T \tag{11.3.3}$$

if $\boldsymbol{v}_t$ is a spatial autoregressive form (11.2.1), and

$$\Omega = \tilde{V} \otimes I_T \tag{11.3.4}$$

if $\boldsymbol{v}_t$ is a spatial moving average form (11.2.2). Conditional on θ or δ, the MLE of $\boldsymbol{\beta}$ is just the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} \boldsymbol{y}), \tag{11.3.5}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \dots, \boldsymbol{y}_T')'$, $X = (X_1', \dots, X_T')'$.

When Ω takes the form of (11.3.3), the log-likelihood function (11.3.2) takes the form

$$\begin{aligned} &T \log \mid I_N - \theta W \mid - \frac{NT}{2} \log \sigma_u^2 \\ &\quad - \frac{1}{2\sigma_u^2}(\boldsymbol{y} - X\boldsymbol{\beta})'[(I_N - \theta W)'(I_N - \theta W) \otimes I_T](\boldsymbol{y} - X\boldsymbol{\beta}). \end{aligned} \tag{11.3.6}$$

The principal difficulty in determining θ is the evaluation of $|I_N - \theta W|$. Ord (1975) notes that if W has eigenvalues $\omega_1, \ldots, \omega_N$, then

$$\mid I_N - \theta W \mid = \prod_{j=1}^{N} (1 - \theta \omega_j), \tag{11.3.7}$$

where ω_j are real even W after row normalization is no longer symmetric. Substituting (11.3.7) into (11.3.6), the log-likelihood values can be evaluated at each possible $(\theta, \boldsymbol{\beta}')$ with an iterative optimization routine. However, when N is large, the computation of the eigenvalues becomes numerically unstable.

When Ω takes the form of (11.3.4), the log-likelihood function takes the form

$$\begin{aligned} &- T \log \mid I_N + \delta W \mid - \frac{NT}{2} \log \sigma_u^2 \\ &\quad - \frac{1}{2\sigma_u^2}(\boldsymbol{y} - X\boldsymbol{\beta})'[(I_N + \delta W)^{-1'}(I_N + \delta W)^{-1} \otimes I_T](\boldsymbol{y} - X\boldsymbol{\beta}). \end{aligned} \tag{11.3.8}$$

Similar to the autoregressive form, the evaluation of (11.3.8) depends on the evaluation of $\mid I_N + \delta W \mid$, which, just like (11.3.7), takes the form

$$\mid I_N + \delta W \mid = \prod_{j=1}^{N} (1 - (-\delta)\omega_j), \tag{11.3.9}$$

where ω_j is the eigenvalue of W.

One can also combine the spatial approach with the error components or fixed-effects specification (e.g., Kapoor, Kelejian, and Prucha 2007; Lee and Yu 2010a, 2010b). For instance, one may generalize the spatial error model by adding the individual-specific effects,

$$\boldsymbol{y} = X\boldsymbol{\beta} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha} + \boldsymbol{v}, \tag{11.3.10}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$. Suppose $\boldsymbol{\alpha}$ are treated as fixed constants and $\boldsymbol{v}$ follows a spatial error autoregressive form (11.2.1); the log-likelihood function is of the form (11.3.2) where Ω is given by (11.3.3) and $\boldsymbol{v} = (\boldsymbol{y} - X\boldsymbol{\beta} - (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha})$. Taking partial derivatives of the log-likelihood function with respect to $\boldsymbol{\alpha}$ and setting it equal to $\mathbf{0}$ yields the MLE estimates of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\beta}$ and θ. Substituting the MLE estimates of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\beta}$ and θ into the log-likelihood function, we obtain the concentrated log-likelihood function

$$\begin{aligned} &- \frac{NT}{2} \log \sigma_u^2 + T \log \mid I_N - \theta W \mid \\ &- \frac{1}{2\sigma_u^2} \tilde{\boldsymbol{v}}'[(I_N - \theta W)'(I_N - \theta W) \otimes I_T]\tilde{\boldsymbol{v}}, \end{aligned} \tag{11.3.11}$$

where the element $\tilde{\boldsymbol{v}}$, $\tilde{v}_{it} = (y_{it} - \bar{y}_i) - (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}$, $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, and $\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$. In other words, the MLE of $\boldsymbol{\beta}$ is equivalent to first taking the covariance transformation of each y_{it} and $\boldsymbol{x}_{it}$ to get rid of the individual-specific effects, α_i, then maximizing (11.3.11) to obtain the MLE of the spatial error model with fixed individual specific effects.

The MLE of $\boldsymbol{\beta}$ and θ are consistent when either N or T or both tend to infinity. However, the MLE of $\boldsymbol{\alpha}$ is consistent only if $T \longrightarrow \infty$. To obtain consistent estimate of $(\boldsymbol{\beta}, \theta, \sigma_u^2)$ with finite T, Lee and Yu (2010a, 2010b) suggest maximizing[4]

$$-\frac{N(T-1)}{2} \log \sigma_u^2 + (T-1) \log |I_N - \theta W| - \frac{1}{2\sigma_u^2} \tilde{\boldsymbol{v}}'[(I_N - \theta W)'(I_N - \theta W) \otimes I_T] \tilde{\boldsymbol{v}}. \tag{11.3.12}$$

When α_i are treated as random and are independent of $\boldsymbol{u}$, the $NT \times NT$ covariance matrix of $\boldsymbol{v}$ takes the form

$$\Omega = \sigma_\alpha^2 (I_N \otimes J_T) + \sigma_u^2 ((B'B)^{-1} \otimes I_T), \tag{11.3.13}$$

if α_i and u_{it} are independent of X and are i.i.d. with mean 0 and variance σ_α^2 and σ_u^2, respectively, and J_T is a $T \times T$ matrix with all elements equal to 1, $B = (I_N - \theta W)$. Using the results in Wansbeek and Kapteyn (1978), one can show that (e.g., Baltagi et al. 2007)

$$\Omega^{-1} = \sigma_u^{-2} \left\{ \frac{1}{T} J_T \otimes \left[T\phi I_N + (B'B)^{-1} \right]^{-1} + E_T \otimes B'B \right\}, \tag{11.3.14}$$

where $E_T = I_T - \frac{1}{T} J$ and $\phi = \frac{\sigma_\alpha^2}{\sigma_u^2}$,

$$| \Omega | = \sigma_u^{2NT} \mid T\phi I_N + (B'B)^{-1} \mid \cdot \mid (B'B)^{-1} \mid^{T-1}. \tag{11.3.15}$$

The MLE of $\boldsymbol{\beta}, \theta, \sigma_u^2$, and σ_α^2 can then be derived by substituting (11.3.14) and (11.3.15) into the log-likelihood function (e.g., Anselin 1988, p. 154).

The feasible generalized least squares estimator of the form (11.3.5) for the random effects spatial error model $\boldsymbol{\beta}$ is to substitute initial consistent estimates of ϕ and θ into (11.3.14). Kapoor et al. (2007) proposed a method of moments estimation with moment conditions in terms of $(\theta, \sigma_u^2, \tilde{\sigma}^2 = \sigma_u^2 + T\sigma_\alpha^2)$.

11.3.2 Dynamic Model

Consider a dynamic panel data model of the form

$$\boldsymbol{y} = \boldsymbol{y}_{-1}\gamma + X\boldsymbol{\beta} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha} + \boldsymbol{v} \tag{11.3.16}$$

where $\boldsymbol{y}_{-1}$ denotes the $NT \times 1$ vector of y_{it} lagged by one period, $\boldsymbol{y}_{-1} = (y_{10}, \ldots, y_{1,T-1}, \ldots, y_{N,T-1})$, X denotes the $NT \times K$ matrix of exogenous variables, $X = (\boldsymbol{x}'_{it})$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$ denotes the $N \times 1$ fixed individual-specific effects. If the error term follows a spatial autoregressive form of (11.2.1), even $| \gamma | < 1$, there could be spatial cointegration if $\gamma + \theta = 1$ (Yu and Lee 2010). Yu et al. (2012) show that the MLE of $(\gamma, \theta, \boldsymbol{\beta}, \boldsymbol{\alpha})$ are $\sqrt{NT}$ consistent with T tends to infinity. However, if $\gamma + \theta = 1$, then the asymptotic covariance matrix of the MLE is singular when the estimator is multiplied by the scale factor $\sqrt{NT}$ because the sum of the spatial and dynamic effects converge at a higher rate (e.g., Yu and Lee 2010).

[4] As a matter of fact, (11.3.12) is derived by the transformation matrix Q^* where $Q^* = \left[F, \frac{1}{\sqrt{T}} I_T\right]$, where F is the $T \times (T-1)$ eigenvector matrix of $Q = I_T - \frac{1}{T}\boldsymbol{e}_T\boldsymbol{e}'_T$ that corresponds to the eigenvalues of 1.

11.4 SPATIAL REGRESSIVE MODELS

When the outcomes of an individual not just are a function of K conditional covariates, but also are due to interactions with other units, a spatial autoregressive model of the form (e.g., Anselin 1988; Ord 1975),

$$\boldsymbol{y} = \rho(W \otimes I_T)\boldsymbol{y} + X\boldsymbol{\beta} + \boldsymbol{u} \tag{11.4.1}$$

is sometimes suggested. On the right-hand side of (11.4.1), $(W \otimes I_T)\boldsymbol{y}$ and $\boldsymbol{u}$ are correlated. Model (11.4.1) can be considered as a special case of the simultaneous-equations model where $\boldsymbol{y}$ is jointly determined given the exogenous factors X and the shock to the system.

When $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2 I_{NT})$, the log-likelihood function is

$$\begin{aligned} &T \log \mid I_N - \rho W \mid - \frac{NT}{2} \log \sigma_u^2 \\ &\quad - \frac{1}{2\sigma_u^2}[\boldsymbol{y} - \rho(W \otimes I_T)\boldsymbol{y} - X\boldsymbol{\beta}]'[\boldsymbol{y} - \rho(W \otimes I_T)\boldsymbol{y} - X\boldsymbol{\beta}], \ \mid \rho \mid < 1. \end{aligned} \tag{11.4.2}$$

When T is fixed, the MLE is $\sqrt{N}$ consistent and asymptotically normally distributed under the assumption that w_{ij} are at most of the order h_N^{-1}, and the ratio $h_N/N \to 0$ as N goes to infinity (Lee 2004). However, when N is large, as with the MLE for (11.3.1), the MLE for (11.4.1) is burdensome and numerically unstable (e.g., Kelejian and Prucha 2001; Lee 2004). The $\mid I_N - \rho W \mid$ is similar in form to (11.3.7). A similar iterative optimization routine as that for (11.3.6) can be evaluated at each possible $(\rho, \boldsymbol{\beta}')$. When N is large, the computation of the eigenvalues becomes numerically unstable.

The parameters $(\rho, \boldsymbol{\beta}')$ can also be estimated by the instrumental variables or generalized method of moments estimator (or two-stage least squares estimator) (Kelejian and Prucha 2001),

$$\begin{pmatrix} \hat{\rho} \\ \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = [Z'H(H'H)^{-1}H'Z]^{-1}[Z'H(H'H)^{-1}H'\boldsymbol{y}], \tag{11.4.3}$$

where $Z = [(W \otimes I_T)\boldsymbol{y}, X]$ and $H = [(W \otimes I_T)X, X]$. Lee (2003) shows that an optimal instrumental variables estimator is to let $H = [(W \otimes I_T)E\boldsymbol{y}, X]$, where $E\boldsymbol{y} = [I_{NT} - \rho(W \otimes I_T)]^{-1}X\boldsymbol{\beta}$. The construction of optimal instrumental variables requires some initial consistent estimators of ρ and $\boldsymbol{\beta}$.

When $w_{ij} = O(N^{-(\frac{1}{2}+\delta)})$, where $\delta > 0, E((W \otimes I_T)\boldsymbol{y}\boldsymbol{u}') = o(N^{-\frac{1}{2}})$, one can ignore the correlations between $(W \otimes I_T)\boldsymbol{y}$ and $\boldsymbol{u}$. Applying the least squares method to (11.4.1) yields a consistent and asymptotically normally distributed estimator of $(\rho, \boldsymbol{\beta}')$ (Lee 2002). However, if W is "sparse," this condition may not be satisfied. For instance, in Case (1991), "neighbors" refers to households in the same district. Each neighbor is given equal weight. Suppose there are r districts and m members in each district, $N = mr$. Then, $w_{ij} = \frac{1}{m-1}$ if i and j are in the same district and $w_{ij} = 0$ if i and j belong to different districts. If $r \longrightarrow \infty$ as $N \longrightarrow \infty$ and N is relatively much larger than r in the sample, one might regard the condition $w_{ij} = O(N^{-(\frac{1}{2}+\delta)})$ being satisfied. On the other hand, if r is relatively much larger than m or $\lim_{N\to\infty} \frac{r}{m} = c \neq 0$, then $w_{ij} = O(N^{-\frac{1}{2}(N+\delta)})$ cannot hold.

For the spatial lag model with individual-specific effects,

$$\boldsymbol{y} = \rho(W \otimes I_T)\boldsymbol{y} + X\boldsymbol{\beta} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha} + \boldsymbol{u}. \tag{11.4.4}$$

If $\boldsymbol{\alpha}$ is treated as fixed constants, the log-likelihood function of (11.4.4) is of similar form as (11.3.11)

$$\begin{aligned} & T \log \mid I_N - \rho W \mid -\frac{NT}{2} \log \sigma_u^2 \\ & -\frac{1}{2\sigma_u^2}\{[\boldsymbol{y} - \rho(W \otimes I_T)\boldsymbol{y} - X\boldsymbol{\beta} - (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha}]' \cdot \\ & \qquad [\boldsymbol{y} - \rho(W \otimes I_T)\boldsymbol{y} - X\boldsymbol{\beta} - (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha}]\}. \end{aligned} \tag{11.4.5}$$

The MLE of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ can be computed similarly as that of (11.3.11).

When α_i are treated as randomly distributed across i with constant variance σ_α^2 and independent of X, then

$$E\left\{(\boldsymbol{u} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha})(\boldsymbol{u} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha})'\right\} = I_N \otimes V^*, \tag{11.4.6}$$

where $V^* = \sigma_u^2 I_T + \sigma_\alpha^2 \boldsymbol{e}_T \boldsymbol{e}_T'$. The MLE or quasi-MLE for the spatial lag model (11.4.1) can be obtained by maximizing

$$\begin{aligned} & T \log \mid I_N - \rho W \mid - \frac{N(T-1)}{2} \log \sigma_u^2 - \frac{N}{2} \log (\sigma_u^2 + T\sigma_\alpha^2) \\ & -\frac{1}{2}(\boldsymbol{y}^* - X\boldsymbol{\beta})'(I_N \otimes V^{*-1})(\boldsymbol{y}^* - X\boldsymbol{\beta}), \end{aligned} \tag{11.4.7}$$

where $\boldsymbol{y}^* = (I_{NT} - \rho(W \otimes I_T))\boldsymbol{y}$. Conditional on ρ, σ_u^2, and σ_α^2, the MLE of $\boldsymbol{\beta}$ is the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = (X'[I_N \otimes V^{*-1}]X)^{-1}(X'(I_N \otimes V^{*-1})(I_{NT} - \rho(W \otimes I_T))\boldsymbol{y}, \tag{11.4.8}$$

where V^{*-1} is given by (2.3.7). Kapoor et al. (2007) have provided moment conditions to obtain initial consistent estimates $\sigma_u^2, \sigma_\alpha^2$, and ρ.

One can also combine the random individual-specific effects specification of $\boldsymbol{\alpha}$ with a spatial specification for the error $\boldsymbol{v}$. For instance, we can let

$$\boldsymbol{y} = \rho(W_1 \otimes I_T)\boldsymbol{y} + X\boldsymbol{\beta} + (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha} + \boldsymbol{v}, \tag{11.4.9}$$

with

$$\boldsymbol{v} = \theta(W_2 \otimes I_T)\boldsymbol{v} + \boldsymbol{u}, \tag{11.4.10}$$

where W_1 and W_2 are $N \times N$ spatial weights matrices and $\boldsymbol{\alpha}$ is an $N \times 1$ vector of individual effects. Let $S(\rho) = I_N - \rho W_1$ and $R(\theta) = I_N - \theta W_2$. Under the assumption that u_{it} is independently normally distributed, the log-likelihood function of (11.4.9) then takes the form

$$\log L = -\frac{NT}{2} \log \sigma_u^2 + T \log \mid S(\rho) \mid + T \log \mid R(\theta) \mid -\frac{1}{2}\tilde{\boldsymbol{v}}^{*\prime}\tilde{\boldsymbol{v}}^*, \tag{11.4.11}$$

where

$$\tilde{\boldsymbol{v}}^* = [R(\theta) \otimes I_T][(S(\rho) \otimes I_T)\boldsymbol{y} - X\boldsymbol{\beta} - (I_N \otimes \boldsymbol{e}_T)\boldsymbol{\alpha}]. \tag{11.4.12}$$

The MLE (or quasi-MLE if u is not normally distributed) can be computed similarly as that of (11.3.11). For detail, see Lee and Yu (2010a, 2010b).

Yu et al. (2012) also consider the estimation of a dynamic spatial lag model with the spatial-time effect,

$$\boldsymbol{y} = (\rho W \otimes I_T)\boldsymbol{y} + \boldsymbol{y}_{-1}\gamma + (\rho^* W \otimes I_T)\boldsymbol{y}_{-1} + X\boldsymbol{\beta} + (\boldsymbol{e}_N \otimes I_T)\boldsymbol{\lambda} + \boldsymbol{v}, \tag{11.4.13}$$

where $\boldsymbol{e}_N$ is an $N \times 1$ vector of $1'$s, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$. Model (11.4.13) is stable if $\gamma + \rho + \rho^* < 1$ and spatially cointegrated if $\gamma + \rho + \rho^* = 1$ but $\gamma \neq 1$. They develop the asymptotics of (quasi-)MLE when both N and T are large and propose a bias correction formula.

11.5 SOME EXTENSIONS

11.5.1 Endogenously Determined Spatial Weight Matrix

So far our discussion has been based on a known fixed spatial weight matrix W. However, Han et al. (2019), Hsieh and Lee (2016), and Qu and Lee (2015) have argued that a spatial autoregressive (SAR) model can also be used to model social networks and the network (spatial) dependence parameter can be interpreted as the strength of peer effects. If economic distance is used to construct the weight matrix, then the elements of the weight matrix are very likely correlated with the final outcome. In other words, the strict exogeneity of the weight matrix W no longer holds.

Consider a spatial autoregressive model of the form

$$\boldsymbol{y}_t = \lambda W \boldsymbol{y}_t + X_{1t}\boldsymbol{\beta} + \boldsymbol{v}_t,\ t = 1, \ldots, T, \tag{11.5.1}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, W is an $N \times N$ spatial weight matrix satisfying the normalization conditions discussed before, X_{1t} denotes the included exogenous variables, and $\boldsymbol{v}_t = (v_{1t}, \ldots, v_{Nt})'$ is the error term. However, if $E(W\boldsymbol{v}_t) \neq 0$, then conditional on W, model (11.5.1) is of the form

$$\boldsymbol{y}_t = \lambda W \boldsymbol{y}_t + X_{1t}\boldsymbol{\beta} + E(\boldsymbol{v}_t|W) + \boldsymbol{v}_t^*. \tag{11.5.2}$$

When only cross-sectional data are available ($T = 1$), Qu and Lee (2015) consider

$$w_{ij} = h_{ij}(Z, c_{ij}) \text{ for } i,\ j = 1, \ldots, N;\ i \neq j, \tag{11.5.3}$$

where $h(\cdot)$ is a bounded function and c_{ij} is some distance measure between unit i and unit j. Qu and Lee (2015) approximate $E(\boldsymbol{v}|W)$ by assuming there exist observable variables, Z and X_2, such that

$$Z = X_2\Gamma + \boldsymbol{\varepsilon}, \tag{11.5.4}$$

where the elements of Z are distinct nonzero elements of w_{ij}. From (11.5.3), $\boldsymbol{\varepsilon}$ and W are correlated, then

$$E(\boldsymbol{v}|W) = (Z - X_2\Gamma)\boldsymbol{\delta} = \boldsymbol{\varepsilon}\boldsymbol{\delta} \tag{11.5.5}$$

where Z is an $N \times P$ matrix of distinct elements of W, X_2 is $N \times R$ matrix of excluded exogenous variables of Equation (11.5.1), Γ is $R \times P$, $\boldsymbol{\delta}$ is a $P \times 1$ vector of constants, and $\boldsymbol{\varepsilon} = (Z - X_2\Gamma)$. Conditional on W and the control variables X_2, (11.5.2) is written as

$$\boldsymbol{y} = \lambda W \boldsymbol{y} + X_1\boldsymbol{\beta} + (Z - X_2\Gamma)\boldsymbol{\delta} + \boldsymbol{v}^*, \tag{11.5.6}$$

where for ease of notations, we drop the subscript t when $t = T = 1$.

Model (11.5.6) can be estimated by first regressing Z on X_2 to obtain a consistent estimate of Γ, then construct the error of $\boldsymbol{\varepsilon}$ as $\hat{\boldsymbol{\varepsilon}} = Z - X_2\hat{\Gamma}$. Substituting $(Z - X_2\hat{\Gamma})$ into (11.5.6), one can obtain the estimator of $(\lambda, \boldsymbol{\beta}, \boldsymbol{\delta})$ by applying the generalized two-stage least squares estimator to the model

$$\boldsymbol{y} = \lambda W \boldsymbol{y} + X_1\boldsymbol{\beta} + \hat{\boldsymbol{\varepsilon}}\boldsymbol{\delta} + \boldsymbol{v}^*. \tag{11.5.7}$$

Qu and Lee (2015) also suggested a quasi-maximum likelihood estimation method that maximizes the joint likelihood function of (11.5.4) and (11.5.6). They show both the two-stage instrumental variable estimator and the QMLE that of λ and $\boldsymbol{\beta}$ are consistent and asymptotically normally distributed.

The panel data have $T > 1$, under the assumption that

$$E(\boldsymbol{v}_t|W) = (Z - X_2\Gamma)\boldsymbol{\delta} = \boldsymbol{a}, \tag{11.5.8}$$

the endogeneity of W can be easily taken care of by considering a model of the form

$$\boldsymbol{y}_t = \lambda W \boldsymbol{y}_t + X_{1t}\boldsymbol{\beta} + \boldsymbol{a} + \boldsymbol{v}_t^*, \quad t = 1, \ldots, T, \tag{11.5.9}$$

where $\boldsymbol{a}$ is an $N \times 1$ vector of constants. One can then estimate $(\lambda, \boldsymbol{\beta}', \boldsymbol{a}')$ by maximizing the quasi-likelihood function of (11.5.9) or by applying the generalized two-stage least squares estimator.

Panel data can also allow a time-varying prespecified spatial weight matrix W_t that depends on $\boldsymbol{y}_t$ by introducing the subscript t to the specification of (11.5.2)–(11.5.4). If $\boldsymbol{v}_t$ is independently distributed over t, then the estimation of $(\lambda, \boldsymbol{\beta}, \Gamma, \boldsymbol{\delta})$ is a straightforward generalization of the estimator suggested by Qu and Lee (2015).

Qu and Lee (2015) assume the edges of the W matrix are continuous to allow a linear approximation of $E(\boldsymbol{v}|W)$,(11.5.4). Alternatively, Han et al. (2019) allow the spatial weight matrix to be time varying and assume the probability density of the elements of the row-normalized $N \times N$ (relative) interactive spatial matrix $W_t = (w_{ij,t})$ taking a logit form[5]

$$f(w_{ij,t}) = \frac{\exp(\Psi_{ij,t})}{1 + \exp(\Psi_{ij,t})}, \tag{11.5.10}$$

where $\Psi_{ij,t}$ are functions of lagged W_{t-1}, $\boldsymbol{y}_{t-1}$, $\boldsymbol{x}_{it}$, and unobserved latent individual and time-specific variables for the panel spatial dynamic model of the form (11.4.13), which is respecified with time varying spatial weight W_t as

$$\boldsymbol{y}_t = \rho W_t \boldsymbol{y}_t + \rho^* W_{t-1}\boldsymbol{y}_{t-1} + \boldsymbol{y}_{t-1}\gamma + X_t\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{v}_t,\ t = 1, \ldots, T. \tag{11.5.11}$$

They then suggest a Bayesian framework for the model (11.5.10)–(11.5.11) and use Markov Chain Monte Carlo (MCMC) sampling steps to obtain the posterior distribution of $(\rho, \rho^*, \gamma, \boldsymbol{\beta})$.

11.5.2 Matrix Exponential Models

The estimation of linear spatial models can be computationally intensive. LeSage and Pace (2006, 2007) make use of the exponential function properties of:

1. $S(\theta^*) = \exp(\theta^* W) = I_N + \theta^* W + \frac{(\theta^*)^2}{2} W^2 + \cdots + \frac{(\theta^*)^\ell}{l!} W^\ell + \cdots,$ (11.5.12)
2. $S(\theta^*)^{-1} = \exp(-\theta^* W),$ (11.5.13)
3. $|\exp(\theta^* W)| = \exp\left(\text{trace}\ (\theta^* W)\right)$ (11.5.14)

[5] Han et al. (2019) actually assume that the N cross-sectional units can be partitioned into G groups. For simplicity, we let $G = 1$ here.

to propose a computationally feasible matrix exponential model (MESS) as a way to take account of cross-sectional dependence through a spatial modeling approach. The matrix exponential models assume that the exponential transformation of $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is linear.

$$S\boldsymbol{y}_t = X_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t,\ t = 1, \ldots, T, \tag{11.5.15}$$

where $X_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$ is an $N \times k$ matrix of strictly exogenous variable, and $\boldsymbol{\varepsilon}_t$ is distributed with $N(\boldsymbol{0}, \sigma^2 I_N)$. The $N \times N$ transformation matrix S is positive definite and takes the form

$$S = e^{\theta^* W} = \sum_{l=0}^{\infty} \left(\frac{1}{l!}\right) (\theta^*)^l W^l, \tag{11.5.16}$$

where $W = (w_{ij})$ is an $N \times N$ nonnegative matrix with $w_{ij} \geq 0$, $w_{ii} = 0$ and $\sum_{j=1}^{N} w_{ij} = 1$, and θ^* is a scalar parameter that measures the strength of spatial dependence. Equation (11.5.16) always converges whatever the value of θ^* is and implies declining weights for observations involving high-order neighboring relationships. Using the norm $\| \| I_N - \theta W \| \| = 1 - \theta$ and $\| \| e^{\theta^* W} \| \| = e^{\theta^*}$ as $N \to \infty$, LeSage and Pace (2006) show that the scale of $\theta^* \approx ln(1-\theta)$ for the linear spatial error model of the form (11.2.1) and (11.3.1).[6]

Since the diagonal elements of W are zero, making use of (11.5.13) and (11.5.14), the Jacobian transformation matrix is equal to 1, and the log-likelihood function for (11.5.15) is proportional to

$$-\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[(S\boldsymbol{y}_t - X_t\boldsymbol{\beta})'(S\boldsymbol{y}_t - X_t\boldsymbol{\beta})]. \tag{11.5.17}$$

Since the powers of W do not rise with the power of l, the power series (11.5.16) converges quickly. Truncating the power series (11.5.16) at order q, then $S\boldsymbol{y}_t$ can be approximated by

$$S\boldsymbol{y}_t \approx Y_t = \left(\sum_{l=0}^{q} \frac{1}{l!}(\theta^*)^l W^l \boldsymbol{y}_t\right). \tag{11.5.18}$$

Substituting (11.5.18) into (11.5.17), the QMLE of θ^* and $\boldsymbol{\beta}$ has a close form solution. Debarsy et al. (2015) find that the QMLE of the MESS model is also robust to the unknown heteroscadasticity in $\boldsymbol{\varepsilon}_t$.

Jin and Lee (2019) generalize the MESS model with prespecified W independent of the exogenous variables X_t to the model of the form

$$S\boldsymbol{y}_t = X_t\boldsymbol{\beta}_1 + W_t\boldsymbol{e}_N\beta_2 + W_tX_t\boldsymbol{\beta}_3 + Z_t\boldsymbol{\beta}_4 + \boldsymbol{v}_t, \quad t = 1, \ldots, T, \tag{11.5.19}$$

where W_t is the $N \times N$ spatial weight matrix at time t, $\boldsymbol{e}_N$ is an $N \times 1$ vector of 1's, and Z_t is an $N \times g$ matrix of endogenous variables. The variables WX_t denote the spatial lags of X_t.

Let $D_t = [X_t, W_t\boldsymbol{e}_N, W_tX_t, Z_t]$, under the assumption that $\boldsymbol{v}_t$ is independently distributed over t with diagonal covariance matrix $E\boldsymbol{v}_t\boldsymbol{v}_t' = \Omega$, the nonlinear two-stage least

[6] Equating the norm of the conventional spatial autoregressive model (11.2.1) and the MESS model, $\theta \approx 1 - \exp(\theta^*)$ (LeSage and Pace 2007).

squares estimator (N2SLS) for the parameter θ, and $\boldsymbol{a} = (\boldsymbol{\beta}_1', \beta_2, \boldsymbol{\beta}_3', \boldsymbol{\beta}_4')'$ with instruments F_t is derived by minimizing the objective function

$$\left[\sum_{t=1}^{T}(S\boldsymbol{y}_t - D_t\boldsymbol{a})'F_t'\right]\left[\sum_{t=1}^{T}F_t\Omega F_t'\right]^{-1}\left[\sum_{t=1}^{T}F_t(S\boldsymbol{y}_t - D_t\boldsymbol{a})\right]. \tag{11.5.20}$$

Provided rank of $F_t \geq$ dimension of $(\theta, \boldsymbol{a}')$, Jin and Lee (2019) show that the N2SLS is consistent and asymptotically normally distributed.

11.6 MIXED SPATIAL AND FACTOR PROCESS

The classical sampling approach assumes that the parameters of a model should be independent of the sample size. However, the spatial approach imposes the condition that the row sum of the spatial weight matrix $W = (w_{ij})$, $\sum_{\substack{i=1 \\ i \neq j}}^{N} w_{ij} = 1$ and the scale parameter θ or δ has absolute value of less than one. When cross-sectional sample size N increases, constant θ (or δ) and w_{ij} can hold only if the cross-sectional units are weakly cross-sectional dependent in the sense that an ith unit can interact only with a finite number of cross-sectional units in the same group, say, g for $y_{it} \in g$th group. This implies that a shock on the ith unit cannot have an aggregate effect. A shock in the ith unit can propagate with only a finite number of other units.

On the other hand, the factor approach discussed in Chapter 10 implies that the cross-sectional units could be strongly cross-correlated. For instance, if $\boldsymbol{b}_i$ are fixed constants and $\boldsymbol{f}_t$ are randomly distributed with

$$v_{it} = \boldsymbol{b}_i'\boldsymbol{f}_t + u_{it}. \tag{11.6.1}$$

The cross-sectional units are correlated if $E(v_{it}v_{jt}) = E(\boldsymbol{b}_i'\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{b}_j) \neq 0$, for any pair of (i, j) units. For instance, when $\boldsymbol{b}_i$ are fixed and $\boldsymbol{f}_t$ are independently identically distributed over t with $E\boldsymbol{f}_t = 0$ and $E\boldsymbol{f}_t\boldsymbol{f}_t' = \Sigma_f$, then

$$Ev_{it}v_{jt} = \boldsymbol{b}_i'\Sigma_f\boldsymbol{b}_j \neq 0. \tag{11.6.2}$$

In other words, the error terms are cross-correlated. Under the assumptions that $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{b}_i\boldsymbol{b}_i' \to \Sigma_b$ and $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t\boldsymbol{f}_t' \to \Sigma_f$ where Σ_b and Σ_f are r-dimensional positive definite matrix, respectively, $\boldsymbol{v}_t$ is a strongly correlated process. In other words, $\boldsymbol{f}_t$ can be considered the dominant shocks that propagate over all cross-sectional units.

There are incidences that shocks in a sector (or group) can only impact the units in the sector or propagate to all units (e.g., Acemoglu et al. 2016; Pesaran and Yang 2020). If we decompose the shocks in a sector with N units as the sum of two components, the component that propagates to all cross-sectional units and the component that only has impacts on a finite number of units, we may consider a mixed spatial error model with a factor model of r common factors (e.g., Pesaran and Yang 2020),

$$\boldsymbol{v}_t = \theta W\boldsymbol{v}_t + \boldsymbol{\eta}_t, \tag{11.6.3}$$

$$\boldsymbol{\eta}_t = \boldsymbol{\alpha} + B\boldsymbol{f}_t + \boldsymbol{u}_t, \tag{11.6.4}$$

where $\boldsymbol{f}_t$ are r-dimensional common factors, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)' = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')'$ is an $N \times 1$ vector of individual-specific effects and, $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)'$ is the $N \times r$ constant factor loading matrix of rank r, and $\frac{1}{N}B'B$ converges to an $r \times r$ nonsingular matrix. Partition $B = (B_1, B_2)$, where B_1 is an $r \times r$ nonsingular matrix, and $\boldsymbol{\eta}_t, \boldsymbol{\alpha}, \boldsymbol{u}_t$ correspondingly,

$$\boldsymbol{\eta}_{1t} = \boldsymbol{\alpha}_1 + B_1 \boldsymbol{f}_t + \boldsymbol{u}_{1t}. \tag{11.6.5}$$

Substituting $\boldsymbol{f}_t = B_1^{-1}(\boldsymbol{\eta}_{1t} - \boldsymbol{\alpha}_1 - \boldsymbol{u}_{1t})$ into $\boldsymbol{\eta}_{2t}$,

$$\boldsymbol{\eta}_{2t} = \boldsymbol{\alpha}_2^* + B_2 B_1^{-1} \boldsymbol{\eta}_{1t} + \boldsymbol{u}_{2t}^*, \tag{11.6.6}$$

where $\boldsymbol{\alpha}_2^* = \boldsymbol{\alpha}_2 - B_2 B_1^{-1} \boldsymbol{\alpha}_1$ and $\boldsymbol{u}_{2t}^* = \boldsymbol{u}_{2t} - B_2 B_1^{-1} \boldsymbol{u}_{1t}$. Then the r-dimensional shocks $\boldsymbol{\eta}_{1t}$ may be viewed as the dominant component shocks in the sector error, $\boldsymbol{v}_{1t}$, that propagate to the other units through $\boldsymbol{\eta}_{2t}$.

Partition $\boldsymbol{v}_t = (\boldsymbol{v}_{1t}, \boldsymbol{v}_{2t})$ and $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ accordingly, where the shock $\boldsymbol{v}_{1t}$ and $\boldsymbol{v}_{2t}$ for model (11.6.3) and (11.6.4) in terms of the dominant shocks $\boldsymbol{\eta}_{1t}$, are,

$$\begin{aligned}
\boldsymbol{v}_{1t} &= C^{-1}\boldsymbol{\eta}_{1t} - \theta C^{-1} W_{12}(I_{N-r} - \theta W_{22})^{-1}\boldsymbol{\eta}_{2t} \\
&= C^{-1}\Big\{[I_r - \theta W_{12}(I_{N-r} - \theta W_{22})^{-1} B_2 B_1^{-1}]\boldsymbol{\eta}_{1t} \\
&\qquad\qquad - \theta W_{12}(I_{N-r} - \theta W_{22})^{-1}(\boldsymbol{\alpha}_2^* + \boldsymbol{u}_{2t}^*)\Big\},
\end{aligned} \tag{11.6.7}$$

$$\begin{aligned}
\boldsymbol{v}_{2t} &= (I_{N-r} - \theta W_{22})^{-1}[\theta W_{21}\boldsymbol{v}_{1t} + \boldsymbol{\eta}_{2t}] \\
&= (I_{N-r} - \theta W_{22})^{-1}\Big\{[B_2 B_1^{-1} + \theta W_{21} C^{-1} \\
&\qquad\qquad (I_r - \theta W_{12}(I_{N-r} - \theta W_{22})^{-1} B_2 B_1^{-1})]\boldsymbol{\eta}_{1t} \\
&\quad + [I_{N-r} - \theta^2 W_{21} C^{-1} W_{12}(I_{N-r} - \theta W_{22})^{-1}](\boldsymbol{\alpha}_2^* + \boldsymbol{u}_{2t}^*)\Big\},
\end{aligned} \tag{11.6.8}$$

where $C = I_r - \theta W_{11} - \theta^2 W_{12}(I_{N-r} - \theta W_{22})^{-1} W_{21}$. The first term on the right-hand side of (11.6.7) and (11.6.8) may be viewed as the impact of dominant shocks $\boldsymbol{\eta}_{1t}$ in the sector that propagate to all units and the second term is the feedback of network effects.

Conditional on $\boldsymbol{f}_t$, under the assumption that $\boldsymbol{u}_t$ is independent normal $N(\boldsymbol{0}, \sigma_u^2 I_N)$, the log-likelihood function of the model (11.3.1), with $\boldsymbol{v}_t$ taking the form (11.6.3) and (11.6.4), becomes

$$\begin{aligned}
&T \log \mid I_N - \theta W \mid - \frac{NT}{2} \log \sigma_u^2 \\
&\quad - \frac{1}{2\sigma_u^2} \sum_{t=1}^{T} [(\boldsymbol{y}_t - X_t\boldsymbol{\beta} - \boldsymbol{\alpha}^* - B^*\boldsymbol{f}_t)'(I_N - \theta W)' \\
&\qquad\qquad (I_N - \theta W)(\boldsymbol{y}_t - X_t\boldsymbol{\beta} - \boldsymbol{\alpha}^* - B^*\boldsymbol{f}_t)],
\end{aligned} \tag{11.6.9}$$

where $\boldsymbol{\alpha}^* = (I_N - \theta W)^{-1}\boldsymbol{\alpha}$, $B^* = (I_N - \theta W)^{-1}B$. Since $\boldsymbol{\alpha}$ and B are unknown constants, we may just take $\boldsymbol{\alpha}^*$ and B^* as unknown fixed constants. The MLE of $(\boldsymbol{\beta}, \theta, \boldsymbol{\alpha}^*, B^*, F)$ can be obtained by maximizing (11.6.9) subject to the normalization conditions $\frac{F'F}{T} = I_r$ and $\frac{1}{N}B^{*\prime}B^*$ diagonal as discussed in Chapter 10, where $F = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$.

The MLE of (11.6.9) under (11.6.3) and (11.6.4) are computationally tedious. Noting that the spatial formulation implies weak cross-sectional dependence, a computationally simpler consistent estimator can be obtained by the following steps:

Step 1: Conditional on $\boldsymbol{\beta}, F$ can be obtained as $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalue of $T \times T$ matrix

$$\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{v}_i^* \boldsymbol{v}_i^{*'}, \tag{11.6.10}$$

where $\boldsymbol{v}_i^* = (\boldsymbol{y}_i - \boldsymbol{e}_T \bar{y}_i) - (X_i - \boldsymbol{e}_T \bar{\boldsymbol{x}}_i')\boldsymbol{\beta} = (v_{i1}^*, \ldots, v_{iT}^*)'$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$, $\boldsymbol{e}_T = (1, \ldots, 1)'$, $\bar{y}_i = \frac{1}{T}\boldsymbol{e}_T'\boldsymbol{y}_i$, and $\bar{\boldsymbol{x}}_i = \frac{1}{T}\boldsymbol{e}_T' X_i$. Compute B^* as

$$B^* = \frac{1}{T} F'(\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_N^*). \tag{11.6.11}$$

Step 2: Estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}^*$ by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i' M_F X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' M_F \boldsymbol{y}_i\right) \tag{11.6.12}$$

and

$$\boldsymbol{\alpha}^{*'} = \frac{1}{T}\, \boldsymbol{e}_T'\, (\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_N^*). \tag{11.6.13}$$

Step 3: Substitute $(\hat{\boldsymbol{\beta}}, \hat{B}^*, \hat{F}, \hat{\boldsymbol{\alpha}}^*)$ into the log-likelihood function (11.6.9) to estimate θ.

Step 4: Repeat steps 1–3 until the solution converges.

If the focus is only on estimating $\boldsymbol{\beta}$ and supposing the data generating process of $\boldsymbol{x}_{it}$ follows

$$\boldsymbol{x}_{it} = \Gamma_i' \boldsymbol{f}_t + \boldsymbol{v}_{it} \tag{11.6.14}$$

where $\boldsymbol{v}_{it}$ follows a linear stationary process with absolute summable autocovariance and

$$\frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \boldsymbol{\beta}_i' \Gamma_i + \boldsymbol{b}_i' \\ \Gamma_i \end{pmatrix} \longrightarrow \bar{C}, \tag{11.6.15}$$

where rank $(\bar{C}) = r$, one could apply Pesaran's (2006) CCE estimator (also see Chapter 10). Alternatively, Pesaran and Tosetti (2011) suggest first to use $(\bar{y}_t, \bar{\boldsymbol{x}}_t)$ to sweep out the impact of dominating factors $\boldsymbol{f}_t$ on $\boldsymbol{y}_i$, then to estimate $\boldsymbol{\beta}$ by the mean group estimator

$$\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{\beta}}_{CCE,i}, \tag{11.6.16}$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{CCE,i} &= (X_i' M_H X_i)^{-1}(X_1' M_H \boldsymbol{y}_i), \\ M_H &= I_T - H(H'H)^{-1}H, \\ H &= (\bar{Y}, \bar{X}), \\ \bar{Y} &= (\bar{y}_1, \ldots, \bar{y}_T)', \bar{X} = (\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_T)', \\ \bar{y}_t &= \frac{1}{N}\sum_{i=1}^{N} y_{it}, \bar{\boldsymbol{x}}_t = \frac{1}{N}\sum_{i=1}^{N} x_{it}. \end{aligned} \tag{11.6.17}$$

They show that as both N and $T \rightarrow \infty$, either approach yields a $\sqrt{NT}$ consistent estimator of $\boldsymbol{\beta}$.

11.7 CROSS-SECTIONAL DEPENDENCE TESTS

Many of the conventional panel data estimators that ignore cross-sectional dependence are inconsistent even when $N \to \infty$ if T is finite. Modeling cross-sectional dependence is a lot more complicated than modeling time-series dependence. So is the estimation of panel data models in the presence of cross-sectional dependence. Therefore, it could be prudent first to test cross-sectional independence and embark only on estimating models with cross-sectional dependence if the tests reject the null hypothesis of no cross-sectional dependence.

11.7.1 Linear Model

Consider a linear model,

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + v_{it}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{11.7.1}$$

where v_{it} are stationary time series.

11.7.1.1 Lagrangian Multiplier Test

Breusch and Pagan (1980) derived a Lagrangian multiplier test statistic for cross-sectional dependence:

$$LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}^2, \tag{11.7.2}$$

where $\hat{\rho}_{ij}$ is the estimated sample cross-correlation coefficient between the least squares residuals $\hat{v}_{it}$ and $\hat{v}_{jt}$, where $\hat{v}_{it} = y_{it} - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}_i$, and $\hat{\boldsymbol{\beta}}_i = (X'_iX_i)^{-1}X_i\boldsymbol{y}_i$. If v_{it} is stationary over t, when N is fixed and $T \to \infty$, (11.7.2) converges to a chi-square distribution with $\frac{N(N-1)}{2}$ degrees of freedom under the null of no cross-sectional dependence. When N is large, the scaled LM statistic (SLM),

$$SLM = \sqrt{\frac{2}{N(N-1)}}LM, \tag{11.7.3}$$

is asymptotically normally distributed with mean zero and variance 1 as $T \to \infty$.

Many panel data sets have N much larger than T. Because $E(T\hat{\rho}_{ij}^2) \neq 0$ for all T, SLM is not properly centered. In other words, when $N > T$, the SLM tends to overeject, often substantially.

To correct for the bias in large N and T panels, Pesaran, Ullah, and Yamagata (2008) propose a bias-adjusted LM test,

$$LM_B = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{(T-k)\hat{\rho}_{ij}^2 - \mu_{ij}}{\nu_{ij}}, \tag{11.7.4}$$

where k is the dimension of $\boldsymbol{x}_{it}$,

$$\mu_{ij} = E[(T-k)\hat{\rho}_{ij}^2] = \frac{1}{T-k}Tr[E(M_iM_j)] \text{ and}$$
$$\nu_{ij}^2 = \text{Var}[(T-k)\hat{\rho}_{ij}^2] = \{Tr[EM_iM_j]\}^2a_1 + 2TrE[(M_iM_j)^2]a_2,$$

$$M_i = I_T - X_i(X_i'X_i)^{-1}X_i',\ a_1 = a_2 - \frac{1}{(T-k)^2},$$

$$a_2 = 3\left[\frac{(T-k-8)(T-k-2)+24}{(T-k+2)(T-k-2)(T-k-4)}\right]^2.$$

Pesaran, Ullah, and Yamagata (2008) show that (11.7.4) is asymptotically normally distributed with mean 0 and variance 1 for all $T > k + 8$.

11.7.1.2 CD Test

Since the adjustment of the finite sample bias of the LM test is complicated, Pesaran (2020) suggests a CD test statistic:

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}\right). \tag{11.7.5}$$

When both N and $T \to \infty$, the CD test converges to a normal distribution with mean 0 and variance 1 under the null of no cross-sectional dependence conditional on $\boldsymbol{x}$. The Monte Carlo conducted in Pesaran (2020) shows that the estimated size is very close to the nominal level for any combinations of N and T considered. However, the CD test has power only if $\frac{1}{N}\sum_{i=1}^{N}\rho_{ij} \neq 0$. On the other hand, the LM test has power even when the average of the correlation coefficient is equal to zero as long as some pairs, $\hat{\rho}_{ij} \neq 0$.

11.7.1.3 Spatial Dependence Coefficient Zero Test

The spatial approach assumes a known correlation pattern among cross-sectional units, W. Under the null of no cross-sectional dependence, $\theta = 0$ for any prespecified W. Therefore, a test for spatial effects is a test of the null hypothesis $H_0: \theta = 0$ (or $\delta = 0$). Burridge (1980) derives the Lagrange Multiplier test statistic for model (11.2.1) or (11.2.2),

$$\tau = \frac{[\hat{\boldsymbol{v}}'(W \otimes I_T)\hat{\boldsymbol{v}}/(\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}/NT)]^2}{tr[(W^2 \otimes I_T) + (W'W \otimes I_T)]} \tag{11.7.6}$$

which is chi-square distributed with one degree of freedom, where $\hat{\boldsymbol{v}} = \boldsymbol{y} - X\boldsymbol{\beta}$.

For error component spatial autoregressive model (11.3.10), Anselin (1988) derived the LM test statistic for $H_0: \theta = 0$,

$$\tau^* = \frac{\left[\frac{1}{\hat{\sigma}_u^2}\hat{\boldsymbol{v}}^{*'}(W \otimes (I_T + \hat{k}(T\hat{k}-2)\boldsymbol{e}_T\boldsymbol{e}_T'))\hat{\boldsymbol{v}}^*\right]}{P}, \tag{11.7.7}$$

which is asymptotically χ^2 distributed with one degree of freedom, where $\boldsymbol{v}^* = \boldsymbol{y} - X\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i'V^{*-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'V^{*-1}\boldsymbol{y}_i\right)$, the usual error component estimator where V^* takes the form of the one in (2.3.7), $\hat{k} = \hat{\sigma}_\alpha^2\left[\hat{\sigma}_u^2\left(1 + T\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_u^2}\right)\right]^{-1}$, and $P = (T^2\hat{k}^2 - 2\hat{k} + T)(trW^2 + trW'W)$. Baltagi et al. (2007) considered various combination of error components and the spatial parameter test. Kelejian and Prucha (2001), and Pinkse (2000) suggested tests of cross-sectional dependence based on the spatial correlation analogue of the Durbin–Watson/Box–Pierce tests for time series correlations.

11.7.2 Linear Dynamic Models

If v_{it} is independently distributed over t, the LM or CD test can be applied to the model of the form,

$$y_{it} = \gamma y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + v_{it}. \tag{11.7.8}$$

Sarafidis, Yamagata, and Robertson (SYR) (2009) proposed a Sargan's (1958) difference test based on the GMM estimator of (11.7.8). As shown in Chapter 3, $\boldsymbol{\theta}' = (\gamma, \boldsymbol{\beta}')$ can be estimated by the GMM method (3.3.17). SYR suggested splitting W_i into two separate sets of instruments,

$$W_{1i}' = \begin{bmatrix} y_{i0} & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & y_{i0} & y_{i1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & y_{i0} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & y_{i0} & \cdot & \cdot & y_{i,T-2} \end{bmatrix}, \tag{11.7.9}$$

and

$$W_{2i}' = \begin{bmatrix} \boldsymbol{x}_i' & \boldsymbol{0}' & \boldsymbol{0}' & \cdot & \cdot \\ \boldsymbol{0}' & \boldsymbol{x}_i' & \boldsymbol{0}' & \cdot\cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \boldsymbol{x}_i' \end{bmatrix}, \tag{11.7.10}$$

where $\boldsymbol{x}_i' = (\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT}')$, W_{1i}' is $(T-1) \times T(T-1)/2$, W_{2i}' is $(T-1) \times KT(T-1)$, and $\boldsymbol{x}_{it}$ is strictly exogenous.[7]

Under the null of no cross-sectional dependence, both sets of moment conditions

$$E[W_{1i}\Delta\boldsymbol{u}_i] = \boldsymbol{0}, \tag{11.7.11}$$

and

$$E[W_{2i}\Delta\boldsymbol{u}_i] = \boldsymbol{0}, \tag{11.7.12}$$

hold. However, if there exists cross-sectional dependence, (11.7.11) may not hold. For instance, suppose u_{it} can be decomposed into the sum of two components, the impact of r time-varying common omitted factors and an idiosyncratic component, ϵ_{it},

$$u_{it} = \boldsymbol{b}_i'\boldsymbol{f}_t + \epsilon_{it}. \tag{11.7.13}$$

For simplicity, we assume ϵ_{it} is independently distributed over i and t. Then the first difference of u_{it},

$$\Delta u_{it} = \boldsymbol{b}_i'\Delta\boldsymbol{f}_t + \Delta\epsilon_{it}, \tag{11.7.14}$$

[7] If $\boldsymbol{x}_{it}$ is predetermined rather than strictly exogenous, a corresponding W_2 can be constructed as

$$W_2 = \begin{bmatrix} \boldsymbol{x}_1', & \boldsymbol{0}' & \boldsymbol{0}' & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \boldsymbol{x}_1' & \boldsymbol{x}_2' & \boldsymbol{0}' & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \boldsymbol{x}_1' & \cdot & \boldsymbol{x}_{1,T-1}' \end{bmatrix}$$

and

$$
\begin{aligned}
y_{it} = {} & \frac{1-\gamma^t}{1-\gamma}\alpha_i + \gamma^t y_{i0} + \sum_{j=0}^{t-1}\gamma^j \boldsymbol{x}'_{i,t-j}\boldsymbol{\beta} \\
& + \boldsymbol{b}'_i \sum_{j=0}^{t-1}\gamma^j \boldsymbol{f}_{t-j} + \sum_{j=0}^{t-1}\gamma^j \epsilon_{i,t-j}.
\end{aligned} \tag{11.7.15}
$$

Under the assumption that $\boldsymbol{f}_t$ are nonstochastic and bounded but $\boldsymbol{b}_i$ are random with mean $\mathbf{0}$ and covariance $E\boldsymbol{b}_i\boldsymbol{b}'_i = \sum_b$, $E(y_{i,t-j}\Delta u_{it})$ is not equal to zero, for $j = 2, \ldots, t$. Therefore, SYR suggest first estimating γ and $\boldsymbol{\beta}$ by the GMM method (3.3.17) using both (11.7.11) and (11.7.12) moment conditions, denoted by $(\hat{\gamma}, \hat{\boldsymbol{\beta}}')$, and constructing estimated residuals $\Delta\boldsymbol{u}_i$ by $\Delta\hat{\boldsymbol{u}}_i = \Delta\boldsymbol{y}_i - \Delta\boldsymbol{y}_{i,-1}\hat{\gamma} - \Delta X_i\hat{\boldsymbol{\beta}}$, where $\Delta\boldsymbol{y}_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$, $\Delta\boldsymbol{y}_{i,-1} = (\Delta y_{i1}, \ldots, \Delta y_{i,T-1})'$ and $\Delta X_i = (\Delta\boldsymbol{x}_{i1}, \ldots, \Delta\boldsymbol{x}_{iT})'$. Then estimate $(\gamma, \boldsymbol{\beta}')$ again using moment conditions (11.7.12) only,

$$
\begin{aligned}
\begin{pmatrix}\tilde{\gamma} \\ \tilde{\boldsymbol{\beta}}\end{pmatrix} = {} & \left\{\left[\sum_{i=1}^{N}\begin{pmatrix}\Delta\boldsymbol{y}'_{i,-1} \\ \Delta X'_i\end{pmatrix} W_{2i}\right]\hat{\Omega}^{-1}\left[\sum_{i=1}^{N} W'_{2i}(\Delta\boldsymbol{y}_{i,-1}, \Delta X_i)\right]\right\}^{-1} \\
& \cdot \left\{\left[\sum_{i=1}^{N}\begin{pmatrix}\Delta\boldsymbol{y}'_{i,-1} \\ \Delta X'_i\end{pmatrix} W_{2i}\right]\hat{\Omega}^{-1}\left[\sum_{i=1}^{N} W'_{2i}\Delta\boldsymbol{y}_i\right]\right\},
\end{aligned} \tag{11.7.16}
$$

where $\hat{\Omega}^{-1} = N^{-1}\sum_{i=1}^{N} W'_{2i}\Delta\hat{\boldsymbol{u}}_i\Delta\hat{\boldsymbol{u}}'_i W_{2i}$. Under the null of cross-sectional independence, both estimators are consistent. Under the alternative, $(\hat{\gamma}, \hat{\boldsymbol{\beta}}')$ may not be consistent, but (11.7.16) remains consistent. Therefore, SYR, following the idea of Sargan (1958) and Hansen (1982), suggest using the test statistic

$$
\begin{aligned}
& N^{-1}\left(\sum_{i=1}^{N}\Delta\hat{\boldsymbol{u}}'_i W_i\right)\hat{\Psi}^{-1}\left(\sum_{i=1}^{N} W'_i\Delta\hat{\boldsymbol{u}}_i\right) \\
& \quad - N^{-1}\left(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{u}}'_i W_{2i}\right)\tilde{\Psi}^{-1}\left(\sum_{i=1}^{N} W'_{2i}\Delta\tilde{\boldsymbol{u}}_i\right)
\end{aligned} \tag{11.7.17}
$$

where $\Delta\tilde{\boldsymbol{u}}_i = \Delta\boldsymbol{y}_i - \Delta\boldsymbol{y}_{i,-1}\tilde{\gamma} - \Delta X_i\tilde{\boldsymbol{\beta}}$, $\hat{\Psi} = \frac{1}{N}\sum_{i=1}^{N} W'_i\Delta\hat{\boldsymbol{u}}_i\Delta\hat{\boldsymbol{u}}'_i W_i$ and $\tilde{\Psi} = \frac{1}{N}\sum_{i=1}^{N} W'_{2i}\Delta\tilde{\boldsymbol{u}}_i\Delta\tilde{\boldsymbol{u}}'_i W_{2i}$. SYR show that under the null of cross-sectional independence, (11.7.17) converges to chi-square distribution with $\frac{T(T-1)}{2}(1+K)$ degrees of freedom as $N \rightarrow \infty$.

The advantage of the SYR test is that the test statistic (11.7.17) has power even when $\sum_{j=1}^{N}\rho_{ij} = 0$. Monte Carlo studies conducted by SYR show that the test statistic (11.7.17) performs well if the cross-sectional dependence is driven by nonstochastic $\boldsymbol{f}_t$ but stochastic $\boldsymbol{b}_i$. However, if the cross-sectional dependence is driven by fixed $\boldsymbol{b}_i$ and stochastic $\boldsymbol{f}_t$, then the test statistic is unlikely to have power because $E(\Delta y_{i,t-j}\Delta u_i) = 0$ if $\boldsymbol{f}_t$ is independently distributed over time.[8]

[8] $E(\Delta y_{i,t-j}\Delta u_{it})$ is not equal to 0 if $\boldsymbol{f}_t$ is serially correlated. However, if $\boldsymbol{f}_t$ is serially correlated, then u_{it} is serial correlated and $y_{i,t-j}$ is not a legitimate instrument if the order of serial correlation is greater than j. Laggard y can be a legitimate instrument only if $E(\Delta u_{it}y_{i,t-s}) = 0$. Then the GMM estimator of (3.3.17) will have to be modified accordingly.

11.7.3 Limited Dependent Variable Model

Many limited dependent variable models take the form of relating observed y_{it} to a latent y_{it}^*, (e.g., Chapters 6 and 7),

$$y_{it}^* = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}, \tag{11.7.18}$$

through a link function $g(\cdot)$

$$y_{it} = g(y_{it}^*). \tag{11.7.19}$$

Two examples are in the binary choice model,

$$g(y_{it}^*) = I(y_{it}^* > 0), \tag{11.7.20}$$

and in the Tobit model,

$$g(y_{it}^*) = y_{it}^* I(y_{it}^* > 0), \tag{11.7.21}$$

where I(A) is an indicator function that takes the value 1 if A occurs and zero otherwise.

There is a fundamental difference between the linear model and limited dependent variable model. There is a one-to-one correspondence between v_{it} and y_{it} in the linear model, but not in the limited dependent variable model. The likelihood for observing $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$,

$$P_t = \int_{A(\boldsymbol{v}_t \mid \boldsymbol{y}_t)} f(\boldsymbol{v}_t) d\boldsymbol{v}_t, \tag{11.7.22}$$

where $A(\boldsymbol{v}_t \mid \boldsymbol{y}_t)$ denotes the region of integration of $\boldsymbol{v}_t = (v_{1t}, \ldots, v_{Nt})'$, which is determined by the realized $\boldsymbol{y}_t$ and the form of the link function. For instance, in the case of the Probit model, $A(\boldsymbol{v}_t \mid \boldsymbol{y}_t)$ denotes the region $(a_{it} < v_{it} < b_{it})$, where $a_{it} = -\boldsymbol{x}_{it}'\boldsymbol{\beta}, b_{it} = \infty$ if $y_{it} = 1$ and $a_{it} = -\infty$, and $b_{it} = -\boldsymbol{x}_{it}'\boldsymbol{\beta}$ if $y_{it} = 0$.

Under the assumption that v_{it} is independently normally distributed across i, Hsiao, Pesaran, and Pick (2012) show that the Lagrangian multiplier test statistic of cross-sectional independence takes an analogous form:

$$LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tilde{\rho}_{ij}^2, \tag{11.7.23}$$

where

$$\tilde{\rho}_{ij} = \frac{T^{-1}\sum_{t=1}^T \tilde{v}_{it}\tilde{v}_{jt}}{\sqrt{T^{-1}\sum_{t=1}^T \tilde{v}_{it}^2}\sqrt{T^{-1}\sum_{t=1}^T \tilde{v}_{jt}^2}}, \tag{11.7.24}$$

and $\tilde{v}_{it} = E(v_{it} \mid y_{it})$, the conditional mean of v_{it} given y_{it}. For instance, in the case of the Probit model,

$$\tilde{v}_{it} = \frac{\phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})}{\Phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})[1 - \Phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})]}[y_{it} - \Phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})]. \tag{11.7.25}$$

In the case of the Tobit model,

$$\tilde{v}_{it} = (y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta})I(y_{it} > 0) - \sigma_i \frac{\phi\left(\frac{\boldsymbol{x}_{it}'\boldsymbol{\beta}}{\sigma_i}\right)}{\Phi\left(-\frac{\boldsymbol{x}_{it}'\boldsymbol{\beta}}{\sigma_i}\right)}[1 - I(y_{it} > 0)], \tag{11.7.26}$$

where $\sigma_i^2 = \text{Var}\ (v_{it}), \phi(\cdot)$ and $\Phi(\cdot)$ denote standard normal and integrated standard normal. Under the null of cross-sectional independence, (11.7.23) converges to a chi-square distribution with $\frac{N(N-1)}{2}$ degrees of freedom if N is fixed and $T \to \infty$. When N is also large,

$$\sqrt{\frac{2}{N(N-1)}}LM \tag{11.7.27}$$

is asymptotically standard normally distributed.

When N is large and T is finite, the LM test statistic is not centered properly. However, for the nonlinear model, the bias correction factor is not easily derivable. Hsiao, Pesaran, and Pick (2012) suggest constructing the Pesaran (2020) CD statistic using $\tilde{v}_{it}$.

Sometimes, the derivation of $\tilde{v}_{it}$ is not straightforward for nonlinear model. Hsiao, Pesaran, and Pick (2012) suggest replacing $\tilde{v}_{it}$ by

$$v_{it}^* = y_{it} - E(y_{it} \mid \boldsymbol{x}_{it}) \tag{11.7.28}$$

in the construction of the LM or CD test statistic. Monte Carlo experiments conducted by Hsiao, Pesaran, and Pick (2012) show that there is very little difference between the two procedures to construct CD tests.

Mao and Shen (2013) consider China's housing price model using 30 provincial-level quarterly data from the second quarter of 2001 to the fourth quarter of 2012 of the logarithm of seasonally adjusted real house price, y_{it}, as a linear function of the logarithm of seasonally adjusted real per capita wage income (x_{1it}), the logarithm of real long-term interest rate (x_{2it}), and the logarithm of urban population (x_{3it}). Table 11.1 provides Mao and Shen (2013) estimates of the mean group estimator $\hat{\boldsymbol{\beta}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i$ for the cross-sectionally independent heterogeneous model (MG),

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + v_{it}; \tag{11.7.29}$$

the Pesaran (2006) common correlated effects heterogeneous model (CCEMG),

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \bar{y}_t c_i + \bar{\boldsymbol{x}}_t'\boldsymbol{d}_i + v_{it}; \tag{11.7.30}$$

and the homogeneous common correlated effects model (CCEP),

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \bar{y}_t\tau_i + \bar{\boldsymbol{x}}_t'\boldsymbol{d}_i + v_{it}. \tag{11.7.31}$$

It can be seen from the results in Table 11.1 that: (i) the estimated slope coefficients, $\boldsymbol{\beta}$, are very sensitive to the adjustment (CCEMG or CCEP) or nonadjustment of

Table 11.1. *Common correlated effects estimation*

	MG			CCEMG			CCEP		
x1	$1.088^{\ddagger}$	$1.089^{\ddagger}$	$0.979^{\ddagger}$	0.264	0.313^{+}	0.308^{+}	$0.388^{\dagger}$	$0.467^{\ddagger}$	$0.449^{\ddagger}$
	(0.058)	(0.056)	(0.114)	(0.176)	(0.173)	(0.170)	(0.169)	(0.165)	(0.170)
x2	–	−0.003	−0.052	–	$6.453^{\dagger}$	4.399	–	4.796	4.387
	–	(0.058)	(0.057)	–	(2.927)	(2.839)	–	(3.943)	(3.401)
x3	–	–	0.718	–	–	−0.098	–	–	0.104
	–	–	(0.484)	–	–	(0.552)	–	(0.130)	
CD	$28.15^{\ddagger}$	$30.39^{\ddagger}$	$27.64^{\ddagger}$	$-4.257^{\ddagger}$	$-.4173^{\ddagger}$	$-4.073^{\ddagger}$	$-4.521^{\ddagger}$	$-4.494^{\ddagger}$	$-4.518^{\ddagger}$

Notes: Symbols $^{+}$, †, and ‡ denote that the corresponding statistics are significant at 10%, 5%, and 1% level, respectively. The values in parenthesis are corresponding standard errors.
Source: Mao and Shen (2013, Table V).

cross-sectional dependence; and (ii) the suggested approach to control the impact of cross-sectional dependence works only if the observed data satisfy the assumptions underlying the approach. (See Section 10.2.3 for the limitation of augmenting regression models by the cross-sectional mean). As one can see from Table 11.1, the Pesaran (2020) CD tests (Equation 11.7.5) of the residuals of (11.7.30) and (11.7.31) indicate that significant cross-sectional dependence remains. Only by further adjusting the common correlated effects model residuals by a spatial model with the spatial weight matrix specified in terms of the geometric distance between region i and j, can Mao and Shen (2013) achieve cross-sectional independence of their model.