CHAPTER 5

Dynamic System

5.1 INTRODUCTION

One of the prominent features of econometric analysis is the incorporation of economic theory into the analysis of numerical and institutional data. Economists, from León Walras onwards, perceive the economy as a coherent system. The interdependence of sectors of an economy is represented by a set of functional relations, each representing an aspect of the economy by a group of individuals, firms, or authorities. The variables entering into these relations consist of a set of *endogenous* (or *joint dependent*) variables, whose formations are conditioning on a set of exogenous variables which the economic theory regards as given. Two approaches have been proposed to model a system of economic behaviors – the structural equation approach and the reduced form approach. The structural approach constructs the system of behavioral equations from a priori assumed "theory," based on behavioral hypotheses and institutional and technological knowledge. The reduced form approach is to construct a statistical mapping between inputs (exogenous variables) and outputs (endogenous variables). Since different theoretical models may generate the same observed phenomena, to ensure the one-to-one relationship between the specified model and the observed phenomena, a priori restrictions need to be imposed to exclude other "observationally equivalent" models[1] (e.g. Dufour and Hsiao 2008; Hsiao 1983). The resulting statistical inference is conditioning on the hypothesized theoretical model. Statistical inference could be grossly misleading if the hypothesized model is not compatible with the data generating process of the observed sample. Sims (1980) has criticized that many models are identified because of the "incredible" prior restrictions. Liu (1960) and Sims (1980), among others, have therefore favored the reduced form approach. Since economic behavior is inherently dynamic because of the institutional, technological, and behavioral rigidities, vector autoregressive models (VAR) have been proposed as a reduced form formulation to take account of both the joint dependence of state variables and their dynamic dependence. We shall discuss panel vector autoregressive modeling when the time series dimension T is fixed and cross-sectional dimension N is large in Section 5.2. However, since the time series properties of a variable behave very differently if the variable is stationary or nonstationary as time series dimension T increases, Section 5.3 discusses the estimation of a cointegrated system when both N and T are large. Section 5.4 focuses on unit root and cointegration tests. Section 5.5 covers the single equation approach to estimating an equation in a dynamic simultaneous equations model.

[1] By observationally equivalent structures we mean all structures that could generate the same observed sample characteristics (e.g., Hsiao 1983).

5.2 PANEL VECTOR AUTOREGRESSIVE MODELS

5.2.1 "Homogeneous" Panel VAR Models

5.2.1.1 Model Formulation

Vector autoregressive models have become a widely used modeling tool in economics (e.g. Hsiao 1979a, 1979b, 1982; Sims 1980). By "homogeneous" panel VAR (PVAR) models, we mean conditional on the unobserved time-invariant individual heterogeneity; the slope coefficients are identical over i and t (e.g. Holtz-Eakin, Newey, and Rosen 1988),

$$\Phi(L)\boldsymbol{w}_{it} = \boldsymbol{w}_{it} - \Phi_1 \boldsymbol{w}_{i,t-1} \ldots - \Phi_p \boldsymbol{w}_{i,t-p} = \boldsymbol{\alpha}_i^* + \boldsymbol{\delta}^* t + \boldsymbol{\epsilon}_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{5.2.1}$$

where $\boldsymbol{w}_{it}$ denotes an $m \times 1$ vector of observed random variables, $\boldsymbol{\alpha}_i^*$ is an $m \times 1$ vector of individual specific constants that vary with $i, \boldsymbol{\delta}^*$ is an $m \times 1$ vector of constants, $\boldsymbol{\epsilon}_{it}$ is an $m \times 1$ vector of random variables that is independently identically distributed over t with mean zero and covariance matrix Ω, and $\Phi(L) = I_m - \Phi_1 L - \ldots - \Phi_p L^p$ is a pth-order polynomial of the lag operator $L, L^s \boldsymbol{w}_t = \boldsymbol{w}_{t-s}$. We assume the order of autoregressive operator p is known. For empirically determining the lag length for (5.2.1), see Han et al. (2017).

With unrestricted intercepts or time trends, the time series property of $\boldsymbol{w}_{it}$ can be different whether $\boldsymbol{w}_{it}$ contains unit roots or not, or if elements of $\boldsymbol{w}_{it}$ are cointegrated as t increases[2] (e.g. Johansen 1995; Pesaran, Shin, and Smith 2000; Phillips 1991; Sims, Stock, and Watson 1990). To make sure that the time series property of $\boldsymbol{w}_{it}$ remains the same whether $\boldsymbol{w}_{it}$ contain unit roots or not, instead of considering (5.2.1) directly, we consider

$$\Phi(L)(\boldsymbol{w}_{it} - \boldsymbol{\eta}_i - \boldsymbol{\delta} t) = \boldsymbol{\epsilon}_{it}, \tag{5.2.2}$$

where the roots of the determinant equation

$$\mid \Phi(\rho) \mid = 0 \tag{5.2.3}$$

are either equal to unity or fall outside the unit circle. Under the assumption that $E\boldsymbol{\epsilon}_{it} = \mathbf{0}$, it follows that

$$E(\boldsymbol{w}_{it} - \boldsymbol{\eta}_i - \boldsymbol{\delta} t) = \mathbf{0}. \tag{5.2.4}$$

To allow for the possibility of the presence of unit roots, we assume that

$$E(\boldsymbol{w}_{it} - \boldsymbol{\eta}_i - \boldsymbol{\delta} t)(\boldsymbol{w}_{it} - \boldsymbol{\eta}_i - \boldsymbol{\delta} t)' = \Psi_t. \tag{5.2.5}$$

Models (5.2.2)–(5.2.5) encompasses many well known panel VAR models (PVAR) as special cases. For instance,

[2] We say that $\boldsymbol{y}_t$ is stationary if $E\boldsymbol{y}_t = \boldsymbol{\mu}$, $E[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-s} - \boldsymbol{\mu})'] = E[(\boldsymbol{y}_{t+q} - \boldsymbol{\mu})(\boldsymbol{y}_{t+q-s} - \boldsymbol{\mu})']$. We say that $\boldsymbol{y}_t$ is integrated of order $d, I(d)$, if $(1-L)^d \boldsymbol{y}_t$ is stationary, $I(0)$. If $\boldsymbol{y}_t \sim I(d)$ but $\boldsymbol{\beta}' \boldsymbol{y}_t \sim I(d-c)$, say $d = 1, c = 1$, then y_t is cointegrated of order c. The maximum number of linearly independent vector $\boldsymbol{\beta}$ is called the rank of cointegration. For any $m \times 1$ $I(d)$ process, the cointegration rank can vary between 0 and $m - 1$ (e.g., Engle and Granger 1987; Intriligator, Bodkin, and Hsiao 1996).

(i) Stationary PVAR with individual-specific effects.

Let $\boldsymbol{\delta} = \mathbf{0}_{m\times 1}$. If all roots of (5.2.3) fall outside the unit circle, (5.2.2) becomes (5.2.1) with $\boldsymbol{\alpha}_i^* = -\Pi\boldsymbol{\eta}_i$ and

$$\Pi = -\left(I_m - \sum_{j=1}^{p} \Phi_j\right). \tag{5.2.6}$$

(ii) Trend-stationary PVAR with individual-specific effects.

If all roots of (5.2.3) fall outside the unit circle and $\boldsymbol{\delta} \neq \mathbf{0}$, we have

$$\Phi(L)\boldsymbol{w}_{it} = \boldsymbol{\alpha}_i^* + \boldsymbol{\delta}^* t + \boldsymbol{\epsilon}_{it}, \tag{5.2.7}$$

where $\boldsymbol{\alpha}_i^* = -\Pi\boldsymbol{\eta}_i + (\Gamma + \Pi)\boldsymbol{\delta}$.

$$\Gamma = -\Pi + \sum_{j=1}^{p} j\Phi_j, \tag{5.2.8}$$

and $\boldsymbol{\delta}^* = -\Pi\boldsymbol{\delta}$.

(iii) PVAR with unit roots (but non-cointegrated) and individual-specific effects.

$$\Phi^*(L)\Delta\boldsymbol{w}_{it} = -\Pi^*\boldsymbol{\delta} + \boldsymbol{\epsilon}_{it} \tag{5.2.9}$$

where $\Delta = (1 - L)$,

$$\Phi^*(L) = I_m - \sum_{j=1}^{p-1} \Phi_j^* L^j, \tag{5.2.10}$$

$\Phi_j^* = -(I_m - \sum_{\ell=1}^{j} \Phi_\ell)$, $j = 1, 2, \ldots, p-1$, and $\Pi^* = -\left(I_m - \sum_{j=1}^{p-1} \Phi_j^*\right)$.

(iv) Cointegrated PVAR with individual-specific effects.

If some roots of (5.2.3) are equal to unity and rank $(\Pi) = r, 0 < r < m$, (5.2.2) may be rewritten in the form of a panel vector error corrections model

$$\Delta\boldsymbol{w}_{it} = \boldsymbol{\alpha}_i^* + (\Gamma + \Pi)\boldsymbol{\delta} + \boldsymbol{\delta}^* t + \Pi\boldsymbol{y}_{i,t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta\boldsymbol{w}_{i,t-j} + \boldsymbol{\epsilon}_{it}, \tag{5.2.11}$$

where $\Gamma_j = -\sum_{s=j+1}^{p} \Phi_s$, $j = 1, \ldots, p-1$, and Π can be decomposed as the product of two $m \times r$ matrices J and $\boldsymbol{\beta}$, with rank r, $\Pi = J\boldsymbol{\beta}'$ and $J_\perp' \boldsymbol{\beta}_\perp$ is of rank $m - r$, where $J_\perp$ and $\boldsymbol{\beta}_\perp$ are $m \times (m - r)$ matrices of full column rank such that $J'J_\perp = \mathbf{0}$ and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$ (Johansen 1995).

The reason of formulating the panel VAR model in terms of (5.2.2)–(5.2.5) rather than (5.2.1) is that it allows putting restrictions on the model intercepts and trend term so that the time series properties of $\boldsymbol{w}_{it}$ remain the same with the presence of unit roots and cointegration. For instance, when $\boldsymbol{\delta} = \mathbf{0}$ and if the roots of (5.2.3) all fall outside the unit circle, $\boldsymbol{w}_{it}$ exhibit no trend growth. However, if one or more roots of (5.2.3) are equal to unity, $\boldsymbol{\alpha}_i^*$ is unrestricted, then $\boldsymbol{w}_{it}$ will exhibit differential trend growth if unit roots are present. If $\boldsymbol{\delta} \neq \mathbf{0}$, (5.2.2) can ensure that the trend growth of $\boldsymbol{w}_{it}$ is linear whether the roots of (5.2.3) are all outside the unit circle or some or all are equal to unity by putting the restrictions on $\boldsymbol{\delta}$. But if the trend term is unrestricted, then $\boldsymbol{w}_{it}$ exhibit a linear trend if the roots of (5.2.3) all fall outside the unit circle and would exhibit quadratic trends if one or more roots of (5.2.3) are equal to unity (e.g. Pesaran, Shin, and Smith 2000).

If $\boldsymbol{\alpha}_i^*$ are assumed to be randomly distributed with a common mean and constant covariance matrix, (5.2.1) is a random-effects PVAR. The random-effects PVAR has the advantages that the number of unknown parameters stay constant as sample size increases and efficient inference of $\Phi(L)$ can be derived by considering the marginal distribution of $(\boldsymbol{w}_{io}, \ldots, \boldsymbol{w}_{it})$,

$$
\begin{aligned}
f(\boldsymbol{w}_{io}, \ldots, \boldsymbol{w}_{iT}) &= \int f(\boldsymbol{w}_{io}, \ldots, \boldsymbol{w}_{iT} \mid \boldsymbol{\alpha}_i^*) dG(\boldsymbol{\alpha}_i^*) \\
&= \int \prod_{t=p}^{T} f(\boldsymbol{w}_{it} \mid \boldsymbol{w}_{i,t-1}, \ldots, \boldsymbol{w}_{i,t-p}, \boldsymbol{\alpha}_i^*) \\
&\quad \times f(\boldsymbol{w}_{ip}, \ldots, \boldsymbol{w}_{io} \mid \boldsymbol{\alpha}_i^*) dG(\boldsymbol{\alpha}_i^*).
\end{aligned} \tag{5.2.12}
$$

However, besides the difficulties of postulating the probability distribution of unobserved effects, $\boldsymbol{\alpha}_i^*$, the derivation of (5.2.12) appears computationally complicated because (5.2.12) involves multiple integration of $m \times (T+1)$ dimensions. On the other hand, treating $\boldsymbol{\alpha}_i^*$ as a fixed constant, $f(\boldsymbol{w}_{it} \mid \boldsymbol{w}_{i,t-1}, \ldots, \boldsymbol{w}_{t-p}; \boldsymbol{\alpha}_i^*)$ is independently distributed over t. Moreover, when even $\boldsymbol{\alpha}_i^*$ are random, the conditional inference of $f(\boldsymbol{w}_{io}, \ldots, \boldsymbol{w}_{iT} \mid \boldsymbol{\alpha}_i^*)$ remains valid, although it is not efficient. We shall therefore focus on inference with $\boldsymbol{\alpha}_i^*$ fixed and discuss conditional inference procedures.

When the time dimension of the panel is short, just as in the single-equation fixed-effects dynamic panel data model (Section 3.5), (5.2.2) raises the classical incidental parameters problem and the issue of modeling initial observations. For ease of exposition, we shall illustrate the estimation and inference by considering $p = 1$, namely, the model of

$$
(I - \Phi L)(\boldsymbol{w}_{it} - \boldsymbol{\eta}_i - \boldsymbol{\delta} t) = \boldsymbol{\epsilon}_{it}, \qquad \begin{aligned} i &= 1 \ldots, N, \\ t &= 1, \ldots, T, \end{aligned} \tag{5.2.13}
$$

We also assume that $\boldsymbol{w}_{i0}$ are available.

5.2.1.2 GMM Estimation

Just as in the single equation case, the individual effects $\boldsymbol{\eta}_i$ can be eliminated by first differencing (5.2.13).

$$
\Delta \boldsymbol{w}_{it} - \boldsymbol{\delta} = \Phi(\Delta \boldsymbol{w}_{i,t-1} - \boldsymbol{\delta}) + \Delta \boldsymbol{\epsilon}_{it}, t = 2, \ldots, T. \tag{5.2.14}
$$

Thus, we have the orthogonality conditions,

$$
E\left\{[(\Delta \boldsymbol{w}_{it} - \boldsymbol{\delta}) - \Phi(\Delta \boldsymbol{w}_{i,t-1} - \boldsymbol{\delta})]\boldsymbol{q}_{it}'\right\} = \mathbf{0}, \qquad t = 2, \ldots, T. \tag{5.2.15}
$$

where

$$
\boldsymbol{q}_{it} = (1, \boldsymbol{w}_{io}', \ldots, \boldsymbol{w}_{i,t-2}')'. \tag{5.2.16}
$$

Stacking the $(T-1)$ (5.2.14) together yields

$$
S_i = R_i \Lambda' + E_i, \quad i = 1, 2, \ldots, N, \tag{5.2.17}
$$

where

$$
\begin{aligned}
S_i &= (\Delta \boldsymbol{w}_{i2}, \Delta \boldsymbol{w}_{i3}, \ldots, \Delta \boldsymbol{w}_{iT})', \quad E_i = (\Delta \boldsymbol{\epsilon}_{i2}, \ldots, \Delta \boldsymbol{\epsilon}_{iT})' \\
R_i &= (S_{i,-1}, \boldsymbol{e}_{T-1}), \quad S_{i,-1} = (\Delta \boldsymbol{w}_{i1}, \ldots, \Delta \boldsymbol{w}_{i,T-1})', \\
\Lambda &= (\Phi, \boldsymbol{a}_1), \quad \boldsymbol{a}_1 = (I_m - \Phi)\boldsymbol{\delta},
\end{aligned} \tag{5.2.18}
$$

and $\boldsymbol{e}_{T-1}$ denotes a $(T-1) \times 1$ vector of ones. Premultiplying (5.2.17) by the $(mT/2+1)$ $(T-1) \times (T-1)$ block-diagonal instrumental variable matrix Q_i,

$$Q_i = \begin{pmatrix} \boldsymbol{q}_{i2} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{q}_{is} & . & . & \\ . & . & . & . & \\ . & . & & . & \\ \mathbf{0} & . & & & \boldsymbol{q}_{iT'} \end{pmatrix}, \tag{5.2.19}$$

one obtains

$$Q_i S_i = Q_i R_i \Lambda' + Q_i E_i, \tag{5.2.20}$$

the transpose of which in vectorized form becomes[3]

$$(Q_i \otimes I_m) \text{ vec } (S_i') = (Q_i R_i \otimes I_m)\boldsymbol{\lambda} + (Q_i \otimes I_m) \text{ vec } (E_i'), \tag{5.2.21}$$

where $\boldsymbol{\lambda} = \text{vec } (\Lambda)$ and vec (.) is the operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. Thus, the GMM estimator of $\boldsymbol{\lambda}$ can be obtained by minimizing (Binder, Hsiao, and Pesaran 2005)

$$\begin{aligned} &\left[\sum_{i=1}^{N} \Big((Q_i \otimes I_m) \text{ vec } (S_i') - (Q_i R_i \otimes I_m)\boldsymbol{\lambda}\Big)\right]' \\ &\cdot \left[\sum_{i=1}^{N} (Q_i \otimes I_m)\tilde{\Omega}(Q_i \otimes I_m)'\right]^{-1} \\ &\cdot \left[\sum_{i=1}^{N} \Big((Q_i \otimes I_m) \text{ vec } (S_i') - (Q_i R_i \otimes I_m)\boldsymbol{\lambda}\Big)\right], \end{aligned} \tag{5.2.22}$$

where

$$\tilde{\Omega} = \begin{bmatrix} 2\Omega & -\Omega & \mathbf{0} & \cdots & \mathbf{0} \\ -\Omega & 2\Omega & -\Omega & & \\ \mathbf{0} & -\Omega & 2\Omega & & \\ \vdots & & & \ddots & \\ \mathbf{0} & & & & 2\Omega \end{bmatrix}. \tag{5.2.23}$$

The moment conditions relevant to the estimation of Ω are given by

$$\begin{aligned} E\{[\Delta \boldsymbol{w}_{it} - \boldsymbol{\delta} - \Phi(\Delta \boldsymbol{w}_{i,t-1} - \boldsymbol{\delta})][\Delta \boldsymbol{w}_{it} - \boldsymbol{\delta} \\ - \Phi(\Delta \boldsymbol{w}_{i,t-1} - \boldsymbol{\delta})]' - 2\Omega\} = \mathbf{0}, t = 2, 3, \ldots, T. \end{aligned} \tag{5.2.24}$$

Also, in the trend-stationary case, upon estimation of $\boldsymbol{a}_1, \boldsymbol{\delta}$ may be obtained as

$$\hat{\boldsymbol{\delta}} = (I_m - \hat{\Phi})^{-1} \hat{\boldsymbol{a}}_1. \tag{5.2.25}$$

The GMM estimator is consistent and asymptotically normally distributed if T is fixed when $N \to \infty$ and all the roots of (5.2.3) fall outside the unit circle. If $\frac{T}{N} \to a \neq 0 < 1/2$ as N increases, then, as shown by Alvarez and Arellano, the GMM estimator is asymptotically biased and the bias is of order $\sqrt{a}$. Moreover, the GMM breaks down if some

[3] Vec $(ABC) = (C' \otimes A)$ vec (B); see Magnus and Neudecker (1999).

roots are equal to unity. To see this, note that a necessary condition for the GMM estimator (5.2.22) to exist is that rank $(N^{-1}\sum_{i=1}^{N} Q_i R_i) = m+1$ as $N \to \infty$. In the case where $\Phi = I_m, \Delta \boldsymbol{w}_{it} = \boldsymbol{\delta} + \boldsymbol{\epsilon}_{it}$, and $\boldsymbol{w}_{it} = \boldsymbol{w}_{io} + \boldsymbol{\delta} t + \sum_{\ell=1}^{t} \boldsymbol{\epsilon}_{i\ell}$. Thus, it follows that for $t = 2, 3, \ldots, T, j = 0, 1, \ldots, t-2$, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} \Delta \boldsymbol{w}_{i,t-1}\boldsymbol{w}_{ij}' \longrightarrow \boldsymbol{\delta}(\boldsymbol{w}_{io} + \boldsymbol{\delta} j)', \tag{5.2.26}$$

which is of rank one. In other words, when $\Phi = I_m$, the elements of $\boldsymbol{q}_{it}$ are not legitimate instruments.

5.2.1.3 *(Transformed) Maximum Likelihood Estimator*

We note that given on $\Delta \boldsymbol{w}_{i1}$ (5.2.14) is well defined for $t = 2, \ldots, T$. However, $\Delta \boldsymbol{w}_{i1}$ is random. Equation (5.2.13) implies that $\Delta \boldsymbol{w}_{i1}$ equals[4]

$$\Delta \boldsymbol{w}_{i1} - \boldsymbol{\delta} = -(I - \Phi)(\boldsymbol{w}_{io} - \boldsymbol{\eta}_i) + \boldsymbol{\epsilon}_{i1}. \tag{5.2.27}$$

We note that by (5.2.4) and (5.2.5), $E(\Delta \boldsymbol{w}_{i1} - \boldsymbol{\delta}) = -(I-\Phi)E(\boldsymbol{w}_{io} - \boldsymbol{\eta}_i) + E\boldsymbol{\epsilon}_{i1} = \boldsymbol{0}$ and $E(\Delta \boldsymbol{w}_{i1} - \boldsymbol{\delta})(\Delta \boldsymbol{w}_{i1} - \boldsymbol{\delta})' = (I-\Phi)\Psi_0(I-\Phi)' + \Omega = \Psi_1$ where $\Psi_0 = E(\boldsymbol{w}_{io} - \boldsymbol{\eta}_i)(\boldsymbol{w}_{io} - \boldsymbol{\eta}_i)'$. Therefore, the joint likelihood of $\Delta \boldsymbol{w}_i' = (\Delta \boldsymbol{w}_{i1}', \ldots, \Delta \boldsymbol{w}_{iT}')$ is well defined and does not involve incidental parameters. Under the assumption that $\boldsymbol{\epsilon}_{it}$ is normally distributed, the likelihood function is given by

$$\prod_{i=1}^{N} (2\pi)^{-\frac{T}{2}} \mid \Omega^* \mid^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{r}_i - H_i\boldsymbol{\phi})'\Omega^{*-1}(\boldsymbol{r}_i - H_i\boldsymbol{\phi})\right], \tag{5.2.28}$$

where

$$\begin{aligned}
\boldsymbol{r}_i &= (\Delta \boldsymbol{w}_i - \boldsymbol{e}_T \otimes \boldsymbol{\delta}),\\
H_i &= G_i' \otimes I_m,\\
G_i &= (\boldsymbol{0}, \Delta \boldsymbol{w}_{i1} - \boldsymbol{\delta}, \ldots, \Delta \boldsymbol{w}_{i,T-1} - \boldsymbol{\delta}),\\
\boldsymbol{\phi} &= \text{vec}\,(\Phi),
\end{aligned}$$

$$\Omega^* = \begin{pmatrix} \Psi_1 & -\Omega & \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ -\Omega & 2\Omega & -\Omega & \boldsymbol{0} & & \\ \boldsymbol{0} & -\Omega & 2\Omega & -\Omega & & \\ \vdots & & & \ddots & & \\ \boldsymbol{0} & & & & & 2\Omega \end{pmatrix}, \tag{5.2.29}$$

and $\boldsymbol{e}_T$ is a $T \times 1$ vector of $(1, \ldots, 1)'$. Maximizing the logarithm of (5.2.28), $\ell(\boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}' = (\boldsymbol{\delta}', \boldsymbol{\phi}', \boldsymbol{\sigma}')'$ where $\boldsymbol{\sigma}$ denotes the unknown element of Ω^* yields the (transformed) MLE that is consistent and asymptotically normally distributed with the asymptotic covariance matrix given by $-\left[E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)\right]^{-1}$ as $N \to \infty$ independent of whether $\boldsymbol{w}_{it}$ contain unit roots or cointegrated.

[4] If $\boldsymbol{w}$ is considered reduced from representation of (5.5.7), then we may partition $\boldsymbol{w} = (\boldsymbol{y}', \boldsymbol{x}')'$ where $\boldsymbol{y}$ are joint dependent and $\boldsymbol{x}$ are exogenous. Using the same arguments as those for the transformed MLE for the single-equation case, we may express

$$\boldsymbol{w}_{io} = A\bar{\boldsymbol{x}}_i + [I - H_1]^{-1}\boldsymbol{\eta}_i + w_i$$

where the data generating process of $\boldsymbol{x}_{it}$ are now of the multivariate form of (3.5.4).

5.2.1.4 Minimum Distance Estimator

We note that conditional on Ω^*, the MLE of Φ and $\boldsymbol{\delta}$ is equivalent to the minimum distance estimator (MDE) that minimizes

$$\sum_{i=1}^{N}(\boldsymbol{r}_i - H_i\boldsymbol{\phi})'\Omega^{*-1}(\boldsymbol{r}_i - H_i\boldsymbol{\phi}). \tag{5.2.30}$$

Furthermore, conditional on $\boldsymbol{\delta}$ and Ω^*, the MDE of Φ is given by

$$\hat{\boldsymbol{\phi}} = \left(\sum_{i=1}^{N} H_i'\Omega^{*-1}H_i\right)^{-1}\left(\sum_{i=1}^{N} H_i'\Omega^{*-1}\boldsymbol{r}_i\right). \tag{5.2.31}$$

Conditional on Φ and Ω^*, the MDE of $\boldsymbol{\delta}$ is equal to

$$\hat{\boldsymbol{\delta}} = (NP\Omega^{*-1}P')^{-1}\left[\sum_{i=1}^{N} P\Omega^{*-1}(\Delta\boldsymbol{w}_i - L_i\boldsymbol{\phi})\right], \tag{5.2.32}$$

where

$$P = (I_m, I_m - \Phi', I_m - \Phi', \ldots, I_m - \Phi'), \tag{5.2.33}$$

and

$$L_i = K_i' \otimes I_m, \text{ and } K_i = (\boldsymbol{0}, \Delta\boldsymbol{w}_{i1}, \ldots, \Delta\boldsymbol{w}_{i,T-1}).$$

Conditional on $\boldsymbol{\delta}$,

$$\hat{\Psi}_1 = \frac{1}{N}\sum_{i=1}^{N}(\Delta\boldsymbol{w}_{i1} - \boldsymbol{\delta})(\Delta\boldsymbol{w}_{i1} - \boldsymbol{\delta})'. \tag{5.2.34}$$

and conditional on $\boldsymbol{\delta}, \Phi$,

$$\hat{\Omega} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}[\Delta\boldsymbol{w}_{it} - \boldsymbol{\delta} - \Phi(\Delta\boldsymbol{w}_{i,t-1} - \boldsymbol{\delta})] [\Delta\boldsymbol{w}_{it} - \boldsymbol{\delta} - \Phi(\Delta\boldsymbol{w}_{i,t-1} - \boldsymbol{\delta})]'. \tag{5.2.35}$$

We may iterate between (5.2.31) and (5.2.35) to obtained the feasible MDE using

$$\hat{\boldsymbol{\delta}}^{(0)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta\boldsymbol{w}_{it}, \tag{5.2.36}$$

and

$$\hat{\Phi}^{(0)} = \left[\sum_{i=1}^{N}\sum_{t=3}^{T}(\Delta\boldsymbol{w}_{it} - \boldsymbol{\delta}^{(0)})(\Delta\boldsymbol{w}_{i,t-2} - \boldsymbol{\delta}^{(0)})'\right] \times \left[\sum_{i=1}^{N}\sum_{t=3}^{T}(\Delta\boldsymbol{w}_{i,t-1} - \boldsymbol{\delta}^{(0)})(\Delta\boldsymbol{w}_{i,t-2} - \boldsymbol{\delta}^{(0)})'\right]^{-1} \tag{5.2.37}$$

to start the iteration.

Conditional on Ω^*, the MDE of $\boldsymbol{\phi}$ and $\boldsymbol{\delta}$ is identical to the MLE. When $\boldsymbol{\delta} = \mathbf{0}$ (no trend term), conditional on Ω^*, the asymptotic covariance matrix of the MLE or MDE of $\boldsymbol{\phi}$ is equal to

$$\left[\sum_{i=1}^{N}(K_i \otimes I_m)\Omega^{*-1}(K_i' \otimes I_m)\right]^{-1}. \tag{5.2.38}$$

When Ω^* is unknown, the asymptotic variance–covariance matrices of MLE and MDE of $\boldsymbol{\phi}$ does not converge to (5.2.38) because when lagged dependent variables appear as regressors, the estimation of Φ and Ω^* is not asymptotically independent. The asymptotic variance covariance matrix of the feasible MDE is equal to the sum of (5.2.38) and a positive semi-definite matrix attributable to the estimation error of Ω^* (Hsiao, Pesaran, and Tahmiscioglu 2002).

Both the MLE and MDE always exist whether $\boldsymbol{w}_{it}$ contain unit roots or not. The MLE and MDE are asymptotically normally distributed independent of whether $\boldsymbol{w}_{it}$ are (trend) stationary, integrated, or cointegrated, as T is fixed and $N \to \infty$. Therefore, a conventional likelihood ratio test statistic or a Wald-type test statistic of unit root or the rank of cointegration can be approximated by chi-square statistics. Moreover, the limited Monte Carlo studies conducted by Binder, Hsiao, and Pesaran (2005) show that both the MLE and MDE perform very well in finite sample and dominate the conventional GMM, in particular, if the roots of (5.2.3) are near unity.

5.2.2 Heterogeneous Vector Autoregressive Models

We shall say a VAR model is heterogeneous if the slope coefficients also vary across individuals,

$$\Phi_i(L)\boldsymbol{w}_{it} = \boldsymbol{\alpha}_i^* + \epsilon_{it}, \quad i = 1, \ldots, N, \tag{5.2.39}$$

where

$$\Phi_i(L) = I_m - \Phi_{i1}L - \cdots - \Phi_{ip_i}L^{p_i} \text{ for } i = 1, \ldots, N. \tag{5.2.40}$$

When $\Phi_i(L) \neq \Phi_j(L)$ for $i \neq j$, there is no way one can get a consistent estimator of $\Phi_i(L)$ if T is fixed. When the root of the determinant equation (5.2.3) falls outside the unit circle, the least squares estimator is consistent (at the speed of $\sqrt{T}$) and is asymptotically normally distributed (e.g., Anderson 1971). When the roots of (5.2.3) contain unit roots, the least squares estimator remains consistent, but its limiting distribution depends on if the coefficients correspond the unit root processes or not. For the coefficient that corresponds to the unit root process, it converges at the speed of T and its limiting distribution is non-standard (e.g. Phillips and Durlauf 1986). Therefore, in this subsection, we restrict the discussion to the stationary case. We defer the discussion of nonstationary case to Section 5.3.

5.2.2.1 *Cross-Sectionally Independent Processes*

If $\boldsymbol{\epsilon}_{it}$ is independent over i with mean $\mathbf{0}$ and unrestricted covariance matrix Ω_i and T is large, applying least squares method to (5.2.39) equation by equation yields consistent and efficient estimates of $\Phi_i(L)$ and $\boldsymbol{\alpha}_i^*$.

5.2.2.2 *Cross-Sectionally Dependent Processes*

When $\boldsymbol{\epsilon}_{it}$ are cross-sectionally dependent, (5.2.39) can be put in Zellner's (1962) seemingly unrelated regression framework if N is fixed and T is large. We can first use i-th individual's time series observation to estimate $\Phi_i(L)$ and $\boldsymbol{\alpha}_i^*$. Then use the estimated $\hat{\Phi}_i(L)$ and $\hat{\boldsymbol{\alpha}}_i^*$ to obtain estimated $\hat{\boldsymbol{\epsilon}}_{it}, i = 1, \ldots, N, t = 1, \ldots, T$ and estimate $\Omega_{ij} = E(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}'_{jt})$ by

$$\hat{\Omega}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\epsilon}}_{it}\hat{\boldsymbol{\epsilon}}'_{jt}. \tag{5.2.41}$$

Given $\hat{\Omega}_{ij}, i, j = 1, \ldots, N$, one can stack all N cross-sectionally equations (5.2.39) together and apply the feasible generalized least squares estimator to obtain efficient estimates of $\Phi_i(L)$ and $\boldsymbol{\alpha}_i^*, i = 1, \ldots, N$.

5.2.2.3 *Global VAR (GVAR)*

To obtain efficient estimates of $\Phi_i(L)$ using Zellner's (1962) seemingly unrelated regression approach requires T to be considerably larger than N. In many macroeconomic applications, the number of time series observations, T, could be of the same magnitude as the number of cross-sectional dimensions, N. When N is large, it is not feasible to stack all Nm equations together as a system. Pesaran, Schuermann, and Weiner (2004) propose a global VAR to accommodate dynamic cross-dependence by considering

$$\Phi_i(L)(\boldsymbol{w}_{it} - \Gamma_i\boldsymbol{w}_{it}^*) = \boldsymbol{\epsilon}_{it}, \quad i = 1, 2, \ldots, N, \tag{5.2.42}$$

where

$$\boldsymbol{w}_{it}^* = \sum_{j=1}^{N} r_{ij}\boldsymbol{w}_{jt}, \tag{5.2.43}$$

$$r_{ii} = 0, \sum_{j=1}^{N} r_{ij} = 1, \text{ and } \sum_{j=1}^{N} r_{ij}^2 \to 0 \text{ as } N \to \infty. \tag{5.2.44}$$

The weight r_{ij} could be $\frac{1}{N-1}$ for $i \neq j$, or constructed from trade value or other measures of some economic distance, and could be time-varying. Just like the cross-sectionally mean augment regression approach discussed in Chapter 11, the global average $\boldsymbol{w}_{it}^*$ is inserted into (5.2.39) to take account of the cross-sectional dependence. When $\boldsymbol{w}_{i,t-s}^*$ can be treated as weakly exogenous (predetermined), the estimation of (5.2.42) for each i can proceed using standard time series estimation techniques (e.g. Pesaran, Shin, and Smith 2000). Pesaran et al. (2004) show that the weak exogeneity assumption of $\boldsymbol{w}_{it}^*$ holds for all countries except for the U.S. because of its dominant position in the world. They also show that (5.2.42) yields better results than (5.2.39) when cross-sectional units are correlated.[5]

[5] The computer program for GVAR and some data sources can be downloaded from www.econ.cam.ac.uk/people-files/emeritus/mhp1/GVAR/GVAR.html.

5.3 COINTEGRATED PANEL MODELS AND VECTOR ERROR CORRECTION

5.3.1 Properties of Cointegrated Processes

Many macro and financial data are nonstationary. The nonstationarity of a time series is usually represented by an integrated process of order $d, d \geq 1, I(d)$. A dth integrated process can be transformed into a stationary process by differencing the variable d times, $(1-L)^d$. (e.g., Box and Jenkins 1970). For instance, suppose all the elements of $\boldsymbol{w}_{it}$ are $I(1)$; then $(1-L)\boldsymbol{w}_{it}$ become stationary ($I(0)$ processes). However, differencing $\boldsymbol{w}_{it}$ also removes the underlying long-run relations among the elements of $\boldsymbol{w}_{it}$, which can have important economic implications. If $\boldsymbol{w}_{it}$ are driven by some common nonstationary variables, one notable feature is that linear combinations of $\boldsymbol{w}_{it}$ can remove these *common trends* and become stationary ($I(0)$). Such linear combinations capture the long-run relations among $\boldsymbol{w}_{it}$ and are called "cointegrating" relations.

Let $\boldsymbol{w}_{it}$ be an $m \times 1$ vector of random variables. We assume that each element of $\boldsymbol{w}_{it}, w_{jit}$, is integrated of order $1, I(1)$,

$$w_{jit} \sim I(1), \quad j = 1, \ldots, m. \tag{5.3.1}$$

Following Engle and Granger (1987), we say that the elements of $\boldsymbol{w}_{it}$ forming $r(\geq 1)$ cointegrating relations if there exist r linearly independent combinations of $\boldsymbol{w}_{it}$ that are stationary $I(0)$,

$$C_i \boldsymbol{w}_{it} = \boldsymbol{u}_{it} \sim I(0), \tag{5.3.2}$$

where C_i denotes the $r \times m$ constant matrix with rank $(C_i) = r$, and $\boldsymbol{u}_{it}$ denotes the $r \times 1$ random vectors with $E(\boldsymbol{u}_{it}) = 0, E(\boldsymbol{u}_{it}\boldsymbol{u}'_{it}) = \tilde{\Omega}_{io}, E(\boldsymbol{u}_{it}\boldsymbol{u}'_{i,t-s}) = \tilde{\Omega}_{is}$, and rank $(\tilde{\Omega}_{io}) = r$.

Rewriting $\boldsymbol{w}_{it}$ as the sum of the impact of $(m-r)$ $I(1)$ common trends $\boldsymbol{z}_{it}$ and stationary components, $\boldsymbol{\xi}_{it}$,

$$\boldsymbol{w}_{it} = A_i \boldsymbol{z}_{it} + \boldsymbol{\xi}_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T. \tag{5.3.3}$$

where

$$\underset{(m-r)\times 1}{\boldsymbol{z}_{it}} \sim I(1), \tag{5.3.4}$$

and $A_i = (\boldsymbol{a}'_i)$ is an $m \times (m-r)$ constant matrix. The cointegrating relations (5.3.2) imply that

$$C_i A_i = \boldsymbol{0}, \quad i = 1, \ldots, N. \tag{5.3.5}$$

If $A_i \neq \boldsymbol{0}$ with rank $(A_i) = m - r, 0 < r < m$ and $\boldsymbol{z}_{it} \neq \boldsymbol{z}_{jt}$, the $m \times 1$ $I(1)$ random variables $\boldsymbol{w}_{it}$ are cointegrated. There exists an error correction representation (VEC) of $\boldsymbol{w}_{it}$ (Engle and Granger 1987),

$$\Delta \boldsymbol{w}_{it} = \Pi_i^* \boldsymbol{w}_{i,t-1} + \sum_{s=1}^{p_i} \Phi_{is}^* \Delta \boldsymbol{w}_{i,t-s} + \boldsymbol{\epsilon}_{it}, i = 1, \ldots, N, \tag{5.3.6}$$

where $\Delta = (1 - L)$ denotes the first difference operation, $\Delta \boldsymbol{w}_{it} = \boldsymbol{w}_{it} - \boldsymbol{w}_{i,t-1}, \boldsymbol{\epsilon}_{it}$ is independent, identically distributed with mean $\mathbf{0}$ and covariance matrix Ω_{ii}, $\Pi_i^* = -(I_m - \sum_{j=1}^{p_i} \Phi_{ij})$, $\Phi_{ij}^* = \left(-\sum_{s=j+1}^{p_i} \Phi_{is}\right)$, $j = 1, 2, \ldots, p_i$, and

$$\text{rank } (\Pi_i^*) = r. \tag{5.3.7}$$

Decompose the $m \times m$ matrix Π_i^* into the product of two rank $r(m \times r)$ matrices, J_i and Λ_i:

$$\Pi_i^* = J_i \Lambda_i', \quad i = 1, \ldots, m. \tag{5.3.8}$$

The decomposition is not unique; $\Pi_i^* = J_i^* \Lambda_i^{*'}$, where $J_i^* = J_i F_i, \Lambda_i^* = \Lambda_i F_i^{'-1}$, for any $r \times r$ nonsingular matrix F_i. To uniquely define the cointegrating relations, we choose the normalization

$$\Lambda_i' = [I_r, \tilde{\Lambda}_i'], \tag{5.3.9}$$

where $\tilde{\Lambda}_i$ is an $(m-r) \times r$ constant matrix. The advantage of considering an error correction representation (5.3.6) rather than PVAR is that one can simultaneously consider the long-run (equilibrium) relations and short-run dynamics of $\boldsymbol{w}_{it}$.

If the stationary components $\boldsymbol{\xi}_{it}$ are cross-sectionally dependent, we can decompose $\boldsymbol{\xi}_{it}$ into the impact of q common factors $\boldsymbol{f}_t$ that affect all cross-sectional units $\boldsymbol{\xi}_{it}$ and the idiosyncratic components $\boldsymbol{\epsilon}_{it}^*$,

$$\boldsymbol{\xi}_{it} = B_i \boldsymbol{f}_t + \boldsymbol{\epsilon}_{it}^*, \quad i = 1, \ldots, N, t = 1, \ldots T, \tag{5.3.10}$$

where B_i is an $m \times q$ constant matrix, $E(\mathbf{f}_t) = \mathbf{0}$, $E(\boldsymbol{f}_t \boldsymbol{f}_t')$ is normalized to be a q-rowed identity matrix, $E(\boldsymbol{\epsilon}_{it}^*) = \mathbf{0}, E(\boldsymbol{\epsilon}_{it}^* \boldsymbol{\epsilon}_{it}^{*'}) = D_i$ is a diagonal matrix.

If the nonstationarity of $\boldsymbol{w}_{it}, i = 1, \ldots, N$, is driven by the same common trends across i, i.e. if $z_{it} = z_{jt} = z_t$, it also implies that each element of $\boldsymbol{w}_{it}$ is cointegrated across cross-sectional units. Let $\boldsymbol{w}_{kt}$ denote the $N \times 1$ vector of the kth element of $\boldsymbol{w}_{it}, w_{kit}, (w_{k1t}, \ldots, w_{kNt})' = \boldsymbol{w}_{kt}$, and $\boldsymbol{\xi}_{kt} = (\xi_{k1t}, \ldots, \xi_{kNt})', \tilde{\boldsymbol{\epsilon}}_{kt} = (\tilde{\epsilon}_{k1t}, \ldots, \tilde{\epsilon}_{kNt})'$. Then

$$\boldsymbol{w}_{kt} = A_k z_t + \tilde{\boldsymbol{\xi}}_{kt}, k = 1, \ldots, m, \tag{5.3.11}$$

where $A_k = (\boldsymbol{a}_{ki}')$ denotes the $N \times (m - r)$ constant matrix of the cross-sectional stacked kth row of A_i. Suppose rank $(A_k) = d_k (\leq m - r)$, there also exists an $(m - d_k) \times m$ matrix C_k with rank $(m - d_k)$ such that

$$C_k A_k = \mathbf{0}, \quad k = 1, \ldots, m. \tag{5.3.12}$$

Then

$$C_k \boldsymbol{w}_{kt} = C_k \tilde{\boldsymbol{\xi}}_{kt} \sim I(0). \tag{5.3.13}$$

In other words, with nonstationary panel data, there could be cointegration relations both over time and across individuals.

5.3.2 Estimation

5.3.2.1 "Homogeneous" Cointegrating Relations

In the case when there is at most one cointegration relation among $I(1)$ $\boldsymbol{w}_{it}, i = 1, \ldots, N$, the "homogeneous" cointegration vector between $(w_{1it}, \tilde{\boldsymbol{w}}_{it}'), (1, \boldsymbol{\beta}')$, with individual-specific effects α_{1i} yields

$$w_{1it} = \tilde{\boldsymbol{w}}_{it}'\boldsymbol{\beta} + \alpha_{1i} + u_{1it}, \tag{5.3.14}$$

where u_{1it} is stationary but independently distributed across i. The cointegrating vector $\boldsymbol{\beta}$ can be estimated by the within estimator,

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)'\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)(w_{1it} - \bar{w}_{1i})\right], \tag{5.3.15}$$

where $\bar{\boldsymbol{w}}_i' = (\bar{w}_{1i}, \bar{\tilde{\boldsymbol{w}}}_i') = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{w}_{it}'$. The least squares estimator converges to $\boldsymbol{\beta}$ in the speed of $T\sqrt{N}$ and is asymptotically normally distributed. (The fast speed of convergence is due to $\tilde{\boldsymbol{w}}_{it}$ is I(1)). However, the exogeneity and unit-root of $\tilde{\boldsymbol{w}}_{it}$ leads to a nonzero asymptotic bias term of order $\frac{1}{T}$ when $\hat{\boldsymbol{\beta}}$ is multiplied by the scale factor, $T\sqrt{N}$ (Kao and Chiang 2000). The idea of Phillips and Hansen (1990) fully modified estimator can be applied to correct the exogeneity effect,

$$w_{1it}^{+} = \tilde{\boldsymbol{w}}_{it}'\boldsymbol{\beta} + \alpha_{1i} + u_{1it}^{+} \tag{5.3.16}$$

where

$$w_{1it}^{+} = w_{1it} - \Omega_{u\Delta\tilde{w}}\Omega_{\Delta\tilde{w}}^{-1}\Delta\tilde{\boldsymbol{w}}_{it}, \tag{5.3.17}$$

$$u_{1it}^{+} = u_{1it} - \Omega_{u\Delta\tilde{w}}\Omega_{\Delta\tilde{w}}^{-1}\Delta\tilde{\boldsymbol{w}}_{it}, \tag{5.3.18}$$

and $\Omega_{u\Delta\tilde{w}} = \sum_{j=-\infty}^{\infty} E(u_{it}\Delta\tilde{\boldsymbol{w}}_{i,t-j}'), \Omega_{\Delta\tilde{w}} = \sum_{j=-\infty}^{\infty} E(\Delta\tilde{\boldsymbol{w}}_{it}\Delta\tilde{\boldsymbol{w}}_{i,t-j}')$, are the long-run covariance matrix of u_{1it} and $\Delta\tilde{\boldsymbol{w}}_{it}$, and $\Delta\tilde{\boldsymbol{w}}_{it}$, respectively. The panel fully modified within the estimator takes the form,

$$\hat{\boldsymbol{\beta}}_{FM} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)'\right]^{-1} \left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}(\tilde{\boldsymbol{w}}_{it} - \bar{\tilde{\boldsymbol{w}}}_i)w_{1it}^{+} - T\Delta_{\Delta\tilde{w}u}\right)\right], \tag{5.3.19}$$

where $\Delta_{\Delta\tilde{w}u} = \sum_{j=0}^{\infty} E(\Delta\tilde{\boldsymbol{w}}_{i,t-j}u_{1it})$. The correction terms $\Omega_{u\Delta\tilde{w}}, \Omega_{\Delta\tilde{w}}, \Delta_{\Delta\tilde{w}u}$ can be replaced by their consistent estimator.[6] Kao and Chiang (2000) show that $T\sqrt{N}(\hat{\boldsymbol{\beta}}_{FM} - \boldsymbol{\beta})$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix $2\sigma_{u|\Delta\tilde{w}}^2\Omega_{\Delta\tilde{w}}^{-1}$, where $\sigma_{u|\Delta\tilde{w}}^2 = \Omega_u - \Omega_{u\Delta\tilde{w}}\Omega_{\Delta\tilde{w}}^{-1}\Omega_{\Delta\tilde{w}u}, \Omega_u = \sum_{j=-\infty}^{\infty} E(u_{1it}u_{1i,t-j})$.

An alternative approach is to apply the dynamic within estimator to the lead-lag adjusted regression model (Saikkonen 1991),

$$w_{1it} = \tilde{\boldsymbol{w}}_{it}'\boldsymbol{\beta} + \sum_{j=-q}^{q}\Delta\tilde{\boldsymbol{w}}_{i,t-j}'\boldsymbol{\gamma}_j + \alpha_{1i} + \tilde{u}_{1it}. \tag{5.3.20}$$

where Westerlund (2005) suggests a data-based choice of truncation lag order q. Kao and Chiang (2000) show that the within estimator of (5.3.20) has the same asymptotic

[6] If cross-sectional units have heteroscedastic long-run variance, Kao and Chiang (2000) suggest using the cross-sectional average in the long-run correction terms.

distribution as the panel fully modified within estimator. The Monte Carlo studies conduced by Kao and Chiang (2000) show that the (lead-lag adjusted) within estimator of (5.3.20) performs better than the panel fully modified estimator, probably because of the failure of obtaining good estimates of $\Omega_{u\Delta\tilde{\boldsymbol{w}}}, \Omega_{\Delta\tilde{w}}$, etc.

When there are more than one linearly independent cointegration relation among the elements of $\boldsymbol{w}_{it}$, in principle, one can normalize the r linearly independent cointegration relation in the form of (5.3.9), $\Lambda' = [I_r\ \tilde{\Lambda}']$, then apply the fully modified within estimator or the lead-lag adjusted within estimator equation by equation.

Alternatively, one can use the methods of estimating the fixed effects PVAR model discussed in Section 5.2 to obtain the consistent estimator of Φ_j's, then solving $\hat{\Lambda}$ from the estimated $\hat{\Phi}_j$'s. The error-correction representation of cointegrated system (5.2.1) gives $\Pi = -(I_m - \sum_{j=1}^{P} \Phi_j)$ (5.2.11). If rank $(\Pi) = r(> 1)$, then we can write $\Pi = J\Lambda'$ where J and Λ are $m \times r$ with rank r. Then the cointegrating matrix Λ' is equal to

$$\Lambda' = (J'J)^{-1}\Gamma'\Pi. \tag{5.3.21}$$

If we choose the normalization $\Lambda' = [I_r\ \tilde{\Lambda}']$, then $J = \Pi_1$ and

$$\tilde{\Lambda}' = (\Pi_1'\Pi_1)^{-1}\Pi_1'\Pi \tag{5.3.22}$$

where Π_1 is the $m \times r$ matrix consisting of the first r columns of Π.

Although the above procedure is consistent, it is not efficient because the first-stage estimator of Φ's has not taken into account the reduced rank restrictions on $\Pi = J\Lambda'$. One way to obtain efficient estimator of Φ_j's is to apply constrained GMM or constrained MDE or constrained (transformed) MLE by minimizing the quadratic form of the moment conditions (5.2.15) or (5.2.30) or maximizing (5.2.28) subject to

$$\Lambda = H\boldsymbol{\delta}_r + B_r, \tag{5.3.23}$$

where H and B_r are, respectively, $m \times (m-r)$ and $m \times r$ matrices with known elements, and $\boldsymbol{\delta}$ is a $(m-r) \times r$ matrix with unknown coefficients. In the case when $\Lambda' = [I_r\ \tilde{\Lambda}']$, $H = [0, I_{m-r}], \boldsymbol{\delta}_r = \tilde{\Lambda}, B_r = [I_r, \mathbf{0}]'$.

5.3.2.2 *Heterogeneous Cointegrating System*

If individual units are independent across i ($B_i = \mathbf{0}$ in (5.3.10)), the Johansen (1991) method can be applied to the time series data for each i to obtain the maximum likelihood estimator (MLE) of (5.3.6). There is no need for pooling.

When $B_i \neq \mathbf{0}$, individual units are correlated. If N is fixed and T is large, the covariance between $\boldsymbol{\epsilon}_{it}$ and $\boldsymbol{\epsilon}_{jt}, \Omega_{ij}$, can simply be estimated by

$$\hat{\Omega}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{\epsilon}}_{it}\hat{\boldsymbol{\epsilon}}_{jt}', \tag{5.3.24}$$

where $\hat{\boldsymbol{\epsilon}}_{it}$ can be constructed from (5.3.6) using the Johansen MLE estimates of $\boldsymbol{\theta}_i' =$ vec $(\Pi_i^*, \Phi_{i1}^*, \ldots, \Phi_{ip_i}^*)'$, where vec $(\cdot)$ denotes the operator that stacks the columns of a matrix successively into a vector. Therefore, we shall be concerned not with the estimation of B_i, but only with the inference of $\boldsymbol{\theta}_i$.

For ease of exposition, we shall first assume that all $\Phi_{is} \equiv \mathbf{0}$, for $s \geq 2$, then

$$\Delta\boldsymbol{w}_{it} = \Pi_i^*\boldsymbol{w}_{i,t-1} + \boldsymbol{\epsilon}_{it}, i = 1, \ldots, N. \tag{5.3.25}$$

Let $\boldsymbol{w}_t = (\boldsymbol{w}'_{1t}, \ldots, \boldsymbol{w}'_{Nt})'$ be the $Nm \times 1$ vector that stacks the N cross-sectionally observed $\boldsymbol{w}_{it}$ one after another. Then (5.3.25) can be written as

$$\Delta \boldsymbol{w}_t = \tilde{\Pi} \boldsymbol{w}_{t-1} + \boldsymbol{\epsilon}_t, \tag{5.3.26}$$

where

$$\tilde{\Pi} = \begin{pmatrix} \Pi_1^* & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \Pi_2^* & . & . & \mathbf{0} \\ . & & . & & \\ . & & & . & \\ \mathbf{0} & & & & \Pi_N^* \end{pmatrix}, \tag{5.3.27}$$

and

$$\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}'_{1t}, \ldots, \boldsymbol{\epsilon}'_{Nt})'.$$

Under the assumption that $\boldsymbol{\epsilon}_t$ is independently normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\underset{Nm \times Nm}{\tilde{\Omega}} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & . & . & \Omega_{1N} \\ . & \Omega_{22} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \Omega_{N1} & . & . & . & \Omega_{NN} \end{pmatrix}, \tag{5.3.28}$$

the log-likelihood function of $\Delta \boldsymbol{w}_t$ is proportional to

$$-\frac{T}{2} \log \mid \tilde{\Omega} \mid - \frac{1}{2} tr \left[\tilde{\Omega}^{-1} \sum_{t=1}^{T} (\Delta \boldsymbol{w}_t - \tilde{\Pi} \boldsymbol{w}_{t-1})(\Delta \boldsymbol{w}_t - \tilde{\Pi} \boldsymbol{w}_{t-1})' \right]. \tag{5.3.29}$$

The MLE of $\tilde{\Omega}$ and $\tilde{\Pi}$ are the solutions that simultaneously satisfy

$$\hat{\tilde{\Omega}} = \frac{1}{T} \sum_{t=1}^{T} (\Delta \boldsymbol{w}_t - \hat{\tilde{\Pi}} \boldsymbol{w}_{t-1})(\Delta \boldsymbol{w}_t - \hat{\tilde{\Pi}} \boldsymbol{w}_{t-1})', \tag{5.3.30}$$

and

$$\begin{pmatrix} \text{vec } (\hat{\Pi}_1^{*\prime}) \\ \text{vec } (\hat{\Pi}_2^{*\prime}) \\ . \\ . \\ \text{vec } (\hat{\Pi}_N^{*\prime}) \end{pmatrix} = \left(\sum_{t=1}^{T} \hat{\tilde{\Omega}}^{-1} \otimes \boldsymbol{w}_{t-1} \boldsymbol{w}'_{t-1} \right)^{-1} \left[\sum_{t=1}^{T} \left(\hat{\tilde{\Omega}}^{-1} \otimes \boldsymbol{w}_{t-1} \right) \text{ vec } (\Delta \boldsymbol{w}_t) \right]. \tag{5.3.31}$$

Conditional on $\hat{\tilde{\Omega}}$, (5.3.31) is in the form of Zellner's (1962) seemingly unrelated regression estimator. Sequentially iterating between (5.3.30) and (5.3.31) until convergence will result in the MLE of Π_i^* and $\tilde{\Omega}$.

However, if $\boldsymbol{w}_{it}$ are cointegrated, then Π_i^* is subject to the restrictions of the form (5.3.8). Substituting (5.3.8) into (5.3.9) and making use of the relations (Magnus and Neudecker 1999, p. 31),

$$\begin{aligned} \text{vec } (J_i \Lambda'_i) &= (I_r \otimes J_i) \text{ vec } (\Lambda'_i) \\ &= (\Lambda_i \otimes I_m) \text{ vec } (J_i), \end{aligned} \tag{5.3.32}$$

we obtain the MLE of $(\text{vec}\,(\Lambda_1')', \ldots, \text{vec}\,(\Lambda_N')')'$ conditional on $\hat{\hat{\Omega}}$ and $\hat{J}_i, i = 1, \ldots, N$,

$$\begin{pmatrix} \text{vec}\,(\hat{\Lambda}_1) \\ \cdot \\ \cdot \\ \text{vec}\,(\hat{\Lambda}_N) \end{pmatrix} = \left\{ \left[\hat{Q}' \left[\hat{\hat{\Omega}}^{-1} \otimes \left(\sum_{t=1}^{T} \boldsymbol{w}_{t-1} \boldsymbol{w}_{t-1}' \right) \right] \hat{Q} \right]^{-1} \right\}^{-1}$$

$$\left\{ \hat{Q}' \left[\hat{\hat{\Omega}}^{-1} \otimes I_{Nm} \right] \text{vec} \left(\sum_{t=1}^{T} \boldsymbol{w}_{t-1} \Delta \boldsymbol{w}_t' \right) \right\}, \tag{5.3.33}$$

and the MLE of $(\text{vec}\,(\hat{J}_1')', \ldots, \text{vec}\,(\hat{J}_N')')'$ conditional on $\hat{\hat{\Omega}}$ and $\hat{\Lambda}_i, i = 1, \ldots, N$,

$$\begin{pmatrix} \text{vec}\,(\hat{J}_1') \\ \cdot \\ \cdot \\ \text{vec}\,(\hat{J}_N)' \end{pmatrix} = \left\{ \hat{P}' \left(\hat{\hat{\Omega}}^{-1} \otimes \sum_{t=1}^{T} \boldsymbol{w}_{t-1} \boldsymbol{w}_{t-1}' \right) \hat{P} \right\}^{-1}$$

$$\cdot \left\{ \hat{P}' \left(\hat{\hat{\Omega}}^{-1} \otimes I_{Nm} \right) \text{vec} \left(\sum_{t=1}^{T} \boldsymbol{w}_{t-1} \Delta \boldsymbol{w}_t' \right) \right\}, \tag{5.3.34}$$

where I_{Nm} denotes the $Nm \times Nm$ identity matrix,

$$\hat{Q} = \left[\left(\boldsymbol{d}_1 \otimes \hat{J}_1 \right) \otimes (\boldsymbol{d}_1 \otimes I_r), \ldots, \left(\boldsymbol{d}_N \otimes \hat{J}_N \right) \otimes (\boldsymbol{d}_N \otimes I_r) \right], \tag{5.3.35}$$

$$\hat{P} = \left[(\boldsymbol{d}_1 \otimes I_m) \otimes \left(\boldsymbol{d}_1 \otimes \hat{\Lambda}_1 \right), \ldots, (\boldsymbol{d}_N \otimes I_m) \otimes \left(\boldsymbol{d}_N \otimes \hat{\Lambda}_N \right) \right], \tag{5.3.36}$$

$\boldsymbol{d}_j$ denotes the jth column of I_N, and $\otimes$ denotes the Kronecker product. Therefore, Groen and Kleibergen (2003) suggest the following iterative scheme to obtain the panel MLE:

1. Construct initial estimates $\hat{\hat{\Omega}}^{(0)}$ and $\hat{J}_i^{(0)}, i = 1, \ldots, N$ from (5.3.30) and (5.3.31), where the initial estimates of $\hat{J}_i^{(0)}$ are simply the first r columns of $\hat{\Pi}_i^{*(0)}$ under the normalization (5.3.9).
2. Construct estimates of $\Lambda_i, i = 1, \ldots, N$ from (5.3.33).
3. Revise the estimate of Π_i^* from (5.3.8) and revise the estimate of $\tilde{\Omega}$ from the revised estimate of $\hat{\Pi}_i^*$ using the formula (5.3.30).
4. Revise the estimate of J_i given the revise estimator of Λ_i and $\tilde{\Omega}$ using (5.3.34).
5. Iterate steps 1 to 4 until the solution converges.

The MLE of $\Lambda_i, i = 1, \ldots, N$ converges to their true values at rate T and their limiting distributions are mixed normal. Therefore, conventional Wald-type test statistics on the null hypothesis of Λ_i is asymptotically chi-square distributed.

When the cointegrating matrix $\Lambda_i = \Lambda_j = \Lambda$, the common cointegrating matrix can be estimated by

$$\text{vec}\,(\Lambda) = \left(\hat{\hat{Q}}' \left(\hat{\hat{\Omega}}^{-1} \otimes \sum_{t=1}^{T} \boldsymbol{w}_{t-1} \boldsymbol{w}_{t-1}' \right) \hat{\hat{Q}} \right)^{-1}$$

$$\left(\hat{\hat{Q}}' \left(\hat{\hat{\Omega}} \otimes I_{Nm} \right) \text{vec} \left(\sum_{t=1}^{T} \boldsymbol{w}_{t-1} \Delta \boldsymbol{w}_t' \right) \right). \tag{5.3.37}$$

Because Λ_i or Λ is asymptotically mixed normally distributed, one can use the likelihood ratio test statistic to test the null hypothesis

$$\begin{aligned} & H_0 \ : \ \Lambda_1 = \Lambda_2 = \ldots = \Lambda_N = \Lambda \\ \text{versus} \quad & H_1 \ : \ \Lambda_i \neq \Lambda_j . \end{aligned}$$

The likelihood ratio statistic is asymptotically chi-square distributed with $(N-1)r(m-r)$ degrees of freedom when $T \longrightarrow \infty$.

When individual units contain an individual specific effects $\boldsymbol{\alpha}_i^*$, and p_i in (5.3.6) different from zero, one can follow Johansen (1991) to concentrate out constants and Φ_{is}^* by regressing $\Delta \boldsymbol{w}_{it}$ and $\boldsymbol{w}_{i,t-1}$ on the constants and $\Delta \boldsymbol{w}_{i,t-s}, s = 1, \ldots, p_i$ respectively, to obtain $\Delta \tilde{\boldsymbol{w}}_{it}$ and $\tilde{\boldsymbol{w}}_{i,t-1}$, then proceed to estimate Γ_i and Λ_i. Once the MLE of Γ_i, Λ_i, and Ω are obtained, the individual-specific constants and Φ_{ij}^* can then be obtained by regressing $\left(\Delta \tilde{\boldsymbol{w}}_{it} - \hat{J}_i \hat{\Lambda}_i' \tilde{\boldsymbol{w}}_{i,t-1}\right)$ on constants, $\Delta \tilde{\boldsymbol{w}}_{i,t-p_i-1}, \ldots, \Delta \tilde{\boldsymbol{w}}_{i,t-1}$ using Zellner's (1962) seemingly unrelated regression method.

When N is large, it is not feasible to use either the likelihood approach or the Zellner (1962) unrelated regression framework to estimate either the homogeneous or the heterogeneous cointegration relations. Pesaran, Schuermann, and Weiner (2004) suggest using a global VAR (GVAR) (Equations 5.2.42–5.2.44) to filter out cross-sectional dependence. When w_{it}^* can be treated as weakly exogenous, the system (5.2.42) subject to the rank condition (5.3.8) for each i can be estimated using standard time series estimator techniques (e.g. Pesaran, Shin, and Smith 2000).

5.4 UNIT ROOT AND COINTEGRATION TESTS

5.4.1 Unit Root Tests

Panels have been used to analyze regional growth convergence (e.g. Bernard and Jones 1996), exchange rate determination (e.g., testing of purchasing power parity hypothesis; Frankel and Rose 1996), business cycle synchronization, etc. In such analysis, the time series property of a variable is of significant interest to economists. The statistical properties of time series estimators also depend on whether the data is stationary or nonstationary. In the case of inference based on time series (i.e., $N = 1$), the limiting distributions of most estimators will be approximately normal when $T \longrightarrow \infty$ if the variables are stationary. Standard normal, chi-square tables can be used to construct confidence intervals or test hypotheses. If the data is nonstationary or contains unit roots, standard estimators will have nonstandard distributions as $T \longrightarrow \infty$. The conventional Wald-type test statistics can not be approximated well by t- or chi-square distributions (e.g., Dickey and Fuller 1979, 1981; Phillips and Durlauf 1986). Computer simulations will have to be used to find the critical values under the null. In cases where N is fixed and T is large, standard time series techniques can be used and the panel aspect of the data does not pose new techniques. If N is large, panel data provides the possibility of invoking a version of the central limit theorem along the cross-section dimension. Hence, contrary to the time series literature, the null hypothesis of panel unit root tests can be either unit root or stationarity. However, since most applications of panel unit root tests still follow the convention that the null of a time series of y_{it} contain a unit root, we shall therefore discuss only tests based on the null of unit root. For reference of tests under the null of stationarity, see, e.g., Kwiatkowski et al. (1992), Hadri (2000), and Hadri and Larsson (2005).

5.4.1.1 Cross-Sectional Independent Data

Since Quah (1994), many people have suggested panel unit root test statistics when N and T are large. When cross-sectional units are independent, following Dickey and Fuller (1979, 1981), Levin and Lin (1993) (LL), and Levin, Lin, and Chu, (LLC) (2002), consider a panel extension of the null hypothesis that each individual time series in the panel contains a unit root against the alternative hypothesis that all individual series are stationary by considering the model

$$\Delta y_{it} = \alpha_i + \delta_i t + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{5.4.1}$$

where ϵ_{it} is assumed to be independently distributed across i and Δ denotes the first difference operator, $1 - L$, with L being the lag operator that shifts the observation by one period, $Ly_{it} = y_{i,t-1}$. If $\gamma_i = 0$, then y_{it} contains a unit root. If $\gamma_i < 0$, y_{it} is stationary. Levin and Lin (1993) specify the null hypothesis as

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_N = 0, \tag{5.4.2}$$

and the alternative hypothesis as

$$H_1 : \gamma_1 = \gamma_2 = \ldots = \gamma_N = \gamma < 0. \tag{5.4.3}$$

To test H_0 against H_1, Levin and Lin (1993) suggest taking out the impact of variables in (5.4.1) that are not directly relevant to the estimation of γ by first regressing Δy_{it} and $y_{i,t-1}$ on the remaining variables in (5.4.1) for each i, providing the residuals $\hat{e}_{it}$ and $\hat{v}_{i,t-1}$, respectively. Then estimate γ by running the regression of the following model

$$\hat{e}_{it} = \gamma \hat{v}_{i,t-1} + \epsilon_{it}. \tag{5.4.4}$$

To adjust for heteroscedasticity across i in (5.4.4), they suggest using the least squares estimate of γ, $\hat{\gamma}$, to compute the variance of $\hat{e}_{it}$,

$$\hat{\sigma}_{ei}^2 = (T - p_i - 1)^{-1} \sum_{t=p_i+2}^{T} (\hat{e}_{it} - \hat{\gamma}\hat{v}_{i,t-1})^2. \tag{5.4.5}$$

Then divide (5.4.4) by $\hat{\sigma}_{ei}$ for each i, to obtain the heteroscedasticity-adjusted model

$$\tilde{e}_{it} = \gamma \tilde{v}_{i,t-1} + \tilde{\epsilon}_{it}, \tag{5.4.6}$$

where $\tilde{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{ei}}, \tilde{\epsilon}_{it} = \frac{\epsilon_{it}}{\hat{\sigma}_{ei}}, \tilde{v}_{i,t-1} = \frac{\hat{v}_{i,t-1}}{\hat{\sigma}_{ei}}$.

The t-statistic for testing $\gamma = 0$ is

$$t_{\tilde{\gamma}} = \frac{\tilde{\gamma}}{sd_{\tilde{\gamma}}}, \tag{5.4.7}$$

where $\tilde{\gamma}$ is the least squares estimates of (5.4.6),

$$sd_{\tilde{\gamma}} = \hat{\sigma}_{\epsilon} \left[\sum_{i=1}^{N} \sum_{t=p_i+2}^{T} \tilde{v}_{i,t-1}^2 \right]^{-1/2}$$

$$\hat{\sigma}_{\epsilon}^2 = (N\tilde{T})^{-1} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} (\tilde{e}_{it} - \tilde{\gamma}\tilde{v}_{i,t-1})^2,$$

$$\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i, \tilde{T} = (T - \bar{p} - 1).$$

Although regressing e_{it} on v_{it} over t for a given i leads to a random variable with nonstandard distributions as $T \longrightarrow \infty$ (e.g., Phillips and Durlauf 1986), averaging over i allows the invocation of the central limit theorem across cross-sectional dimension when N is large. However, $t_{\tilde{\gamma}}$ is not centered at 0. To correct the asymptotic bias, Levin and Lin (1993) suggest adjusting (5.4.7) by

$$t^* = \frac{t_{\tilde{\gamma}} - N\tilde{T}S_{N,T}\hat{\sigma}_{\epsilon}^{-2} \cdot sd_{\tilde{\gamma}} \cdot \mu_{\tilde{T}}}{\sigma_{\tilde{T}}}, \tag{5.4.8}$$

where

$$S_{NT} = N^{-1}\sum_{i=1}^{N} \frac{\hat{\omega}_{yi}}{\hat{\sigma}_{ei}}, \tag{5.4.9}$$

and $\hat{\omega}_{yi}^2$ is an estimate of the long-run variance of y_i, say,

$$\hat{\omega}_{yi}^2 = (T-1)^{-1}\sum_{t=2}^{T} \Delta y_{it}^2 + 2\sum_{L=1}^{\bar{K}} W_{\bar{K}}(L) \left((T-1)^{-1} \sum_{t=L+2}^{T} \Delta y_{it} \Delta y_{i,t-L} \right), \tag{5.4.10}$$

where $W_{\bar{K}}(L)$ is the lag kernel to ensure the positivity of $\hat{\omega}_{yi}^2$; for instance, Newey and West (1987) suggest that

$$W_{\bar{K}}(L) = \begin{cases} 1 - \frac{L}{T} & \text{if } L < \bar{K}, \\ 0 & \text{if } L \geq \bar{K}. \end{cases} \tag{5.4.11}$$

The $\mu_{\tilde{T}}$ and $\sigma_{\tilde{T}}$ are mean and standard deviation adjustment terms which Newey and West computed by Monte Carlo simulation and tabulated in their paper. They show that provided the augmented Dickey–Fuller (1981) lag order p increases at some rate T^p where $0 \leq p \leq 1/4$, and the lag truncation parameter $\bar{K}$ increases at rate T^q where $0 < q < 1$, the panel test statistic $t_{\tilde{\gamma}}$ under the null of $\gamma = 0$ converges to a standard normal distribution as both T and $N \rightarrow \infty$.

In the special case that $\alpha_i = \delta_i = \phi_{i\ell} = 0$, and ϵ_{it} is independently identically distributed with mean 0 and variance σ_{ϵ}^2, Levin and Lin (1993) and Levin, Lin, and Chu (2002) show that under the null of $\gamma = 0$, $T\sqrt{N}\hat{\gamma}$ of the pooled least squares estimator, $\hat{\gamma}$, converges to a normal distribution with mean 0 and variance 2 and the t-statistic of $\hat{\gamma}$ converges to a standard normal as $\sqrt{N}/T \rightarrow 0$ as $(N, T) \rightarrow \infty$ (i.e., the time dimension can expand more slowly than the cross-section). When T is fixed and $N \rightarrow \infty, \sqrt{N}\hat{\gamma}$ is asymptotically biased. For the correction of asymptotic bias of the t-statistics, see Harris and Tzavalis (1999).

Im, Pesaran, and Shin (2003) (IPS) relax the LL, LLC strong assumption of homogeneity for (5.4.1) under the alternative (i.e. allowing $\gamma_i \neq \gamma_j$) by postulating the alternative hypothesis as

$$H_A^* : \gamma_i < 0 \text{ for some } i, \text{ and } \frac{N_0}{N} = c > 0, \tag{5.4.12}$$

where N_0 denotes the number of cross-sectional units with $\gamma_i < 0$.[7] Thus, instead of pooling the data, Im, Pesaran, and Shin suggest taking the average of separate unit root tests for N individual cross-section units of the augmented Dickey–Fuller (ADF; Dickey and Fuller 1981) t-ratios $\tau_i, \bar{\tau}$. Since τ_i are independent across i, $\bar{\tau}$ converges to a normal distribution under the null with mean $E(\bar{\tau})$ and variance, Var $(\bar{\tau}_N)$, as $T \longrightarrow \infty$ and $N \longrightarrow \infty$. However, the statistic $\bar{\tau}/\sqrt{\text{Var }(\bar{\tau})}$ is equivalent to multiplying each τ_i by $\sqrt{N}$. Although the limiting distribution of each ADF statistic as $T \longrightarrow \infty$ is well defined, multiplying τ_i by $\sqrt{N}$ introduces a nonnegligble bias term as N also goes to ∞. Therefore, IPS suggest using the statistics

$$Z = \frac{\sqrt{N}(\bar{\tau} - E(\tau_i))}{\sqrt{\text{Var }(\tau_i)}}. \tag{5.4.13}$$

Since $E(\tau_i)$ and $\text{Var}(\tau_i)$ will vary as the lag length in the ADF regression varies, IPS tabulate $E(\tau_i)$ and var (τ_i) for different lag lengths. They show in their Monte Carlo studies that their test is more powerful than the LLC test under certain cases. However, if the null is rejected, all we can say is that a fraction of cross-section units is stationary. It does not provide explicit guidance as to the size of this fraction or the identity of cross-sectional units that are stationary.

Alternatively, Maddala and Wu (1999) (MW) and Choi (2001) suggest using the Fisher (1932) P_λ to test the null (5.4.2) against (5.4.12) by combining the evidence from several independent tests. The idea is as follows: Suppose there are N unit root tests as in Im, Pesaran, and Shin. Let P_i be the observed significance level (P-value) for the ith test. Then

$$P_\lambda = -2 \sum_{i=1}^{N} \log P_i, \tag{5.4.14}$$

has a chi-square distribution with $2N$ degrees of freedom as $T_i \rightarrow \infty$ (Rao 1952, p. 44). When N is large, Choi (2001) proposes a modified P_λ test,

$$P_m = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (-2 \log P_i - 2)}{2}, \tag{5.4.15}$$

because $E(-2 \log P_i) = 2$ and Var $(-2 \log P_i) = 4$. Using a sequential limit argument ($T_i \rightarrow \infty$ followed by $N \rightarrow \infty$), Choi (2001) shows that the P_m test is asymptotically normally distributed with mean 0 and variance 1.

The LL test is based on homogeneity of the autoregressive parameter (although it allows heterogeneity in the error variances and the serial correlation structure of the errors). Thus, the test is based on pooled regressions. On the other hand, both the MW (or Choi) test or the IPS test are based on the heterogeneity of the autoregressive parameter under the

[7] Technically speaking, the alternative can be formulated for at least one i, $\gamma_i < 0$. However, the test will have no power if $c \rightarrow 0$ as $N \rightarrow \infty$.

alternative. The tests amount to a combination of different independent tests. However, they are all consistent tests against either alternative. The advantage of the MW (or Choi) test is that it does not require a balanced panel, nor the identical lag length in the individual ADF regressions. In fact, it can be carried out for any unit root test derived. It is nonparametric. Whatever test statistic we use for testing for a unit root for each individual unit, we can get the P-values, P_i. The disadvantage is that the P-values have to be derived by Monte Carlo simulation. On the other hand, the LL and the IPS tests are parametric. Although the distribution of the $t_{\tilde{\gamma}}$ or $\bar{\tau}$ statistic involves the adjustment of the mean and variance, they are easy to use because ready tables are available from their papers. However, these tables are valid only for the ADF test.

The heterogeneity of panel data introduces an asymmetry in the way the null and the alternative hypotheses are treated, which is not present in the univariate time series models. This is because the same null hypothesis is imposed across i, but the alternative could be either homogeneous across i or be allowed to vary with i. The IPS and MW tests allow the alternative to be heterogeneous. The drawback of their approach is that the model is overly parameterized, in the sense that the information contained in the unit-specific t-statistics is not used in an efficient way. The tests will have power only if $\frac{\sqrt{N}}{T} = c < \infty$. Since the interest is only in whether the null holds or not, Westerlund and Larsson (2012) consider a random specification of $\rho_i = 1 + \gamma_i$ where ρ_i is assumed independently identically distributed with mean μ_ρ and variance σ_ρ^2. Under the null of unit root,

$$H_0 : \mu_\rho = 1 \text{ and } \sigma_\rho^2 = 0 \tag{5.4.16}$$

The alternative is

$$\begin{aligned} H_1 : \mu_\rho &\neq 1 \\ &\text{or } \sigma_\rho^2 > 0 \\ &\text{or both.} \end{aligned} \tag{5.4.17}$$

The Lagrangian Multiplier (LM) test statistics of the null (5.4.16) takes the form

$$LM_{\mu_\rho, \sigma_\rho^2} = LM_{\mu_\rho | \sigma_\rho^2} + LM_{\sigma_\rho^2 | \mu_\rho}, \tag{5.4.18}$$

where

$$LM_{\mu_\rho | \sigma_\rho^2} = \frac{(A_{NT})^2}{B_{NT}} + \frac{1}{2} \frac{(C_{NT})^2}{D_{NT}}, \tag{5.4.19}$$

$$A_{NT} = \frac{1}{\sqrt{N}T} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta e_{it} v_{i,t-1}, \tag{5.4.20}$$

$$B_{NT} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{i,t-1}^2, \tag{5.4.21}$$

$$C_{NT} = \frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left[(\Delta e_{it})^2 - 1 \right] v_{i,t-1}^2, \tag{5.4.22}$$

$$D_{NT} = \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\Delta e_{it})^2 v_{i,t-1}^4. \tag{5.4.23}$$

The $LM_{\mu_\rho|\sigma_\rho^2}$ can be viewed as an LM test statistic for testing $\mu_\rho = 1$ versus $\mu_\rho \neq 1$ given $\sigma_\rho^2 = 0$. The $LM_{\sigma_\rho^2|\mu_\rho}$ is the LM test statistic for testing $\sigma_\rho^2 = 0$ versus $\sigma_\rho^2 > 0$ given $\mu_\rho = 1$. Westerlund and Larsson (2012) show that when both N and $T \to \infty$ and $\frac{T}{N} = N^{\theta-1}$ for $\theta > 1$, under the null (5.4.16), $LM_{\sigma_\rho^2|\mu_\rho}$ is chi-square distributed with one degree of freedom. The distribution of $LM_{\sigma_\rho^2|\mu_\rho}$ is asymptotically independent of the limiting distribution of $LM_{\mu_\rho|\sigma_\rho^2}$ and is asymptotically equal to $\frac{5(\kappa-1)}{24}$ times a chi-square degree 1 variable, where $\kappa = 1 + \frac{1}{T}\sum_{t=p+1}^{T}[E(\epsilon_{it}^4) - 1]$. If $\epsilon_{it} \sim N(0,1)$, then $\kappa = 3$ and $\frac{12}{5}LM_{\sigma_\rho^2|\mu_\rho}$ is asymptotically chi-square distributed with one degree of freedom.

When e_{it} and $v_{i,t-1}$ are unknown, one can construct the feasible LM test using $\hat{e}_{it}$ and $\hat{v}_{i,t-1}$ as in the LL test. The feasible LM is asymptotically equal to the LM test. (i.e., $\delta_i = 0$ in (5.4.1)).

Assuming the ρ_i as random has the advantage that the number of parameters needed to be estimated is reduced. Neither does it rule out that under the alternative, some of the units may be explosive. Moreover, by considering not only the mean of ρ_i, but also the variance, random coefficient formulation takes into account more information and hence is more powerful in detecting the alternative than the LL or IPS test.

In a large N or small T case, it only makes sense to discuss powerful unit root tests which are informative if either all cross-sectional units have the same dynamic response or a significant fraction of cross-sectional units reject the null under heterogeneity. However, identifying the exact proportion of the sample for which the null hypothesis is rejected requires T being very large (for further discussion, see Pesaran 2012).

5.4.1.2 Cross-Sectionally Correlated Data

When v_{it} are cross-sectionally dependent, the above tests are subject to severe size distortion (e.g., Banerjee, Marcellino, and Osbat 2005; Breitung and Das 2008). Chang (2002) suggests using some nonlinear transformation of the lagged level variable, $y_{i,t-1}$, $F(y_{i,t-1})$ as an instrument (IV) for $y_{i,t-1}$ for the usual Dickey–Fuller type regression (5.4.1). As long as $F(\cdot)$ is regularly integreable, say $F(y_{i,t-1}) = y_{i,t-1}e^{-c_i|y_{i,t-1}|}$, where c_i is a positive constant, the IV t-ratio,

$$Z_i = \frac{\hat{\gamma}_i - 1}{s(\hat{\gamma}_i)}, \tag{5.4.24}$$

will converge to a standard normal when $T_i \longrightarrow \infty$, where $s(\hat{\gamma}_i)$ is the standard error of the IV estimator $\hat{\gamma}_i$ and T_i denotes the time series observation of the ith unit because $F(y_{i,t-1})$ tends to zero and $y_{i,t-1}$ tends to $\pm\infty$, the nonstationarity of $y_{i,t-1}$ is eliminated. Moreover, as long as the product of $F(y_{i,t-1})$ and $F(y_{j,t-1})$ from different cross-sectional units i and j are asymptotically uncorrelated even though $y_{i,t-1}$ and $y_{j,t-1}$ are correlated, therefore the average IV t-ratio statistic

$$S_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i \tag{5.4.25}$$

possesses a standard normally limiting distribution as N also tends to infinity.

When y_{it} are correlated across cross-sectional units, Bai and Ng (2004, 2010), Moon and Perron (2004) consider that the cross-correlations are driven by some common factors,

$\boldsymbol{f}_t' = (f_{it}, \ldots, f_{kt})$ that vary over time. Rewrite y_{it} as the sum of the impact of the common factors, $\boldsymbol{b}_i' \boldsymbol{f}_t$ and the idiosyncratic component, u_{it},

$$y_{it} = \sum_{j=1}^{k} b_{ij} f_{jt} + u_{it}, \tag{5.4.26}$$

where $\boldsymbol{b}_i' = (b_{i1}, \ldots, b_{ik})$ is a $1 \times k$ vector of constants, u_{it} is independent across i with mean zero. If $\boldsymbol{b}_i = \boldsymbol{b}_j = \boldsymbol{b}$, then $\lambda_t = \boldsymbol{b}' \boldsymbol{f}_t$ is the common time-specific effects. The impact of common factors, $\boldsymbol{f}_t$, can be eliminated by subtracting the cross-sectional mean $y_t = \frac{1}{N} \sum_{i=1}^{N} y_{it}$ from y_{it}. If N is large, the demeaned series can be treated as if they are uncorrelated. We shall therefore consider the case that $\boldsymbol{b}_i \neq \boldsymbol{b}_j$.

Bai and Ng (2004, 2010) and Moon and Perron (2004) propose testing unit roots in both $\boldsymbol{f}_t$ and u_{it}. The factors are estimated from k principal components of Δy_{it} (e.g. Section 10.3). Augmented Dickey–Fuller tests (ADF) are then applied to test if $\boldsymbol{f}_t$ is integrated or not (I(1) versus I(0)). If it is found that the estimated factors contain unit roots and are not cointegrated, the N series are integrated of order 1. If the presence of a unit root in the factors is rejected, panel unit root tests such as the Maddala and Wu (1999) P_λ or the Choi (2001) P_m test is then applied to the defactored observations, $u_{it}, (y_{it} - \hat{\boldsymbol{b}}_i' \hat{\boldsymbol{f}}_t)$.[8]

If $\boldsymbol{b}_i \neq \boldsymbol{b}_j$ and $k = 1$, Pesaran (2007) suggests augmenting (5.4.1) by the cross sectional averaged values of $\bar{y}_{t-1} = \frac{1}{N} \sum_{i=1}^{N} y_{i,t-1}$ and $\Delta \bar{y}_{t-j} = \frac{1}{N} \sum_{i=1}^{N} \Delta y_{i,t-j}$,

$$\begin{aligned} \Delta y_{it} = {} & \alpha_i + \delta_i t + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} \\ & + c_i \bar{y}_{t-1} + \sum_{\ell=1}^{p_i} d_{i\ell} \Delta \bar{y}_{t-\ell} + e_{it}, \end{aligned} \tag{5.4.27}$$

to filter out the cross-sectional dependency, then taking the average of separate unit root tests for N individual cross-section units of the augmented Dickey–Fuller t-ratios, $\tau_i, \bar{\tau}$, as in Im, Pesaran, and Shin (2003).

When $k > 1$, if there exist $K (\geq k-1)$ observed time series $\boldsymbol{w}_{it}$, Kapetanios, Pesaran, and Yamagata (2011) and Pesaran, Smith, and Yamagata (2013) suggest further augmenting (5.4.1) by the cross-sectional mean of $\boldsymbol{w}_t$

$$\begin{aligned} \Delta y_{it} = {} & \alpha_i + \delta_i t + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} \\ & + \sum_{\ell=1}^{p_i} d_{i\ell} \Delta \bar{y}_{t-\ell} \\ & + W_t \boldsymbol{\theta} + \epsilon_{it}, \quad i = 1, \ldots, N \end{aligned} \tag{5.4.28}$$

where $W_t = (\Delta \boldsymbol{w}_t', \Delta \boldsymbol{w}_{t-1}', \ldots, \Delta \boldsymbol{w}_{t-p_i-1}', \boldsymbol{w}_{t-1}')$ and $\boldsymbol{\theta}$ is an $mp_i \times 1$ vector of constants. Pesaran, Smith, and Yamagata (2013) suggest either constructing an IPS-type test statistic by taking the average of t-statistics for $\gamma_i, i = 1, \ldots, N$ or taking the average of the Sargan and Bhargava test statistic (SB) (1983),

[8] This approach is called PANIC (panel analysis of nonstationarity in idiosyncratic and common components) by Bai and Ng (2004, 2010).

$$SB_i = T^{-2} \sum_{t=1}^{T} S_{it}^2 / \hat{\sigma}_i^2, \quad i = 1, \ldots, \tag{5.4.29}$$

where

$$S_{it} = \sum_{s=1}^{t} \hat{\epsilon}_{is}, \hat{\sigma}_i^2 = \frac{\sum_{t=1}^{T} \hat{\epsilon}_{it}^2}{T - K^*},$$

$\hat{\epsilon}_{it}$ is the least squares residual of (5.4.28), and K^* is the number of unknown constants in (5.4.28). Each of the SB_i converges to a function of Brownian motion which is independent of the factors as well as their loadings. Pearson, Smith, and Yamagata (2013) have provided the critical values of the augmented IPS and SB tests for $K = 0, 1, 2, 3$ and $p_i = 0, 1, \ldots, 4$ for $N, T = 20, 30, 50, 70, 100, 200$.

When the process of driving cross-section dependence is unknown, Choi and Chue (2007) suggest using a subsample testing procedure and grouping the sample into a number of overlapping blocks of b time periods. Using all $(T - b + 1)$ possible overlapping blocks, the critical value of a test statistic is estimated by the empirical distribution of the $(T - b + 1)$ test statistics computed. Although the null distribution of the test statistic may depend on the unknown nuisance parameters when $T \longrightarrow \infty$ and N is fixed, the subsample critical values will converge in probability to the true critical values. The Monte Carlo simulation conducted by Choi and Chue (2007) shows that the size of the subsample test is indeed robust against various forms of cross-section dependence.

All the unit root tests discussed above assume there are no structural breaks. Structural changes are likely to happen over long horizons. Failing to consider the presence of structural breaks may lead to misleading conclusions about the order of integration of a time series. For instance, it is well known that a stationary time series that evolves around a broken trend might be regarded as a nonstationary process (e.g., Perron 1989). For panel unit root tests allowing structural breaks with cross-sectional independent data, see, e.g., Im, Lee, and Tieslau (2005); with cross-sectional dependent data, see, e.g. Bai and Carrion-i-Silvestre (2009).

Nasreen et al. (2018) investigated the relationship between economic growth, freight transport, and energy consumption for 63 developing countries over the period 1929–2016. They divided the global panel into three subpanels, lower-middle-income countries (LMIC), upper-middle-income countries (UMIC), and high-income countries (MIC). Using the IPS (2003) and LLC (2002) tests, they rejected the null of stationarity based on the 10% significance level. Under the assumption that all series are nonstationary, they tested the null of stationarity for the first-differenced series. They accepted the null of stationarity. Based on the GMM estimates, they showed that there were bidirectional Granger causal relationships between economic growth and freight transport for all subpanels and between economic growth and energy consumption for the HIC and UMIC countries, but unidirectional running from energy consumption to economic growth for LMIC countries.

5.4.2 Tests of Cointegration

5.4.2.1 Residual-Based Tests

One can test the existence of cointegration relationships by testing if the residuals of regressing one element of $\boldsymbol{w}_{it}$ on the rest of the elements of $\boldsymbol{w}_{it}$ have a unit root or not. For instance, if the residuals of (5.3.14) contain a unit root, $\boldsymbol{w}_{it}$ are not cointegrated.

If the residuals are stationary, $\boldsymbol{w}_{it}$ are cointegrated. In a time series framework, the null of time series unit root tests is unit root. In other words, the null is no cointegration. Under the null, time series regression of $\boldsymbol{\beta}$ fails to converge (Phillips 1987). On the other hand, panels with large N and large T can yield a convergent estimate of $\boldsymbol{\beta}$. If there is no cointegration, Kao (1999), using a sequential limit argument (first $T \to \infty$, followed by $N \to \infty$), has shown that the least squares estimator of (5.3.14) converges to $\Omega_{\Delta\tilde{w}}^{-1}\Omega_{\Delta\tilde{w}u}$. If there is cointegration, then the panel fully modified within estimator (5.3.19) or the within estimator for the lead-lag adjusted regression model (5.3.20) is consistent and asymptotically normally distributed. Moreover, since u_{it} is independently distributed across i, a version of the central limit theorem can be invoked on the cross-sectional average of a statistic. That is, even though a statistic could have different asymptotic properties depending on whether u_{it} contains a unit root or not along a time series dimension, the properly scaled cross-sectional average of such a statistic is asymptotically normally distributed as $N \to \infty$. What this implies is that the panel null for the distribution of u_{it} could either be unit root (no cointegration) or stationary (cointegration).

McCoskey and Kao (1998) proposed to test the null of cointegration using the statistic

$$\overline{LM} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}S_{it}^{+2}}{s^{+2}}, \tag{5.4.30}$$

where

$$S_{it}^{+} = \sum_{s=1}^{t}\hat{u}_{1is},\ s^{+2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{1it}^{+2},$$

$\hat{u}_{1it}$ are the estimated residuals of (5.3.14) with $\boldsymbol{\beta}$ estimated by either (5.3.19) or applying the within estimator to (5.3.20) and $\hat{u}_{1it}^{+}$ is the estimated residual of (5.3.18). McCoskey and Kao (1998) show that $\sqrt{N}(\overline{LM})$ is asymptotically normally distributed with mean μ and variance σ_v^2, if individual units are independent of each other where μ and σ_v^2 can be found through Monte Carlo simulations.

Kao (1999) and Pedroni (2004) propose test cointegration under the null of no cointegration (residuals contain unit root). Kao proposes modified Dickey–Fuller–type test statistics to take account of correlations between regressors and error or serial correlations. Pedroni (2004) considers the unit root test statistic against "homogeneous" alternatives or the "heterogeneous" alternative. For surveys of panel unit root and cointegration tests, see Baltagi and Kao (2000), Banerjee (1999), Breitung and Das (2008), and Choi (2002a, 2002b, 2006).

5.4.2.2 Likelihood Approach

The residual-based tests can tell only if there exist cointegration relations among elements of $\boldsymbol{w}_{it}$; they cannot tell the cointegration rank, nor account for the correlations across cross-section units. For the "homogeneous" PVAR model, a test of the cointegration rank can be conducted by maximizing the (transformed) likelihood function (5.2.28) subject to either

$$\Lambda_r = H_r\boldsymbol{\delta}_r + B_r, \tag{5.4.31}$$

or

$$\Lambda_{r+1} = H_{r+1}\boldsymbol{\delta}_{r+1} + B_{r+1}, \tag{5.4.32}$$

where H_r and H_{r+1} are, respectively, $m \times (m-r)$ and $m \times (m-r-1)$ matrices with known elements, and B_r and B_{r+1} are, respectively, $m \times r$ and $m \times (r+1)$ matrices with known elements, and $\boldsymbol{\delta}_r$ and $\boldsymbol{\delta}_{r+1}$ are, respectively, $(m-r) \times r$ and $(m-r-1) \times (r+1)$ matrices with unknown elements. When T is fixed and $N \to \infty$, the likelihood ratio test statistic of cointegration rank r versus rank $(r+1)$ is asymptotically chi-square distributed with $(m-r)^2 - (m-r-1)^2 = 2(m-r) - 1$ degrees of freedom (Binder et al. 2005). (Imposing Π to be of rank r leaves $m^2 - (m-r)^2$ unrestricted coefficients in Π).

For the case of "heterogeneous" PVAR, the panel VEC approach discussed in the previous section can provide a basis to test for a common rank of cointegration on all individual units while allowing different dynamics across the individuals and interdependencies between the different individuals.

When rank $(\Pi_i^*) = m$, (5.3.6) implies $\boldsymbol{w}_{it}$ are stationary. For given i, Johansen (1991, 1995) provided the likelihood ratio tests for each of the restrictions $r = 0, \ldots, m-1$ versus stationarity (rank $(\Pi_i^*) = m$). He showed that as $T \longrightarrow \infty$, the likelihood ratio statistic of testing rank $(\Pi_i^*) = r$ versus rank $(\Pi_i^*) = m$ converges in distribution to

$$tr\left(\int dB_{m-r}B'_{m-r}\left[\int B_{m-r}B'_{m-r}\right]^{-1}\int B_{m-r}dB'_{m-r}\right), \tag{5.4.33}$$

where B_{m-r} is an $(m-r)$-dimensional vector Brownian motion with an identity covariance matrix. Larsson et al. (2001) presented a likelihood-based panel test of cointegrating rank for heterogeneous panels based on the average of the individual (Johansen) rank trace statistics. When N is fixed and T is large, Groen and Kleibergen (2003) show that the likelihood ratio test of common cointegration rank $r(r = 0, 1, \ldots, m-1)$ across cross-sectional units versus $r = m$ (stationarity) converges in distribution to

$$\begin{aligned}&\sum_{i=1}^{N} LR_i(r \mid m)\\ &\quad = \sum_{i=1}^{N} tr\left(\int dB_{m-r,i}B'_{m-r,i}\left[\int B_{m-r,i}B'_{m-r,i}\right]^{-1}\int B_{m-r,i}dB'_{m-r,i}\right),\end{aligned} \tag{5.4.34}$$

where $B_{m-r,i}$ is an $(m-r)$-dimensional Brownian motion for individual i. They show that (5.4.34) is robust with respect to cross-sectional dependence. Using a sequential limit argument ($T \to \infty$ followed by $N \to \infty$), Groen and Kleibergen (2003) show that

$$\frac{\overline{LR}(r \mid m) - E(\overline{LR}(r \mid m))}{\sqrt{\text{Var}\,(\overline{LR}(r \mid m))}} \tag{5.4.35}$$

converges to a standard normal, where $\overline{LR}(r \mid m) = \frac{1}{N}\sum_{i=1}^{N} LR_i(r \mid m)$.

5.5 DYNAMIC SIMULTANEOUS EQUATION MODELS

5.5.1 The Model

The discussions in previous sections, in particular, the panel VAR model (5.2.1), can be considered as a reduced form specification of the panel dynamic simultaneous equations model of the form (assuming $\boldsymbol{\delta} = \mathbf{0}$ for ease of exposition),

$$\Gamma(L)\boldsymbol{w}_{it} + C\boldsymbol{x}_{it} = \boldsymbol{\alpha}_i^* + \boldsymbol{\epsilon}_{it}, \tag{5.5.1}$$

where $\boldsymbol{x}_{it}$ is a $K \times 1$ vector satisfying $E(\boldsymbol{x}_{it}\boldsymbol{\epsilon}'_{is}) = \mathbf{0}$,

$$\Gamma(L) = \Gamma_0 + \Gamma_1 L + \cdots + \Gamma_p L^p, \tag{5.5.2}$$

and

$$\Gamma_0 \neq I_m, \tag{5.5.3}$$

with $C = \mathbf{0}$. For simplicity, we maintain the assumption that $\boldsymbol{\epsilon}_{it}$ is independently, identically distributed (i.i.d.) across i and over t with covariance matrix Ω. Model (5.5.1), in addition to the issues of (i) the presence of time-invariant individual effects and (ii) the assumption about initial observations, also contains (iii) contemporaneous dependence among elements of $\boldsymbol{w}_{it}$. Since statistical inference can be made only in terms of observed data, the joint dependence of observed variables raises the possibility that many observational equivalent theoretical structures could generate the same observed phenomena (e.g., Hood and Koopmans 1953). To uniquely identify the model of (5.5.1)–(5.5.3) from observed data, prior restrictions are needed. However, although the presence of $\boldsymbol{\alpha}_i^*$ creates correlations between $\boldsymbol{w}_{it}$ and all future or past $\boldsymbol{w}_{is}, \boldsymbol{\alpha}_i^*$ can be removed from the system through a linear transformation. For instance, taking the first difference of (5.5.1) yields a system of

$$\Gamma(L)\Delta\boldsymbol{w}_{it} + C\Delta\boldsymbol{x}_{it} = \Delta\boldsymbol{\epsilon}_{it}. \tag{5.5.4}$$

System (5.5.4) is in the form of Coweles Commission dynamic simultaneous equations model with first-order moving average error terms; hence, the usual rank condition is necessary and sufficient to identify an equation in the system (e.g., Hsiao 1983) if we assume that the prior information takes the form of exclusion restrictions, i.e., some variables are excluded from the gth equation. Let $[\Gamma_{0g}, \ldots, \Gamma_{pg}, C_g]$ be the matrix formed from the columns of $[\Gamma_0, \ldots, \Gamma_p, C]$ that are zero on the gth row. Then the necessary and sufficient condition for the identification of the gth equation of (5.5.1) is (e.g., Hsiao 1983),[9]

$$\text{rank}\left[\Gamma_{og}, \ldots, \Gamma_{pg}, C_g\right] = m - 1. \tag{5.5.5}$$

In other words, the presence of individual specific effects $\boldsymbol{\alpha}_i^*$ does not change the necessary and sufficient conditions for the identification of an equation in a system of m equations.

For ease of exposition, we assume $p = 1$, then (5.5.1) becomes

$$\Gamma_0\boldsymbol{w}_{it} + \Gamma_1\boldsymbol{w}_{i,t-1} + C\boldsymbol{x}_{it} = \boldsymbol{\alpha}_i^* + \boldsymbol{\epsilon}_{it}. \tag{5.5.6}$$

Premultiplying Γ_0^{-1} to (5.5.6) yields the reduced form

$$\boldsymbol{w}_{it} = H_1\boldsymbol{w}_{i,t-1} + H_2\boldsymbol{x}_{it} + \boldsymbol{\eta}_i + \boldsymbol{v}_{it}, \tag{5.5.7}$$

where $H_1 = -\Gamma_0^{-1}\Gamma_1, H_2 = -\Gamma_0^{-1}C, \boldsymbol{\eta}_i = \Gamma_0^{-1}\boldsymbol{\alpha}_i^*, \boldsymbol{v}_{it} = \Gamma_0^{-1}\boldsymbol{\epsilon}_{it}$.

Taking the first difference of (5.5.6) to eliminate $\boldsymbol{\alpha}_i^*$ yields

$$\Gamma_0\Delta\boldsymbol{w}_{it} + \Gamma_1\Delta\boldsymbol{w}_{i,t-1} + C\Delta\boldsymbol{x}_{it} = \Delta\boldsymbol{\epsilon}_{it}, \quad t = 2, \ldots, T, \quad i = 1, \ldots, N. \tag{5.5.8}$$

The basic idea of estimating a dynamic simultaneous equation can be illustrated by considering the Anderson-Rubin (1950)-type limited information approach in which only

[9] The vector error correction representation (5.3.6) subject to (5.3.7) is a reduced form specification with the prior knowledge of the rank of cointegrations among $\boldsymbol{w}_{it}$. For the identification for a structural VAR and dichotomization between long-run equilibrium and short-run dynamics, see Hsiao (2001).

the prior information on the gth equation is utilized. Let $g = 1$. Since only the prior restrictions of the first equation are utilized in inference, there is no loss of generality to write (5.5.6) in the form,

$$\Gamma_0 = \begin{bmatrix} 1 & -\boldsymbol{\gamma}_0' \\ \mathbf{0} & I_{m-1} \end{bmatrix}, \Gamma_1 = -\begin{bmatrix} \gamma_{11} & \boldsymbol{\gamma}_{12}' \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1' \\ \Pi_2 \end{bmatrix},$$
$$C = -\begin{bmatrix} \boldsymbol{c}_1' \\ \Pi_{23} \end{bmatrix}. \tag{5.5.9}$$

We assume that the prior restrictions are in the form of exclusion restrictions and there are at least $(m-1)$ elements in the vector $(\boldsymbol{\gamma}_0', \gamma_{11}, \boldsymbol{\gamma}_{12}', \boldsymbol{c}_1')$ which are zero and the rank condition for the identification of the first equation is satisfied. For ease of notations, we assume all m elements of $\boldsymbol{w}_{it}$ appear in the first equation.

Suppose the first equation of a dynamic simultaneous equations model (5.5.6),

$$w_{1it} = \boldsymbol{\gamma}_0' \boldsymbol{w}_{2it} + \boldsymbol{\gamma}_1' \boldsymbol{w}_{i,t-1} + \boldsymbol{c}_1' \boldsymbol{x}_{1it} + \alpha_{1i}^* + \epsilon_{1it}, \tag{5.5.10}$$

where $\boldsymbol{w}_{it} = (w_{1it}, \boldsymbol{w}_{2it}')', \boldsymbol{x}_{it} = (\boldsymbol{x}_{1it}', \boldsymbol{x}_{2it}')$. Taking the first difference to get rid of α_{1i}^* yields

$$\Delta w_{1it} = \boldsymbol{\gamma}_0' \Delta \boldsymbol{w}_{2it} + \boldsymbol{\gamma}_1' \Delta \boldsymbol{w}_{i,t-1} + \boldsymbol{c}_1' \Delta \boldsymbol{x}_{1it} + \Delta \epsilon_{1it}, \quad t = 2, \ldots, T. \tag{5.5.11}$$

5.5.2 Method of Moments Approach

5.5.2.1 Simple Instrumental Variable Estimator (IV)

Generalizing the Anderson and Hsiao (1981, 1982) simple IV estimator to (5.5.11), the moment conditions are:

$$E\left[\begin{pmatrix} \boldsymbol{w}_{i,t-2} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \Delta \epsilon_{1it}\right] = \mathbf{0}, \tag{5.5.12}$$

or

$$E\left[\begin{pmatrix} \Delta \boldsymbol{w}_{i,t-2} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \Delta \epsilon_{1it}\right] = \mathbf{0}. \tag{5.5.13}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0', \boldsymbol{\gamma}_1', \boldsymbol{c}_1')$ and $\boldsymbol{z}_{1it} = (\boldsymbol{w}_{2it}', \boldsymbol{w}_{i,t-1}')'$, the simple IV takes the form:

$$\hat{\boldsymbol{\theta}}_{IV} = \left[\sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} \boldsymbol{z}_{1,it} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_{1i,t-2}' & \Delta \boldsymbol{x}_{1it}' \end{pmatrix}\right]^{-1} \left[\sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} \boldsymbol{z}_{1i,t-2} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \Delta w_{1it}\right], \tag{5.5.14}$$

or

$$\hat{\boldsymbol{\theta}}_{IV} = \left[\sum_{i=1}^N \sum_{t=3}^T \begin{pmatrix} \boldsymbol{z}_{1,it} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{z}_{1i,t-2}' & \Delta \boldsymbol{x}_{1it}' \end{pmatrix}\right]^{-1} \left[\sum_{i=1}^N \sum_{t=3}^T \begin{pmatrix} \Delta \boldsymbol{z}_{1i,t-2} \\ \Delta \boldsymbol{x}_{1it} \end{pmatrix} \Delta w_{1it}\right]. \tag{5.5.15}$$

5.5.2.2 Panel Two Stage Least Squares Estimator (P2SLS)

Let $\Delta \boldsymbol{\varepsilon}_{1i} = (\Delta \varepsilon_{1i2}, \ldots, \Delta \varepsilon_{1iT})'$. Noting

$$E(\Delta \boldsymbol{\varepsilon}_{1i} \Delta \boldsymbol{\varepsilon}_{1i}') = \sigma_1^2 \tilde{A}, \tag{5.5.16}$$

where

$$\tilde{A} = \begin{bmatrix} 2 & -1 & 0 & \cdot & \cdot & 0 \\ -1 & 2 & -1 & \cdot & \cdot & \cdot \\ 0 & -1 & 2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & -1 & 2 \end{bmatrix} \tag{5.5.17}$$

and suppose $\boldsymbol{q}_{1it}$ be a vector of instruments that satisfy $E(\boldsymbol{q}_{1it}\Delta\varepsilon_{1it}) = 0$. The P2SLS takes the form

$$\hat{\boldsymbol{\theta}} = \left\{\sum_{i=1}^{N}\begin{pmatrix} Z_{1i} \\ \Delta X'_{1i}\end{pmatrix} Q'_{1i}\left[\sum_{i=1}^{N} Q_{1i}\tilde{A}Q'_{1i}\right]^{-1} Q'_{1i}\left(Z'_{1i} \quad \Delta X'_{1i}\right)\right\}^{-1}$$
$$\left\{\sum_{i=1}^{N}\begin{pmatrix} Z_{1i} \\ \Delta X_{1i}\end{pmatrix} Q'_{1i}\left[\sum_{i=1}^{N} Q_{1i}\tilde{A}Q'_{1i}\right]^{-1} Q'_{1i}\Delta\boldsymbol{w}_{1i}\right\} \tag{5.5.18}$$

where $Z_{1i} = (z_{1,i1}, \ldots, z_{1,iT})$, $\Delta X_{1i} = (\Delta\boldsymbol{x}_{1,i1}, \ldots, \Delta\boldsymbol{x}_{1,iT})$, $Q_{1i} = (\boldsymbol{q}_{1,i1}, \ldots, \boldsymbol{q}_{1,iT})$, and $\Delta\boldsymbol{w}_{1i} = (\Delta w_{1,i1}, \ldots, \Delta w_{1,iT})'$.

Both the simple IV and P2SLS are consistent and asymptotically normally distributed centered at the true value when $NT \to \infty$, whether N or T is finite or both N and T go to infinity.

5.5.2.3 Generalized Method of Moments Estimator

Just like the single equation case, $\boldsymbol{w}_{i,t-2}$ or $\Delta\boldsymbol{w}_{i,t-2}$ are not the only instruments that satisfy (5.5.12) or (5.5.13). All $\boldsymbol{w}_{i,t-2-j}$ (or $\Delta\boldsymbol{w}_{i,t-2-j}$), $j = 1, \ldots, t-2$ and $\Delta\boldsymbol{x}'_i = (\Delta\boldsymbol{x}'_{i2}, \ldots, \Delta\boldsymbol{x}'_{iT})$ are legitimate instruments. Let $\boldsymbol{q}'_{it} = (\boldsymbol{w}'_{i,t-2}, \ldots, \boldsymbol{w}'_{i0}, \Delta\boldsymbol{x}'_i)$, and let

$$D_i = \begin{bmatrix} \boldsymbol{q}_{i2} & \mathbf{0} & \mathbf{0} & \cdot & 0 \\ \mathbf{0} & \boldsymbol{q}_{i3} & \mathbf{0} & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \cdot & \cdot & \cdot & \boldsymbol{q}_{iT} \end{bmatrix}, \tag{5.5.19}$$

be the $R \times (T-1)$ block dimensional matrix, where R denotes the dimension of $(\boldsymbol{q}'_{i2}, \ldots, \boldsymbol{q}'_{iT})'$, then

$$E(D_i\Delta\boldsymbol{\epsilon}_{1i}) = 0, \quad i = 1, \ldots, N, \tag{5.5.20}$$

where $\Delta\boldsymbol{\epsilon}'_{1i} = (\Delta\epsilon_{1i2}, \ldots, \Delta\epsilon_{1iT})$ with $E(\Delta\boldsymbol{\epsilon}_{1i}) = \mathbf{0}$ and $E(\Delta\boldsymbol{\epsilon}_{1i}\Delta\boldsymbol{\epsilon}'_{1i}) = \sigma_1^2\tilde{A}$, $\sigma_1^2 = E(\epsilon_{1it}^2)$. The Arellano–Bond (1991) type GMM estimator of $\boldsymbol{\theta}$ is to find $\hat{\boldsymbol{\theta}}$ that minimizes

$$\left(\sum_{i=1}^{N}\Delta\boldsymbol{\epsilon}'_{1i}D'_i\right)\left(\sum_{i=1}^{N}D_i\tilde{A}D'_i\right)^{-1}\left(\sum_{i=1}^{N}D_i\Delta\boldsymbol{\epsilon}_{1i}\right), \tag{5.5.21}$$

which yields an estimator of the form,

$$\hat{\boldsymbol{\theta}}_{GMM} = \left\{\left[\sum_{i=1}^{N} \tilde{Z}_i D_i'\right]\left(\sum_{i=1}^{N} D_i \tilde{A} D_i'\right)^{-1}\left[\sum_{i=1}^{N} D_i \tilde{Z}_i\right]\right\}^{-1}$$
$$\times \left\{\left[\sum_{i=1}^{N} \tilde{Z}_i D_i'\right]\left(\sum_{i=1}^{N} D_i \tilde{A} D_i'\right)^{-1}\left[\sum_{i=1}^{N} D_i \Delta \boldsymbol{w}_{1i}\right]\right\}, \tag{5.5.22}$$

and an asymptotic covariance matrix,

$$\sigma_1^2 \left\{\left[\sum_{i=1}^{N} \tilde{Z}_i' D_i'\right]\left(\sum_{i=1}^{N} D_i \tilde{A} D_i'\right)^{-1}\left[\sum_{i=1}^{N} D_i \tilde{Z}_i\right]\right\}^{-1}, \tag{5.5.23}$$

where $\tilde{Z}_i = (\Delta W_{2i}, \Delta \tilde{W}_{i,-1}, \Delta X_{1i}), \Delta W_{2i}, \Delta \tilde{W}_{i,-1}$ denote the $T-1$ time series stacked $\Delta \boldsymbol{w}_{2it}$ and $\Delta \boldsymbol{w}_{i,t-1}$ that appear in the first equation, respectively. When T is fixed and $N \to \infty, \sqrt{NT}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})$ is asymptotically normally distributed with mean $\boldsymbol{\theta}$. However, because

$$E\left[\begin{pmatrix} \Delta W_{2i}' \\ \Delta \tilde{W}_{i,-1}' \end{pmatrix} D_i' \left(\sum_{i=1}^{N} D_i \tilde{A} D_i'\right)^{-1} D_i \Delta \boldsymbol{\epsilon}_{1i}\right] \neq \mathbf{0}, \tag{5.5.24}$$

the process of removing individual-specific effects creates a second order bias that is of order $\log T$. If T increases with N, it is shown by Akashi and Kunitomo (2011, 2012) that the GMM is inconsistent if $\frac{T}{N} \to c \neq 0$ as $N \to \infty$. Even when $c \to 0$, as long as $\frac{T^3}{N} \neq 0$ as $N \to \infty, \sqrt{NT}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})$ is biased of order $\sqrt{\frac{T^3}{N}}$. Monte Carlo studies conducted by Akashi and Kunitomo (2011, 2012) show that the GMM estimator is badly biased when N and T are both large.

Akashi and Kunitomo (2011, 2012) propose a panel least variance ratio estimator (PLVAR) which is a panel generalization of Anderson and Rubin's (1950) limited information maximum likelihood estimator formula.[10] They show that when $\frac{T}{N} \to 0$ as $N \to \infty$, the panel least variance ratio estimator is asymptotically normally distributed centered at the true value. They have also derived the asymptotic bias of PLVAR when $\frac{T}{N} \to c \neq 0$. Monte Carlo studies confirm that the PLVAR is almost median-unbiased after correcting the bias.

5.5.2.4 *Jackknife Instrumental Variable Estimator (JIVE)*

Whether an estimator of $\boldsymbol{\theta}$ multiplied by the scale factor $\sqrt{NT}$ (the magnitude of the inverse of the standard error) is centered at the true value or not has important implications in hypothesis testing. A consistent but asymptotically biased estimator could lead to significant size distortion in hypothesis testing (Hsiao and Zhang 2015).

Noting that although $E(D_i \Delta \boldsymbol{\epsilon}_{1i}) = \mathbf{0}$, the source of asymptotic bias of the Arellano–Bond type GMM is (5.5.24) in the use of cross-sectional mean $\frac{1}{N}\sum_{i=1}^{N} D_i \Delta \boldsymbol{\epsilon}_{1i}$ to

[10] Akashi and Kunitomo (2012) actually called their estimator PLIML. However, it appears more appropriate to call their estimator panel least variance ratio estimator.

approximate the population moments $E(D_i \Delta\boldsymbol{\epsilon}_{1i})$. When $(\boldsymbol{\epsilon}_i)$ are independently distributed over i, Hsiao and Zhou (2019) note that

$$E\left[\begin{pmatrix}\Delta W_{2i}' \\ \Delta \tilde{W}_{i,-1}'\end{pmatrix} D_i' \left(\sum_{i=1}^{N} D_i \tilde{A} D_i'\right)^{-1} D_j \Delta \boldsymbol{\varepsilon}_{1j}\right] = 0,$$

and propose a bias reduced JIVE estimator,[11]

$$\hat{\boldsymbol{\theta}}_{JIVE} = \left\{\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{i\neq j}\begin{pmatrix}\boldsymbol{z}_{1it} \\ \Delta^*\boldsymbol{x}_{1it}\end{pmatrix}\boldsymbol{q}_{it}'\left(\sum_{i=1}^{N}\boldsymbol{q}_{it}\boldsymbol{q}_{it}'\right)^{-1}\boldsymbol{q}_{jt}(\boldsymbol{z}_{1jt}' \; \Delta^*\boldsymbol{x}_{1jt}')\right\}^{-1}$$
$$\cdot \left\{\sum_{t=2}^{T}\sum_{j=1}^{N}\sum_{i\neq j}\begin{pmatrix}\boldsymbol{z}_{1it} \\ \Delta^*\boldsymbol{x}_{1it}\end{pmatrix}\boldsymbol{q}_{it}'\left(\sum_{i=1}^{N}\boldsymbol{q}_{it}\boldsymbol{q}_{it}'\right)^{-1}\boldsymbol{q}_{jt}\Delta^* w_{1jt}\right\}, \quad (5.5.25)$$

where Δ^* denotes the forward differencing operator (see Remark 3.3.2 in Section 3.3) and $\boldsymbol{f}_{it}' = (y_{ie}, \cdot\cdot, y_{i,t-1}, \boldsymbol{x}_i')$.

Hsiao and Zhou (2018) show that $\hat{\boldsymbol{\theta}}_{JIVE}$ is asymptotically normally distributed centered at the true value of $\boldsymbol{\theta}$ when $\frac{T}{N} \to c \neq 0$ as $N \to \infty$.

5.5.3 Likelihood Approach

Just like the single equation model discussed in Chapter 3, the likelihood approach assumes that the data generating process of $\boldsymbol{w}_{i0}$ is no different from the data generating process of any $\boldsymbol{w}_{it}$; hence,

$$\boldsymbol{w}_{i0} = [I - H_1 L]^{-1}[H_2\boldsymbol{x}_{i0} + \boldsymbol{\eta}_i^* + \boldsymbol{v}_{i0}], \quad (5.5.26)$$

where L denotes the lag operator, $L\boldsymbol{w}_{it} = \boldsymbol{w}_{i,t-1}$. Under the assumption that the data generating process of $\boldsymbol{x}_{it}$ is homogeneous (e.g., (3.4.12) or (3.5.3)), we can write[12]

$$[I - H_1 L]^{-1} H_2 \boldsymbol{x}_{i0} = A\bar{\boldsymbol{x}}_i + \boldsymbol{\omega}_i, \quad (5.5.27)$$

where $\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}$, and $\boldsymbol{\omega}_i$ is i.i.d. across i. Substituting (5.5.27) into (5.5.26) yields

$$\boldsymbol{w}_{i0} = A\bar{\boldsymbol{x}}_i + [I - H_1]^{-1}\boldsymbol{\eta}_i^* + \boldsymbol{\omega}_i + [I - H_1 L]^{-1}\boldsymbol{v}_{i0}. \quad (5.5.28)$$

When $\boldsymbol{\eta}_i^*$ are random and uncorrelated with $\boldsymbol{x}_{is}$ with known distribution, maximizing the joint likelihood of (5.5.6) and (5.5.28) is consistent and asymptotically normally distributed, either N or T or both are large. However, the MLE involves multiple integrations over t, which can be computationally complicated. On the other hand, conditional on $\boldsymbol{\eta}_i^*$, $\boldsymbol{\epsilon}_{it}$ is independently distributed across i and over t. If both N and T are large, the MLE of structural parameters and $\boldsymbol{\eta}_i^*$ are consistent and asymptotically normally distributed. However, if T is finite, $\boldsymbol{\eta}_i^*$ cannot be consistently estimated. Therefore, we shall consider a limited information MLE for the transformed model that eliminates $\boldsymbol{\eta}_i^*$. Although the first difference can remove $\boldsymbol{\eta}_i^*$, the transformed errors follow a first-order moving average process. Inverting a matrix of the form (5.5.17) can be computationally tedious. Therefore, Hsiao and Zhou (2015), following the suggestion of Grassetti (2011) to take the long difference, $\boldsymbol{w}_{it}^* = \boldsymbol{w}_{it} - \boldsymbol{w}_{i0}$, to eliminate $\boldsymbol{\eta}_i^*$ and transform the dynamic

[11] The forward difference error u_{it}^* (see Remark 3.3.2) has $Eu_{it}^*u_{is}^* = 0$ if $t \neq s$ and $Eu_{it}^{*2} = \sigma_u^2$.

[12] For a stationary invertible MA process, $\boldsymbol{X}_t$ can be equivalently written $\boldsymbol{X}_{it} = A(F)\boldsymbol{X}_{i,t+1} + \varepsilon_{it}$, $\sum_{j=1}^{\infty}|A_j| < \infty$ and F denotes the forward operator (Box and Jenkins (1970), ch.6). For ease of notations, we use $\bar{\boldsymbol{X}}_i$ in lieu of $\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots$.

simultaneous equations model into a model that has the variance–covariance matrix of the error terms to have the form of a random-effects model.

Multiplying Γ_0 to (5.5.28) yields[13]

$$\Gamma_0 \boldsymbol{w}_{i1}^* + C\boldsymbol{x}_{i1} + A^*\bar{\boldsymbol{x}}_i = \boldsymbol{\xi}_i + \boldsymbol{\epsilon}_{i1}, \tag{5.5.29}$$

where $A^* = \Gamma_0 A$, $\boldsymbol{\xi}_i = \Gamma_0\boldsymbol{\omega}_i - \Gamma_0(I - H_1 L)^{-1}\boldsymbol{v}_{i0}$. The long difference of the system (5.5.6) using (5.5.29) becomes,

$$\begin{aligned} \Gamma_0 \boldsymbol{w}_{it}^* + \Gamma_1 \boldsymbol{w}_{i,t-1}^* + C\boldsymbol{x}_{it} + A^*\bar{\boldsymbol{x}}_i &= \boldsymbol{\xi}_i + \boldsymbol{\epsilon}_{it}, \\ & i = 1, \ldots, N \\ & t = 1, \ldots, T. \end{aligned} \tag{5.5.30}$$

The $(mT \times 1)$ error term $\boldsymbol{v}_i = [(\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_{i1})', \ldots, (\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_{iT})']'$ has mean $\mathbf{0}$ and covariance matrix of the form analogous to the single equation random effects covariance matrix,

$$\Omega = E\boldsymbol{v}_i\boldsymbol{v}_i' = \Omega_\epsilon \otimes I_T + \Omega_\xi \otimes \boldsymbol{e}_T\boldsymbol{e}_T', \tag{5.5.31}$$

where $\Omega_\epsilon = E(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'), \Omega_\xi = E(\boldsymbol{\xi}_i\boldsymbol{\xi}_i')$ and $\boldsymbol{e}_T$ is a $T \times 1$ vector $(1, \ldots, 1)'$. Rewrite the covariance matrix in terms of the eigenvalues Ω_ϵ and $\Omega_{\xi^*} = \Omega_\epsilon + T\Omega_\xi$ (Chapter 4, Section 4.2; Avery 1977),

$$\Omega = \Omega_\epsilon \otimes Q + \Omega_{\xi^*} \otimes J, \tag{5.5.32}$$

where $Q = I_T - \frac{1}{T}\boldsymbol{e}_T\boldsymbol{e}_T'$ and $J = \frac{1}{T}\boldsymbol{e}_T\boldsymbol{e}_T'$. It follows that

$$\Omega^{-1} = \Omega_\epsilon^{-1} \otimes Q + \Omega_{\xi^*}^{-1} \otimes J. \tag{5.5.33}$$

Thus, the log-likelihood function of the transformed system (5.5.30) takes the multivariate analogue of the single-equation log-likelihood function (3.5.7),

$$\begin{aligned} & -\frac{N(T-1)}{2} \log \mid \Omega_\epsilon \mid - \frac{N}{2} \log \mid \Omega_{\xi^*} \mid \\ & -\frac{1}{2}\sum_{i=1}^{N} \Big\{ [\boldsymbol{w}_{1i}^{*'} - \tilde{\boldsymbol{\theta}}' Z_i', \text{ vec } (W_{2i}^{*'})' - \text{ vec } (\Pi')'(I_{m-1} \otimes \tilde{X}_i')]\Omega^{-1} \\ & \quad [\boldsymbol{w}_{1i}^{*'} - \tilde{\boldsymbol{\theta}}' Z_i', \text{ vec } (W_{2i}^{*'})' - \text{ vec } (\Pi')'(I_{m-1} \otimes \tilde{X}_i')]' \Big\}. \end{aligned} \tag{5.5.34}$$

where $W_i^* = (\boldsymbol{w}_{1i}^*, W_{2i}^*), W_{2i}^* = (\boldsymbol{w}_{2i}^*, \ldots, \boldsymbol{w}_{mi}^*), \boldsymbol{w}_{ji}^{*'} = (\boldsymbol{w}_{ji1}^*, \ldots, \boldsymbol{w}_{jiT}^*), j = 1, \ldots, m, Z_i = (W_{2i}^*, \tilde{W}_{i,-1}^*, X_{1i}, \boldsymbol{e}_T\bar{\boldsymbol{x}}_i'), \tilde{X}_i = (W_{i,-1}^*, X_i, \boldsymbol{e}_T\bar{\boldsymbol{x}}_i'), W_{i,-1}^* = (\tilde{W}_{i,-1}^*, \tilde{\tilde{W}}_{i,-1}^*), \tilde{W}_{i,-1}^*$ and $\tilde{\tilde{W}}_{i,-1}$ denote the $T \times \tilde{m}$ and $T \times (m - \tilde{m})$ matrix of $\boldsymbol{w}_{ji,-1}^* = (0, w_{ji1}^*, \ldots, w_{ji,T-1}^*)'$ that appear or are excluded from equation 1, respectively; $X_i = (X_{1i}, X_{2i})$, X_{1i}, and X_{2i} denote the $T \times k$ and $T \times (K - k)$ matrix of included and excluded $\boldsymbol{x}_{it}$ in the first equation, respectively; and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}', \boldsymbol{a}_1^{*'}), \Pi = (\Pi_{21}, \Pi_{22}, \Pi_{23}, A_2^*)$, where $\boldsymbol{a}_1^{*'}$ and A_2^* denote the $1 \times K$ vector and $(m-1) \times K$ matrix of $A^* = \begin{pmatrix} \boldsymbol{a}_1^{*'} \\ A_2^* \end{pmatrix}$, respectively.

The panel limited information maximum likelihood estimator (PLIML) can be obtained by iterating between

[13] Should (3.5.4) be assumed as a data generating process for $\boldsymbol{x}_{it}$, then (5.5.29) in terms of first difference should be specified as $\Gamma_0 \Delta \boldsymbol{w}_{i1} + C\boldsymbol{x}_{i1} + A^*\Delta\bar{\boldsymbol{x}}_i = \boldsymbol{\xi}_i + \epsilon_{i1}$, where $\Delta\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=2}^{T} \Delta \boldsymbol{x}_{it}$.

$$\hat{\Omega}_{\epsilon} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \Big\{ \big[\boldsymbol{w}_{1i}^{*'} - \hat{\tilde{\boldsymbol{\theta}}}'_1 Z'_i,\ \text{vec}\,(W_{2i}^{*'}) - \ \text{vec}\,(\Pi')'(I_{m-1} \otimes \tilde{X}'_i) \big] \\ \times \big[I_m \otimes Q \big] \big[\boldsymbol{w}_{1i}^{*'} - \hat{\tilde{\boldsymbol{\theta}}}' Z'_i,\ \text{vec}\,(W_{2i}^{*'})' - \ \text{vec}\,(\Pi')'(I_{m-1} \otimes \tilde{X}'_i) \big]' \Big\}, \tag{5.5.35}$$

$$\hat{\Omega}_{\xi *} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \hat{\bar{u}}_{i1}^2 & \hat{\bar{u}}_{i1} \hat{\bar{\boldsymbol{u}}}'_{2i} \\ \hat{\bar{\boldsymbol{u}}}_{2i} \hat{\bar{u}}_{i1} & \hat{\bar{\boldsymbol{u}}}_{2i} \hat{\bar{\boldsymbol{u}}}'_{2i} \end{bmatrix}, \tag{5.5.36}$$

and

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_1 \\ \text{vec}\,(\hat{\Pi}') \end{pmatrix} = \left\{ \sum_{i=1}^{N} \begin{pmatrix} Z'_i & \mathbf{0} \\ \mathbf{0} & I_{m-1} \otimes \tilde{X}'_i \end{pmatrix} \hat{\Omega}^{-1} \begin{pmatrix} Z_i & \mathbf{0} \\ \mathbf{0} & I_{m-1} \otimes \tilde{X}_i \end{pmatrix} \right\}^{-1} \\ \left\{ \sum_{i=1}^{N} \begin{pmatrix} Z'_i & \mathbf{0} \\ \mathbf{0} & I_{m-1} \otimes \tilde{X}'_i \end{pmatrix} \hat{\Omega}^{-1} \begin{pmatrix} \boldsymbol{w}_{1i}^{*} \\ \text{vec}\,(W_{2i}^{*'}) \end{pmatrix} \right\}, \tag{5.5.37}$$

until the solution converges, where $\hat{\Omega} = \hat{\Omega}_{\varepsilon} \otimes I_T + \hat{\Omega}_{\xi *} \otimes \boldsymbol{e}_T \boldsymbol{e}'_T$, $\bar{u}_{1i} = \bar{w}_{1i}^{*} - \bar{Z}_{1i} \hat{\boldsymbol{\delta}}_1$, $\bar{\boldsymbol{u}}_{2i} = \bar{\boldsymbol{w}}_{2i}^{*} - (I_{m-1} \otimes \tilde{\boldsymbol{x}}'_i)\ \text{vec}\,(\Pi')$, $\bar{\boldsymbol{w}}_i^{*} = \frac{1}{T} \sum_{i=1}^{T} \boldsymbol{w}_{it}^{*}$, $\bar{Z}_{1i} = \frac{1}{T} \sum_{i=1}^{T} Z_{1it}$, and $\bar{\boldsymbol{x}}_i = \frac{1}{T} \sum_{i=1}^{T} \boldsymbol{x}_{it}$. The PLMLE is consistent and asymptotically normally distributed when $N \to \infty$, whether T is fixed or also goes to infinity.