USC Dornsife Institute for New Economic Thinking

Working Paper No. 14-02

Panel Macroeconometric Modeling

Cheng Hsiao

September 29, 2014

Panel Macroeconometric Modeling*

Cheng Hsiao

Department of Economics, University of Southern California

Department of Quantitative Finance, National Tsing Hua University,

and

WISE, Xiamen University

September 29, 2014

Abstract

This paper provides a selective survey of the panel macroeconometric techniques that focus on controlling the impact of “unobserved heterogeneity” across individuals and over time to obtain valid inference for “structures” that are common across individuals and over time. We consider issues of (i) estimating vector autoregressive models; (ii) testing of unit root or cointegration; (iii) statistical inference for dynamic simultaneous equations models; (iv) policy evaluation; and (v) aggregation and prediction.

Keywords: Panel macro-modeling; Unit root; Cointegration; Aggregation; Prediction

JEL code: C10; C33; C51

*This paper is dedicated to P.C.B. Phillips for his creative and lasting contributions to econometrics. The work is partially supported by the China NSF grant #71131008. I would like to thank P.C.B. Phillips and a referee for their helpful comments.

1. Introduction

One of the distinct features of macroeconometric modeling is that an economy as a whole is not always predictable from the parts which compose the entire *system*. The progress of macroeconometric modelling is principally fueled by four developments: “The improvement in computational capacity, improvements in the quality and availability of economic data, developments in econometric theory, and the virtuous circle of improvements in macroeconomic theory and the evolution of macroeconometric models” (Bodkin, Klein and Marnah (1991, p. 527)). However, a model is not a mirror. In order to capture the essential relationships of the real world phenomena, a model must simplify. The details are reduced through suitable *aggregation* and *stochastic modeling*, namely, all the omitted factors that affect the outcomes are put into the error terms, ε .

Panel data contain more degrees of freedom and more sample variability than a cross-sectional or time series sample, hence, in principal, can yield more powerful inference (e.g. Hsiao (2007, 2014)). However, panel data, by nature, focus on individual outcomes. Factors affecting individual outcomes are numerous. The conventional “homogeneity” assumption at the suitable “aggregate” level and the probabilistic assumption of the error term ε conditional on certain observable factors may not hold. This paper intends to provide a selective survey of the panel macro econometric techniques that focus on controlling the impact of “unobserved heterogeneity” across individuals and over time to obtain valid inference on structures that are common across individuals and over time.

Panel data contain at least two dimensions, a cross-section dimension and a time series dimension. The multidimensional nature of panel data raises issues of statistical inference (Phillips and Moon (1999, 2000a,b)). Moreover, because of the inertia in human behavior and institutional, and technological rigidities, a behavioral equation is more appropriately specified in *dynamic* form (Nerlove (2000)). Therefore, the focus of this paper is on the issues of statistical inference for a vector of dynamic interdependent variables. We shall consider both the reduced form and structural form approaches for modeling dynamic

response. We shall consider issues of (i) estimating vector autoregressive model (VAR); (ii) testing of unit root or cointegration; (iii) statistical inference for dynamic structural equation; (iv) policy evaluation; and (v) Aggregation and prediction. Because panel data is multi-dimensional, we pay special attention to the way that sample size increases on inference. In section 2 and 3 we discuss the reduced form approaches for modeling system of equations by considering panel stationary vector autoregressive models (VAR), and nonstationary system, respectively. Unit root and cointegration tests are briefly discussed in section 4. Section 5 discusses Cowles Commission structural modeling approach by decomposing w_{it} into joint dependent variables y_{it} and exogenous variables x_{it} . Section 6 demonstrates how one can use panel data to provide a “measurement without theory” approach to evaluate the policy impact. Section 7 discusses issues of aggregation and prediction. Concluding remarks are made in section 8.

2. Stationary Panel Vector Autoregressive Models

Economic magnitudes are often determined by the joint work of many factors. To take account of both the joint dependence of state variables as well as their dependence on past behavior, vector autoregressive models (VAR) have been proposed as a data based approach to conveniently summarize the information contained in the data ((e.g. Hsiao 1979a,b, 1982), Sims (1980)).

It is well known that time series properties of an $m \times 1$ vector, w_{it} , depend critically on whether the underlying processes are (trend) stationary or integrated or cointegrated. To ensure the time series property of w_{it} containing intercepts and time trend term remains the same whether w_{it} contains unit roots or not, we consider a model of the form,

$$\Phi_i(L)(w_{it} - \eta_i - \delta t) = \epsilon_{it}, \quad (2.1)$$

where

$$\Phi_i(L) = I_m - \Phi_{i1}L - \dots - \Phi_{ip_i}L^{p_i}, \quad (2.2)$$

L denotes the lag operator, $Lw_{it} = w_{i,t-1}$, and ϵ_{it} are independently distributed over i and

t with mean 0 and covariance matrix Ω_i . For ease of exposition, we shall assume $p_i = 1$ and $\delta = 0$. When the roots of $|\Phi_i(z)| = 0$ all lie outside the unit circle, the process is (trend) stationary. When some or all roots of $|\Phi_i(z)| = 0$ lie inside the unit circle, we say the process is nonstationary. We consider the case of stationary process in this section.

The individual-specific effects, η_i , and $\Phi_i(L)$ can either be treated as *fixed constants* or *random variables*. If they can be viewed as random draws from a common population and inference is on the population properties from which those in the data are considered to be a random sample, then it is more appropriate to consider the marginal inference on the observables by treating them as random variables (*random effects* inference). If η_i and $\Phi_i(L)$ are from heterogeneous population, then it is more appropriate to make conditional inference, conditional on the *fixed effects*, η_i , and $\Phi_i(L)$. Unfortunately, without the specific knowledge of how the observables w_{it} interact with the unobservables, it is hard to make a choice between the two.

The random effects inference needs to postulate the distribution of unobservables η_i and Φ_i , $G(\eta_i, \Phi_i)$, and the computation of the marginal distribution can be complicated because

$$\begin{aligned} f(w_{i0}, \dots, w_{iT}) &= \int f(w_{i0}, \dots, w_{iT} \mid \eta_i, \Phi_i) dG(\eta_i, \Phi_i) \\ &= \int \prod_{t=1}^T f(w_{it} \mid w_{i,t-1}, \eta_i, \Phi_i) f(w_{i0} \mid \eta_i, \Phi_i) dG(\eta_i, \Phi_i) \end{aligned} \quad (2.3)$$

involves multiple integration of $m(T+1)$ dimensions. However, conditional on η_i, Φ_i , under the i.i.d. assumption about ϵ_{it} , the distribution of $f(w_{it} \mid w_{i,t-1}, \Phi_i, \eta_i)$ is independently distributed over time. Furthermore, the conditional inference, although is less efficient than the marginal inference, remains consistent even η_i and Φ_i are indeed random so we shall concentrate on the *fixed effects* inference.

(a) “Homogeneous” panel VAR (PVAR) models

By “homogeneous” PVAR models we mean apart from the time-invariant but individual-varying intercept $\alpha_i = -(I_m - \Phi_i)\eta_i$, the slope coefficients, $\Phi_i = \Phi_j = \Phi$.

(i) Generalized Method of Moments (GMM)

Just as in the single equation case (e.g. Anderson and Hsiao (1981, 1982)), $\alpha_i = (I - \Phi)\eta_i$ can be eliminated by first differencing,

$$\Delta w_{it} = \Phi \Delta w_{i,t-1} + \Delta \varepsilon_{it}, \quad t = 2, \dots, T, \quad (2.4)$$

where $\Delta = 1 - L$. Then we have the orthogonality conditions

$$E \left\{ (\Delta w_{it} - \Phi \Delta w_{i,t-1}) q'_{it} \right\} = 0, \quad \text{for } t = 2, \dots, T, \quad (2.5)$$

where $q_{it} = (w'_{i0}, \dots, w'_{i,t-2})'$. Therefore, an Arellano-Bond (1991) GMM method can be applied to obtain the consistent and asymptotically normally distributed estimator of Φ when T is fixed and N tends to infinity, provided none of the roots of $|I - \Phi z| = 0$ equals 1. However, if T also increases with N and $\frac{T}{N} \rightarrow c \neq 0$ as $N \rightarrow \infty$, then moment conditions (2.4) increase at order T^2 . Following the derivation of Alvarez and Arellano (2003), one can show that the GMM estimator is asymptotically biased of order $\sqrt{\frac{T}{N}}$, when it is multiplied by the scale factor $\sqrt{NT}$. This asymptotic bias term leads to severe size distortion and needs to be corrected for valid inference (e.g. Hsiao and Zhang (2013) Hsiao and Zhou (2014)).

(ii) (Transformed) Likelihood Approach¹

We note that (2.4) is well defined given Δw_{i1} . However, Δw_{i1} is random. Equation (2.1) implies that

$$\Delta w_{i1} = -(I - \Phi)w_{i0} + \varepsilon_{i1}. \quad (2.6)$$

Therefore, $E(\Delta w_{i1} \Delta w'_{i2}) = -\Omega$ and $E(\Delta w_{i1} \Delta w'_{it}) = 0$ for $t > 2$. Assuming w_{i0} has constant covariance matrix Ψ_0 , then $E(\Delta w_{i1} \Delta w'_{i1}) = (I - \Phi)\Psi_0(I - \Phi)' + \Omega = \Psi_1$. The joint likelihood is well defined and does not involve incidental parameters, η_i or α_i . The (quasi) likelihood function of $\Delta w'_i = (\Delta w'_{i1}, \dots, \Delta w'_{iT})$ is given by

$$\prod_{i=1}^N (2\pi)^{-\frac{T}{2}} |\Omega^*|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\Delta w_i - H_i \phi)' \Omega^{*-1} (\Delta w_i - H_i \phi) \right\} \quad (2.7)$$

¹See Han and Phillips (2013) for the properties of transformed MLE when $| \Phi(z) | = 0$ contains unit roots or near unit roots.

where

$$\begin{aligned}
H_i &= G'_i \otimes I_m, \\
G_i &= (0, \Delta w_{i1}, \dots, \Delta w_{i,T-1}), \\
\phi &= \text{vec}(\Phi'), \\
\Omega^* &= \begin{bmatrix} \Psi_1 & -\Omega & 0 & \cdot & 0 \\ -\Omega & 2\Omega & -\Omega & \cdot & \cdot \\ 0 & -\Omega & 2\Omega & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & -\Omega \\ 0 & 0 & \cdot & \cdot & 2\Omega \end{bmatrix}.
\end{aligned}$$

Maximizing the logarithm of (2.7), $\ell(\theta)$ with respect to $\theta' = (\phi', \sigma')$, where σ denotes the unknown elements of Ω^* yields the (transformed) (quasi) maximum likelihood estimator (QMLE) (Binder, Hsiao and Pesaran (2005)). The QMLE with properly formulated initial conditions is asymptotically unbiased independent of the way N or T or both tend to infinity (Hsiao and Zhou (2014)).

Conditional on Ω^* , the MLE of Φ is equivalent to the minimum distance estimator (MDE) that minimizing

$$\sum_{i=1}^N (\Delta w_i - H_i \phi)' \Omega^{*-1} (\Delta w_i - H_i \phi), \quad (2.8)$$

which yields

$$\hat{\phi} = \left(\sum_{i=1}^N H_i' \Omega^{*-1} H_i \right)^{-1} \left(\sum_{i=1}^N H_i \Omega^{*-1} \Delta w_i \right). \quad (2.9)$$

The GMM estimator breaks down when some of the roots of $|I - \Phi z| = 0$ are equal to unity. To see this, suppose $\Phi = I_m$, then $\Delta w_{it} = \Delta \epsilon_{it}, q_{it}$ are no longer legitimate instruments. On the other hand, both the MLE and MDE always exist and are consistent whether w_{it} contains unit roots or not. The MLE and MDE are asymptotically normally distributed independent of whether w_{it} is (trend) stationary, integrated or cointegrated as $N \rightarrow \infty$ and is asymptotically unbiased whether T is fixed or $\rightarrow \infty$. Moreover, if the roots of $|I - \Phi z| = 0$ lie outside the unit circle, the MLE is asymptotically normally distributed with mean equal to Φ independent of the way N or T go to infinity (Hsiao

and Zhang (2013)). Therefore, the conventional likelihood ratio test statistic or 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 $|I - \Phi z| = 0$ are near unity.

(b) “Heterogeneous” VAR Models

We shall say a VAR model “heterogeneous” if slope coefficients also vary across individuals, $\Phi_i \neq \Phi_j$ if $i \neq j$.

(i) Cross-Sectionally Independent Processes

When ε_{it} is independent over i with mean 0 and unrestricted covariance matrix Ω_i and T is large, applying the least squares method to the system (2.1) equation by equation yields consistent and efficient estimates of $\Phi_i(L)$ and α_i if $T \rightarrow \infty$.

(ii) Cross-Sectionally Dependent Processes

When ε_{it} are cross-sectionally dependent, (2.1) can be put in Zellner’s (1962) seemingly unrelated regression framework if N is fixed and T is large. We can first use the i -th individual’s time series observation to estimate $\Phi_i(L)$ and α_i . Then use the estimated $\hat{\Phi}_i(L)$ and $\hat{\alpha}_i$ to obtain estimated $\hat{\varepsilon}_{it}, i = 1, \dots, N, t = 1, \dots, T$ and estimate $\Omega_{ij} = E(\varepsilon_{it}\varepsilon_{jt}')$ by

$$\hat{\Omega}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt}'. \quad (2.10)$$

Given $\hat{\Omega}_{ij}, i, j = 1, \dots, N$, one can stack all N cross-sectionally equations (2.1) together and apply the feasible generalized least squares estimator to obtain the efficient estimate of $\Phi_i(L)$ and $\alpha_i, i = 1, \dots, N$.

(iii) Global VAR (GVAR)

To obtain efficient estimates of $\Phi_i(L)$ using Zellner (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

$$\Phi_i(L)(w_{it} - \Gamma_i w_{it}^*) = \epsilon_{it}, \quad i = 1, 2, \dots, N, \quad (2.11)$$

where

$$w_{it}^* = \sum_{j=1}^N r_{ij} w_{jt}, \quad (2.12)$$

$$r_{ii} = 0, \sum_{j=1}^N r_{ij} = 1, \text{ and } \sum_{j=1}^N r_{ij}^2 \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (2.13)$$

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 proposed by Pesaran (2004), the global average w_{it}^* is inserted into (2.1) to take account of the cross-sectional dependence. When $w_{i,t-s}^*$ can be treated as weakly exogenous (predetermined), the estimation of (2.11) 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 w_{it}^* hold for practically all countries except for the U.S. because of the U.S.'s dominate position in the world. They also show that (2.11) yields better results than (2.1) when cross-sectional units are correlated.

3. Cointegrated Panel Models and Vector Error Correction²

3.1 Properties of Cointegrated Processes

Many macro and financial data are nonstationary. The non-stationarity of a time series is usually represented by an integrated process of order $d, d \geq 1, I(d)$. A d -th integrated process can be transformed into a stationary process by differencing the variable d -times, $(1 - L)^d$. (e.g. Box and Jenkins (1971)). For instance, suppose all the elements of w_{it} are $I(1)$, then $(1 - L)w_{it}$ become stationary ($I(0)$ processes). However, differencing

²This section is largely drawn from Hsiao (2014 Chapter 10.3)

w_{it} also removes the underlying long-run relations among the elements of w_{it} , which can have important economic implications. If w_{it} are driven by some common nonstationary variables, one notable feature is that linear combinations of w_{it} can remove these *common trends* and become stationary ($I(0)$). Such linear combinations capture the long-run relations among w_{it} and are called “cointegrating” relations.

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

$$w_{jit} \sim I(1), \quad j = 1, \dots, m. \quad (3.1)$$

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

$$C_i w_{it} = u_{it} \sim I(0), \quad (3.2)$$

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

Rewriting w_{it} as the sum of the impact of $(m - r)$ $I(1)$ common trends z_{it} and stationary components, ξ_{it} ,

$$w_{it} = A_i z_{it} + \xi_{it}, \quad i = 1, \dots, N, t = 1, \dots, T. \quad (3.3)$$

where

$$\begin{matrix} z_{it} \\ (m - r) \times 1 \end{matrix} \sim I(1), \quad (3.4)$$

and $A_i = (a'_i)$ is an $m \times (m - r)$ constant matrix. The cointegrating relations (3.2) imply that

$$C_i A_i = 0, \quad i = 1, \dots, N. \quad (3.5)$$

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

w_{it} , (Engle and Granger (1987)),

$$\Delta w_{it} = \Pi_i^* w_{i,t-1} + \varepsilon_{it}, i = 1, \dots, N, \quad (3.6)$$

where $\Delta = (1 - L)$ denotes the first difference operation, $\Delta w_{it} = w_{it} - w_{i,t-1}$, ε_{it} is independent, identically distributed with mean 0 and covariance matrix Ω_{ii} , $\Pi_i^* = -(I_m - \Phi_i)$,

$$\text{rank}(\Pi_i^*) = r. \quad (3.7)$$

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

$$\Pi_i^* = J_i \Lambda_i', \quad i = 1, \dots, m. \quad (3.8)$$

The decomposition is not unique, $\Pi_i^* = J_i^* \Lambda_i^{*'}$, where $J_i^* = \Gamma_i F_i$, $\Lambda_i^* = \Lambda_i F_i'^{-1}$, for any $r \times r$ nonsingular matrix F_i . To uniquely define the cointegrating relations, Phillips (1991) chooses the normalization

$$\Lambda_i' = [I_r, \tilde{\Lambda}_i'], \quad (3.9)$$

where $\tilde{\Lambda}_i$ is an $(m - r) \times r$ constant matrix. The advantage of considering a vector error correction representation rather than PVAR ((2.4)) is that one can simultaneously consider the long-run (equilibrium) relations and short-run dynamics of w_{it} .

If the nonstationarity of w_{it} , $i = 1, \dots, N$, 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 w_{it}, w_{kt} is cointegrated across cross-sectional units. Let w_{kt} denote the $N \times 1$ vector of the k th element of $w_{it}, w_{kit}, (w_{k1t}, \dots, w_{kNt})' = w_{kt}$, and $\xi_{kt} = (\xi_{k1t}, \dots, \xi_{kNt})', \varepsilon_{kt} = (\varepsilon_{k1t}, \dots, \varepsilon_{kNt})'$. Then

$$w_{kt} = A_k z_t + \xi_{kt}, k = 1, \dots, m, \quad (3.10)$$

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

$$C_k A_k = 0, \quad k = 1, \dots, m. \quad (3.11)$$

Then

$$C_k w_{kt} = C_k \xi_{kt} \sim I(0). \quad (3.12)$$

In other words, with nonstationary panel data, there could be cointegration relations both over time and across individuals (Breitung and Pesaran (2008)).

3.2 Estimation

3.2.a “Homogeneous” Cointegrating Relations

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

$$w_{1it} = \tilde{w}'_{it} \beta + \alpha_{1i} + u_{1it}, \quad (3.13)$$

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

$$\hat{\beta}_{\sim} = \left[\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{it} - \bar{\tilde{w}}_i)(\tilde{w}_{it} - \bar{\tilde{w}}_i)' \right]^{-1} \left[\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{it} - \bar{\tilde{w}}_i)(w_{1it} - \bar{w}_{1i}) \right], \quad (3.14)$$

where $\bar{\tilde{w}}'_i = (\bar{w}_{1i}, \bar{\tilde{w}}'_i) = \frac{1}{T} \sum_{t=1}^T w'_{it}$. The least squares estimator converges to $\beta_{\sim}$ in the speed of $T\sqrt{N}$ and is asymptotically normally distributed if $N \rightarrow \infty$. However, the endogeneity and unit-root of $\tilde{w}_{it}$ leads to a nonzero asymptotic bias term of order $\frac{1}{T}$ when $\hat{\beta}_{\sim}$ is multiplied by the scale factor, $T\sqrt{N}$ (Kao and Chiang (2000)). The idea of Park (1992) and Phillips and Hansen (1990) fully modified estimator is suggested to correct the endogeneity effect,

$$w_{1it}^+ = \tilde{w}'_{it} \beta + \alpha_{1i} + u_{1it}^+ \quad (3.15)$$

where

$$w_{1it}^+ = w_{1it} - \Omega_{u\Delta\tilde{w}} \Omega_{\Delta\tilde{w}}^{-1} \Delta \tilde{w}_{it}, \quad (3.16)$$

$$u_{1it}^+ = u_{1it} - \Omega_{u\Delta\tilde{w}} \Omega_{\Delta\tilde{w}}^{-1} \Delta \tilde{w}_{it}, \quad (3.17)$$

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

$$\hat{\beta}_{\tilde{FM}} = \left[\sum_{i=1}^N \sum_{t=1}^T (\tilde{w}_{it} - \tilde{w}_i)(\tilde{w}_{it} - \tilde{w}_i)' \right]^{-1} \left[\sum_{i=1}^N \left(\sum_{t=1}^T (\tilde{w}_{it} - \tilde{w}_i)w_{1it}^+ - T\Delta_{\Delta\tilde{w}u} \right) \right], \quad (3.18)$$

where $\Delta_{\Delta\tilde{w}u} = \sum_{j=0}^{\infty} E(\Delta\tilde{w}_{i,t-j}u_{1it})$. The correction terms $\Omega_{u\Delta\tilde{w}}, \Omega_{\Delta\tilde{w}}, \Delta_{\Delta\tilde{w}u}$ can be replaced by their consistent estimator. Kao and Chiang (2000) show that $T\sqrt{N}(\hat{\beta}_{\tilde{FM}} - \beta)$ is asymptotically normally distributed with mean 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 (Phillips and Loretan (1991), Saikkonen (1991)),

$$w_{1it} = \tilde{w}'_{it}\beta + \sum_{j=-q}^q \Delta\tilde{w}'_{i,t-j}\gamma_j + \alpha_{1i} + \tilde{u}_{1it}. \quad (3.19)$$

where Westerlund (2005) suggests a data based choice of truncation lag order q . Kao and Chiang (2000) show that the within estimator of (3.19) has the same asymptotic distribution as the panel fully modified within estimator. The Monte Carlo studies conducted by Kao and Chiang (2000) show that the (lead-lag adjusted) within estimator of (3.19) performs better than the panel fully modified estimator, probably because of the failure to obtain good estimates of $\Omega_{u\Delta\tilde{w}}, \Omega_{\Delta\tilde{w}}$, etc.

When there are more than one linearly independent cointegration relations among the elements of w_{it} , in principle, one can follow Phillips (1991) to normalize the r linearly independent cointegration relations in the form of (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 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 (2.6) gives $\Pi = -(I_m - \Phi)$. If $\text{rank}(\Pi) = r(> 1)$, then we can write $\Pi = J\Lambda'$ where J and Λ are $m \times r$ with $\text{rank } r$. Then the cointegrating matrix Λ' is equal to

$$\Lambda' = (J'J)^{-1}J'\Pi. \quad (3.20)$$

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

$$\tilde{\Lambda}' = (\Pi_1'\Pi_1)^{-1}\Pi_1'\Pi, \quad (3.21)$$

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_j 's has not taken into account the reduced rank restrictions on $\Pi = J\Lambda'$. One way to obtain efficient estimator of Φ_j is to apply constrained GMM or constrained MDE or constrained (transformed) MLE by minimizing the quadratic form of the moment conditions (2.5) or (2.8) or maximizing (2.7) subject to

$$\Lambda = H\delta_r + B_r, \quad (3.22)$$

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

3.2.b. Heterogeneous Cointegrating System

If individual units are independent across i , the Johansen (1991) method can be applied to the time series data for each i to obtain the maximum likelihood estimator (MLE) of (3.6). There is no need for pooling.

When individual units are correlated, if N is fixed and T is large the covariance between ϵ_{it} and ϵ_{jt} , Ω_{ij} , can simply be estimated by

$$\hat{\Omega}_{ij} = \frac{1}{T} \sum_{t=1}^T \epsilon_{it} \epsilon_{jt}', \quad (3.23)$$

where $\hat{\xi}_{it}$ can be constructed from (3.6) using the Johansen MLE estimates. Stacking the N cross-sectionally observed w_{it} one after another yields

$$\Delta w_t = \tilde{\Pi} w_{t-1} + u_t, \quad (3.24)$$

where $w'_t = (w'_{1t}, \dots, w'_{Nt})$,

$$\tilde{\Pi} = \begin{pmatrix} \Pi_1^* & 0 & \cdot & \cdot & 0 \\ 0 & \Pi_2^* & \cdot & \cdot & 0 \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ 0 & & & & \Pi_N^* \end{pmatrix}, \quad (3.25)$$

and

$$\epsilon_t = (\epsilon'_{1t}, \dots, \epsilon'_{Nt})'.$$

Under the assumption that ϵ_t is independently normally distributed with mean 0 and covariance matrix

$$\tilde{\Omega}_{Nm \times Nm} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdot & \cdot & \Omega_{1N} \\ \cdot & \Omega_{22} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \Omega_{N1} & \cdot & \cdot & \cdot & \Omega_{NN} \end{pmatrix}, \quad (3.26)$$

the log-likelihood function of Δw_t is proportional to

$$-\frac{T}{2} \log |\tilde{\Omega}| - \frac{1}{2} tr \left[\Omega^{-1} \sum_{t=1}^T (\Delta w_t - \tilde{\Pi} w_{t-1})(\Delta w_t - \tilde{\Pi} w_{t-1})' \right]. \quad (3.27)$$

The MLE of Ω and Π are the solutions that simultaneously satisfy

$$\hat{\hat{\Omega}} = \frac{1}{T} \sum_{t=1}^T (\Delta w_t - \hat{\hat{\Pi}} w_{t-1})(\Delta w_t - \hat{\hat{\Pi}} w_{t-1})', \quad (3.28)$$

and

$$\begin{pmatrix} \text{vec } (\hat{\hat{\Pi}}_1^{*'}) \\ \text{vec } (\hat{\hat{\Pi}}_2^{*'}) \\ \cdot \\ \cdot \\ \text{vec } (\hat{\hat{\Pi}}_N^{*'}) \end{pmatrix} = \left(\sum_{t=1}^T \hat{\hat{\Omega}}^{-1} \otimes w_{t-1} w'_{t-1} \right)^{-1} \left[\sum_{t=1}^T \left(\hat{\hat{\Omega}}^{-1} \otimes w_{t-1} \right) \text{vec } (\Delta w_t) \right]. \quad (3.29)$$

Conditional on $\hat{\Omega}$, (3.19) is in the form of Zellner's (1962) seemingly unrelated regression estimator. Sequentially iterating between (3.28) and (3.29) until convergence will result in the MLE of Π_i^* and $\tilde{\Omega}$.

However, if w_{it} are cointegrated, then Π_i is subject to the restrictions of the form (3.8). Substituting (3.28) into (3.29) and making use of the relations (Magnus and Neudeck (1999, p.31))

$$\begin{aligned} \text{vec} (J_i \Lambda_i') &= (I_r \otimes J_i) \text{vec} (\Lambda_i') \\ &= (\Lambda_i \otimes I_m) \text{vec} (J_i), \end{aligned} \quad (3.30)$$

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

$$\begin{aligned} \begin{pmatrix} \text{vec} (\hat{\Lambda}_1) \\ \vdots \\ \text{vec} (\hat{\Lambda}_N) \end{pmatrix} &= \{\hat{Q}' [\hat{\Omega}^{-1} \otimes (\sum_{t=1}^T w_{t-1} w_{t-1}')] \hat{Q}\}^{-1} \\ &\quad \cdot \{\hat{Q}' [\hat{\Omega}^{-1} \otimes I_{Nm}] \text{vec} (\sum_{t=1}^T w_{t-1} \Delta w_t')\}, \end{aligned} \quad (3.31)$$

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

$$\begin{aligned} \begin{pmatrix} \text{vec} (\hat{J}_1') \\ \vdots \\ \text{vec} (\hat{J}_N') \end{pmatrix} &= \{\hat{P}' (\hat{\Omega}^{-1} \otimes \sum_{t=1}^T w_{t-1} w_{t-1}') \hat{P}\}^{-1} \\ &\quad \cdot \{\hat{P}' (\hat{\Omega}^{-1} \otimes I_{Nm}) \text{vec} (\sum_{t=1}^T w_{t-1} \Delta w_t')\}, \end{aligned} \quad (3.32)$$

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

$$\hat{Q} = \left[(d_1 \otimes \hat{J}_1) \otimes (d_1 \otimes I_r), \dots, (d_N \otimes \hat{J}_N) \otimes (d_N \otimes I_r) \right], \quad (3.33)$$

$$\hat{P} = \left[(d_1 \otimes I_m) \otimes (d_1 \otimes \hat{\Lambda}_1), \dots, (d_N \otimes I_m) \otimes (d_N \otimes \hat{\Lambda}_N) \right], \quad (3.34)$$

d_j denotes the j th 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 $\hat{\Omega}^{(0)}$ and $\hat{J}_i^{(0)}, i = 1, \dots, N$ from (3.28) and (3.20), where the initial estimates of $\hat{J}_i^{(0)}$ are simply the first r columns of $\hat{\Pi}_i^{*(0)}$ under the normalization (3.9).

2. Construct estimates of $\Lambda_i, i = 1, \dots, N$ from (3.31).
3. Revise the estimate of Π_i^* from (3.29) and revise the estimate of Ω from the revised estimate of $\hat{\Pi}_i^*$ using the formula (3.28).
4. Revise the estimate of J_i given the revised estimator of Λ_i and Ω using (3.32).
5. Iterate steps 1 to 4 until the solution converges.

The MLE of $\Lambda_i, i = 1, \dots, N$ 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

$$\text{vec}(\Lambda) = \left(\tilde{Q}' \left(\hat{\Omega}^{-1} \otimes \sum_{t=1}^T \underline{w}_{t-1} \underline{w}_{t-1}' \right) \tilde{Q} \right)^{-1} \left(\tilde{Q}' \left(\hat{\Omega}^{-1} \otimes I_{Nm} \right) \text{vec} \left(\sum_{t=1}^T \underline{w}_{t-1} \Delta \underline{w}_t' \right) \right). \quad (3.35)$$

where $\tilde{Q}$ is of the form of $\hat{Q}$ with $\hat{J}_i$ replaced by i -th row of $I_N, i = 1, \dots, N$. Because Λ_i or Λ is asymptotically mixed normally distributed, one can use the likelihood ratio test statistic to test the null hypothesis

$$H_0 : \Lambda_1 = \Lambda_2 = \dots = \Lambda_N = \Lambda$$

$$\text{versus } H_1 : \Lambda_i \neq \Lambda_j.$$

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

When individual units contain an individual specific effects α_i and p_i in (2.2) different from zero, one can follow Johansen (1991) to concentrate out constants and Φ_{is}^* by regressing $\Delta \underline{w}_{it}$ and $\underline{w}_{i,t-1}$ on the constants and $\Delta \underline{w}_{i,t-s}, s = 1, \dots, p_i$, respectively, to obtain $\Delta \tilde{\underline{w}}_{it}$ and $\tilde{\underline{w}}_{i,t-1}$, then proceed to estimate J_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 $\left(\Delta \tilde{\underline{w}}_{it} - \hat{J}_i \hat{\Lambda}_i' \tilde{\underline{w}}_{i,t-1} \right)$ on constants, $\Delta \tilde{\underline{w}}_{i,t-p_i-1}, \dots, \Delta \tilde{\underline{w}}_{i,t-1}$ 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, Shuermann and Weiner (2004) suggest using a global VAR (GVAR) ((2.11) - (2.13)) to filter out cross-sectional dependence. When w_{it}^* can be treated as weakly exogenous, the system (3.6) subject to the rank condition (3.9) for each i can be estimated using standard time series estimator techniques (e.g. Pesaran, Shin and Smith (2000)).

4. Unit Root and Cointegration Tests

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 interests to economists. If observed data are nonstationary, or contain unit roots, standard estimators will have nonstandard distributions as $T \rightarrow \infty$. 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 (1986), 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 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 under the null of unit root (no cointegration for the residual based tests of cointegration) or under the null of stationarity (cointegration for the residual based tests of cointegration) when $N \rightarrow \infty$ with either T fixed or $\rightarrow \infty$. When the null is formulated as containing a unit root, it is usually

formulated in the form of

$$\Delta y_{it} = \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} + \alpha_i + \epsilon_{it}, \quad (4.1)$$

with the null,

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_N = 0. \quad (4.2)$$

Under the assumption that ϵ_{it} are independently, identically distributed (i.i.d.) across i and over t , the test statistics are usually based on the average of augmented Dickey-Fuller test statistics (1979, 1981) (e.g. Levin, Lin, Chu (2002) (LL), Im, Pesaran and Shin (2003) (IPS), Westerlund and Larsson (2012), etc.).

When ϵ_{it} are cross-sectionally dependent, the tests based on i.i.d. are subject to severe size distortion (e.g. Breitung and Das (2005)). 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 (4.1). As long as $F(\cdot)$ is regularly integrable, 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

$$Z_i = \frac{\hat{\gamma}_i - 1}{s(\hat{\gamma}_i)}, \quad (4.3)$$

will converge to a standard normal when $T_i \rightarrow \infty$ 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 i th unit because $F(y_{i,t-1})$ tends to zero as $y_{i,t-1}$ tends to $\pm\infty$, the nonstationarity of $y_{i,t-1}$ is eliminated. Moreover, 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

$$S_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i \quad (4.4)$$

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,

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

$$y_{it} = \sum_{j=1}^k b_{ij} f_{jt} + u_{it}, \quad (4.5)$$

where $\underline{b}_i = (b_{i1}, \dots, b_{ik})'$ is a $k \times 1$ vector of constants, u_{it} is independent across i with mean zero. Bai and Ng (2004, 2010) and Moon and Perron (2004) propose to test unit roots in both $\underline{f}_t$ and u_{it} . The factors are estimated from k principal components of Δy_{it} . Augmented Dickey-Fuller tests (ADF) are then applied to test if $\underline{f}_t$ 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 Maddala and Wu (1999) P_λ or Choi (2001) P_m test is then applied to the defactored observations, $u_{it}, (y_{it} - \underline{b}'_i f_t)$.

If $b_i \neq b_j$ and $k = 1$, Pesaran (2007) suggests augmenting (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 $\Delta \bar{y}_{t-j} = \frac{1}{N} \sum_{i=1}^N \Delta y_{i,t-j}$,

$$\begin{aligned} \Delta y_{it} = & \alpha_i + \delta_i t + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \phi_{i\ell} \Delta y_{i,t-\ell} \\ & + c_i \bar{y}_{t-1} + \sum_{\ell=1}^{p_i} d_{i\ell} \Delta \bar{y}_{t-\ell} + e_{it}, \end{aligned} \quad (4.6)$$

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 Fully 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 w_{it} , Kapetanios, Pesaran and Yamagata (2011), Pearan, Smith and Yamagata (2012) suggest further augmenting (4.1) by cross-sectional mean of w_t and $w_{t-\ell}$,

$$\begin{aligned} \Delta y_{it} = & \alpha_i + \delta_{it} + \gamma_i y_{i,t-1} + \sum_{\ell=1}^{p_i} \Delta y_{i,t-\ell} \\ & + \sum_{\ell=1}^{p_i} d_{i\ell} \Delta \bar{y}_{t-\ell} \\ & + W_t \underline{\theta} + \epsilon_{it}, \quad i = 1, \dots, N \end{aligned} \quad (4.7)$$

where $W_t = (\Delta w'_t, \Delta w'_{t-1}, \dots, \Delta w'_{t-p_i-1}, w'_{t-1})$ and ϱ is an $mp_i \times 1$ vector of constants. Pesaran, Smith and Yamagata (2012) suggest either constructing an IPS type test statistic by taking the average of t -statistics for $\gamma_i, i = 1, \dots, N$ or taking the average of the Sargan and Bhargava test statistic (SB) (1983),

$$SB_i = T^{-2} \sum_{t=1}^T s_{it}^2 / \hat{\sigma}_i^2, \quad i = 1, \dots, N, \quad (4.8)$$

where

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

$\hat{\epsilon}_{it}$ is the least squares residual of (4.7), and K^* is the number of the unknown constants in (4.7). Each of the SB_i converges to a functional of Brownian motion which is independent of the factors as well as their loadings. Pesaran, Smith and Yamagata (2012) have provided the critical values of the augmented IPS and SB tests for $K=0,1,2,3$ and $p_i=0,1,\dots, 4$ 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. They suggest to group 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 $T \rightarrow \infty$ and N fixed, the subsample critical values will converge in probability to the true critical values. The Monte Carlo simulation conducted by Choi and Chue (2007) show that the size of the subsample test is indeed robust against various forms of cross-section dependence.

When the null is formulated as of stationarity. The test statistics is usually constructed based on the average of Sargan-Bhargava (1983) statistics, which is formulated in the form of partial sum of y_{it} relative to the variance of Δy_{it} ,

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^N (K_i - E(K_i)), \quad (4.9)$$

where

$$K_i = \frac{1}{T^2 \sigma_i^2} \sum_{t=1}^T S_{it}^2, \quad S_{it} = \frac{1}{\sqrt{T}} \sum_{s=1}^t y_{is},$$

and σ_i^2 is the variance of Δy_{it} , (e.g. Kwiaskowski, et.al. (1992), Hadri (2000), Hadri and Larsson (2005)).

Simulations conducted by Choi (2001, 2002) show that the empirical size of all the unit root tests is reasonably close to the nominal size and they all become more powerful as N increases. For further details, see Hsiao (2014, Chapter 10.4).

4.2 Cointegration Tests

The existence of cointegration can be tested if the residuals of regressing one element of w_{it} on the rest elements of w_{it} contain a unit root or not. If the residuals are stationary, 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 of no cointegration, time series regression of the slope coefficients $\beta_{\sim}$ fails to converge (Phillips (1987)). On the other hand, panels with large N and large T can yield convergent estimate of β . If there is no cointegration, Kao (1999), using a sequential limit argument (first $T \rightarrow \infty$, followed by $N \rightarrow \infty$), has shown that the least squares estimator of (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 (3.18) or the within estimator for the lead-lag adjusted regression model (3.19) 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 time series dimension, the properly scaled cross-sectional average of such a statistic is asymptotically normally distributed when $N \rightarrow \infty$. 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) propose to test the null of cointegration using the statistic

$$\overline{LM} = \frac{\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T S_{it}^{+2}}{s^{+2}}, \quad (4.10)$$

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 (3.13) with β estimated by either (3.18) or applying the within estimator to (3.19) and $\hat{u}_{1it}^+$ is the estimated residual of (3.17). 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 (1999) proposes a modified Dicky-Fully 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 “heterogeneous” alternative. For excellent surveys of panel unit root and cointegration tests, see Baltagi and Kao (2000), Banerjee (1999), Breitung and Pesaran (2008), Choi (2006), etc.

The residual-based tests can only tell if there exist cointegration relations among elements of w_{it} , they cannot tell the cointegration rank. For the “homogeneous” PVAR model, a test of the cointegration rank can be conducted by maximizing the (transformed) likelihood function (2.4) subject to either

$$\Lambda_r = H_r \delta_r + B_r, \quad (4.11)$$

or

$$\Lambda_{r+q} = H_{t+1} \delta_{r+1} + B_{r+1}, \quad (4.12)$$

where H_r and H_{r+1} are, respectively, $m \times (m-r)$ and $m \times (m-r-1)$ matrices with known elements, B_r and B_{r+1} are, respectively, $m \times r$ matrices with known elements, δ_r and δ_{r+1}

are, respectively, $r \times m$ and $(r+1) \times m$ matrices with unknown elements. When T is fixed and $N \rightarrow \infty$, 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 coefficient in Π).

5. Dynamic Simultaneous Equations Models

The discussions in previous sections, in particular, the panel VAR model (2.1), can be considered as a reduced form specification of the model

$$\Phi(L)w_{it} = \alpha_i + \varepsilon_{it}, \quad (5.1)$$

where

$$\Phi(L) = \Phi_0 + \Phi_1 L + \dots + \Phi_p L^p, \quad (5.2)$$

and

$$\Phi_0 \neq I_m. \quad (5.3)$$

For simplicity, we maintain the assumption that ε_{it} is independently, identically distributed (i.i.d) across i and over t with covariance matrix Ω . When $\Phi_0 \neq I_m$, in addition to dynamic dependence, (5.1) also introduces contemporaneous dependence among elements of w_{it} . One will need prior information to identify an equation in the system, say the g th. If we assume that the prior information takes the form of exclusion restrictions, i.e. some variables are excluded from the g th equation. Let $[\Phi_{0g}, \dots, \Phi_{pg}]$ be the matrix formed from the columns of $[\Phi_0, \dots, \Phi_p]$ that are zero on the g -th row. The presence of α_i does not lead to any change in the necessary and sufficient condition for the identification of the g th equation of (5.1), i.e. (e.g. Hsiao (1983), Hsiao and Zhou (2014)),

$$\text{rank } [\Phi_{0g}, \dots, \Phi_{pg}] = m - 1. \quad (5.4)$$

When an equation in (5.1) is identified, one can estimate that equation either through a method of moments approach or a likelihood approach. As an illustration. let $w'_{it} =$

$(\underline{y}'_{it}, \underline{x}'_{it})$, where $\underline{y}_{it}$ is $G \times 1$ and $\underline{x}_{it}$ is $(m - G) \times 1$. Assume $p = 1$ and the first equation of the system (5.1) takes the form

$$y_{1it} = Y'_{it}\phi_{11,0} + Y'_{i,t-1}\phi_{11,1} + y_{1i,t-1}\gamma + \underline{x}'_{it}\beta + \alpha_{1i} + \epsilon_{1it} \quad (5.5)$$

where (y_{1it}, Y'_{it}) denotes the sub-vector of $\underline{y}_{it}$ that appears in the first equation. For notational ease we assume $w'_{i0} = (\underline{y}'_{i0}, \underline{x}'_{i0})$ are observable.

Taking the first difference of (5.5) to eliminate the individual-specific effects, α_{1i} , yields

$$\begin{aligned} \Delta y_{1it} &= \Delta Y'_{it}\phi_{11,0} + \Delta Y'_{i,t-1}\phi_{11,1} + \Delta y_{1i,t-1}\gamma + \Delta \underline{x}'_{it}\beta + \Delta \epsilon_{1it} \\ &= \Delta \underline{z}'_{it}\theta + \Delta \epsilon_{1it}, \quad \begin{matrix} i = 1, \dots, N, \\ t = 2, \dots, T, \end{matrix} \end{aligned} \quad (5.6)$$

where $\Delta = (1 - L)$, L is the lag operator that shifts the variable back by one period, $Ly_{1it} = y_{1i,t-1}$, and $\underline{z}'_{it} = (Y'_{it}, Y'_{i,t-1}, y_{1i,t-1}, \underline{x}'_{it})$, $\theta = (\phi'_{11,0}, \phi'_{11,1}, \gamma, \beta')$. Under the assumption that ϵ_{it} is independently, identically distributed and $E(\underline{x}_{it}\epsilon'_{it}) = 0$, we have

$$Eq_{it}\Delta\epsilon_{it} = 0, \quad (5.7)$$

where $q_{it} = (y_{1i0}, Y'_{i0}, \dots, y_{1i,t-2}, Y'_{1i,t-2}, \underline{x}'_{i0}, \dots, \underline{x}'_{iT})'$. Let

$$W_i = \begin{bmatrix} q_{i2} & 0 & 0 & \cdot & 0 \\ 0 & q_{i3} & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & q_{iT} \end{bmatrix} \quad (5.8)$$

be the $Q \times (T-1)$ block dimensional matrix, where Q denotes the dimension of $(q'_{i2}, \dots, q'_{iT})'$.

Then

$$E(W_i\Delta\epsilon_{1i}) = 0, \quad i = 1, \dots, N, \quad (5.9)$$

where $\Delta\epsilon'_{1i} = (\Delta\epsilon_{1i2}, \dots, \Delta\epsilon_{1iT})$ with $E(\Delta\epsilon_{1i}) = 0$ and $E(\Delta\epsilon_{1i}\Delta\epsilon'_{1i}) = \sigma_1^2\tilde{A}$, $\sigma_1^2 = E(\epsilon_{1it}^2)$,

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

The Arellano-Bond (1991) GMM estimator of θ is to find $\hat{\theta}$ that minimizes

$$\left(\sum_{i=1}^N \Delta \epsilon'_{1i} W'_i \right) \left(\sum_{i=1}^N W_i \tilde{A} W'_i \right)^{-1} \left(\sum_{i=1}^N W_i \Delta \epsilon_{1i} \right). \quad (5.10)$$

When T is fixed and $N \rightarrow \infty$, $\sqrt{N}(\hat{\theta}_{GMM} - \theta)$ is asymptotically normally distributed with mean 0 and covariance matrix $\sigma_1^2 \left(\frac{1}{N} \sum_{i=1}^N W_i \tilde{A} W'_i \right)^{-1}$. When T is also large and $\frac{T}{N} \rightarrow c \neq 0 < \infty$ as $N, T \rightarrow \infty$, the process of removing individual-specific effects creates a second order bias that becomes serious for large T because

$$E \left[\begin{pmatrix} \Delta Y'_i \\ \Delta Y'_{i,-1} \\ \Delta y_{1i,01} \end{pmatrix} W'_i (W_i \tilde{A} W'_i)^{-1} W_i \Delta \epsilon_{1i} \right] \neq 0, \quad (5.11)$$

and is of order $\log T$. Akashi and Kunitomo (2012, 2014) show that the Arellano-Bond type GMM is inconsistent and asymptotically biased. Monte Carlo studies conducted by Akashi and Kunitomo (2012, 2014) show that the GMM estimator is badly biased when N and T are both large.

Alternatively, Hsiao and Zhou (2014) show that the likelihood approach with properly formulated initial value distribution for y_{i0} yields an estimator that is asymptotically unbiased independent of the way N or T or both tend to infinity. However, maximizing the (transformed) likelihood function of $(\Delta y_{i2}, \dots, \Delta y_{iT}), i = 1, \dots, N, t = 2, \dots, T$ under the false assumption that Δy_{i1} are fixed constants is asymptotically unbiased only if N is fixed and T is large. If both N and T are large, the false (quasi) maximum likelihood estimator is asymptotically biased of order $\sqrt{\frac{N}{T}}$. Monte Carlo studies conducted by Hsiao and Zhou (2014) show that the reliability of statistical inference depends critically on the estimator that is asymptotically unbiased.

6. Macro Policy Evaluation

Denote the outcome of the i th unit that received the treatment at time t by y_{it}^{1*} and the outcome not receiving the treatment by y_{it}^{0*} . If both y_{it}^{1*} and y_{it}^{0*} are observable, then treatment effect is simply

$$\Delta_{it} = y_{it}^{1*} - y_{it}^{0*}. \quad (6.1)$$

One of the issues of evaluating the effectiveness of a macro policy is that we do not simultaneously observe y^{1*} and y^{0*} . Often, we only have time series data of the outcomes before the implementation of a policy and the outcomes after the implementation of a policy. However, many conditioning events have changed over time. For instance, one knows the outcomes of the U.S. economy before the 2008 financial crisis and the outcomes after the implementation of Fed's quantity easing policy, but one does not know what would have happened to the U.S. economy in the absence of Fed's quantity easing policy. The parametric approach to construct the counterfactuals, y_{it}^{0*} , after the year 2008 requires the conditional variables not affected by the changes in policy. The availability of panel data provides the possibility of measuring policy impact without relying on economic theory or macroeconometric model. Hsiao, Ching and Wan (2012) propose a *measurement without theory* approach to evaluate the impact of macro policies by exploiting the dependence among cross-sectional units and noting that panel data not only contain observations of those before and after the implementation of a policy, but also contain time series information on those not subject to the policy intervention.

Let d_{it} be 1 if the i th unit received the treatment at time t and 0 if not. Then the observed y_{it} takes the form

$$y_{it} = d_{it}y_{it}^{1*} + (1 - d_{it})y_{it}^{0*}. \quad (6.2)$$

In other words, at time t one either observes y_{it}^{1*} or y_{it}^{0*} , not both. To measure the impact of a policy intervention at time t , one either has to construct the counterfactual of y_{it}^{1*} or y_{it}^{0*} . Suppose all N units at time t , $\underline{y}_t = (y_{1t}, \dots, y_{Nt})$ did not receive the treatment for $t = 1, \dots, T_1$, i.e. $\underline{y}_t = \underline{y}_t^{0*}$, but from time period $T_1 + 1$ onwards, the first unit received treatment, $y_{1t} = y_{1t}^{1*}, t = T_1 + 1, \dots, T$, while the rest of individuals did not, $y_{it} = y_{it}^{0*}, t = 1, \dots, T$ for $i = 2, \dots, N$. As long as

$$E(\epsilon_{it} \mid d_{1t}) = 0, \quad i = 2, \dots, N, \quad (6.3)$$

one can write

$$\begin{aligned} y_{1t}^{0*} &= E(y_{1t}^{0*} | \tilde{y}_t) + \eta_{1t}, \quad t = 1, \dots, T, \\ &= a + \underline{c}' \tilde{y}_t + \eta_{1t}, \end{aligned} \tag{6.4}$$

where $\tilde{y}_t = (y_{2t}, \dots, y_{Nt})'$ and $E(\eta_{1t} | \tilde{y}_t) = 0$. It is shown by Hsiao, Ching and Wan (2012) that minimizing

$$\frac{1}{T_1} (\underline{y}_1^0 - \underline{e}a - Y\underline{c})' A (\underline{y}_1^0 - \underline{e}a - Y\underline{c}) \tag{6.5}$$

yields consistent estimates of a and $\underline{c}$, where $\underline{y}_1^0 = (y_{11}, \dots, y_{1T_1})'$, $\underline{e}$ is a $T_1 \times 1$ vector of 1's, Y is a $T_1 \times (N - 1)$ matrix of T_1 time series observations of $\tilde{y}_t$, and A is a $T_1 \times T_1$ positive definite matrix. From the estimates $(\hat{a}, \hat{\underline{c}}')$, one can construct the predicted value of the first unit in the absence of treatment, y_{1t}^{0*} , by

$$\hat{y}_{1t}^{0*} = \hat{a} + \hat{\underline{c}}' \tilde{y}_t, \quad t = T_1 + 1, \dots, T. \tag{6.6}$$

The treatment effect on the first unit can then be estimated by

$$\hat{\Delta}_{1t} = y_{1t} - \hat{y}_{1t}^{0*}, \quad t = T_1 + 1, \dots, T. \tag{6.7}$$

Cross-sectional data provide measurement of policy intervention as a once-and-for-all impact. Panel data allow the policy impact to be evolutionary. If Δ_{1t} is serially correlated, but stationary, one can further model the time-varying treatment effects by an autoregressive moving average model using Box-Jenkins (1970) methodology

$$a(L) \hat{\Delta}_{1t} = \mu + \theta(L) \eta_t \tag{6.8}$$

where L is the lag operator, η_t is an i.i.d. process with zero mean and constant variance and the roots of $\theta(L) = 0$ lie outside the unit circle. If the roots of $a(L) = 0$ all lie outside the unit circle, the treatment effect is stationary, and the long-term treatment effect is

$$\Delta_1 = a(L)^{-1} \mu = \mu^*. \tag{6.9}$$

Condition (6.3) makes no claim about the relationship between d_{1t} and ϵ_{1t} . They can be correlated or uncorrelated. Here, all we need is that the j th unit's idiosyncratic components are independent of d_{1t} for $j \neq 1$.

7. Aggregation and Predictions

Macro-econometrics modeling depends on the existence of stable relations among aggregate variables. Aggregation is a good simplification tool as long as micro relations are “homogeneous” (i.e. have identical parameter values) or the distribution of micro variables remains constant over time (e.g. Stocker (1993), Theil (1954)). However, if the micro relation is dynamic and “heterogeneous”, the dynamics of aggregate variables can be very different from the dynamics of disaggregate variables (e.g. Lewbel (1994) Zaffaroni (2001)). For instance, consider a simple dynamic relation for micro units,

$$y_{it} = \gamma_i y_{i,t-1} + \mathbf{x}'_{it} \beta_i + \alpha_i + u_{it}, \quad |\gamma_i| < 1, \quad i = 1, \dots, N. \quad (7.1)$$

where u_{it} is independently, identically distributed with mean 0 and constant variance σ_u^2 .

Then the long-run relation between y and $\mathbf{x}$ for each i is given by

$$y_{it} = \frac{1}{1 - \gamma_i} \mathbf{x}'_{it} \beta_i + \frac{\alpha_i}{1 - \gamma_i} + \frac{1}{1 - \gamma_i} u_{it}. \quad (7.2)$$

Let $\bar{y}_t = \sum_{i=1}^N y_{it}$, $\bar{\mathbf{x}}_t = \sum_{i=1}^N \mathbf{x}_{it}$, then

$$\bar{y}_t - \bar{\mathbf{x}}'_t \mathbf{b} - \alpha = v_t \quad (7.3)$$

has a stable relation (i.e. v_t is covariance stationary) if and only if either of the following conditions hold (Hsiao, Shen and Fujiki (2004)):

$$(i) \quad \frac{1}{1 - \gamma_i} \beta_i = \frac{1}{1 - \gamma_j} \beta_j \text{ for all } i \text{ and } j, \quad (7.4)$$

or

$$(ii) \text{ if } \frac{1}{1 - \gamma_i} \beta_i \neq \frac{1}{1 - \gamma_j} \beta_j,$$

then $\underline{x}'_t = (\underline{x}'_{1t}, \dots, \underline{x}'_{Nt})$ must lie on the null space of D for all t , where

$$D' = \left(\frac{1}{1 - \gamma_1} \beta'_1 - \underline{b}', \frac{1}{1 - \gamma_2} \beta'_2 - \underline{b}', \dots, \frac{1}{1 - \gamma_N} \beta'_N - \underline{b}' \right).$$

Panel data provide information on micro units over time that can be used to explore if conditions for stable aggregate relations hold. For instance, Hsiao, Shen and Fujiki (2004) find that the estimated aggregate relations between Japan's money demand, GDP and bond rate are unstable and sensitive to the time period covered. Depending on the sample period used, the estimated relationships are either of wrong sign or statistically insignificant or yield unbelievable magnitudes. They find the estimated long-run income elasticities are 75.23 for M1 and 11.04 for M2, respectively. They attribute these incredible results to the aggregation of “heterogeneous” micro units.

If micro units are “heterogeneous”, two issues arise. First, shall one focus the analysis on micro units? Second, to predict the aggregate outcomes, shall one rely on the summation of micropredictions or rely on the prediction generated from the aggregate variables? The problem with micro analysis is that there are numerous micro relations to estimate while decision makers are only interested in the average outcomes. Moreover, there may not have enough time series observations to provide reliable estimates of micro relations. It appears that a random coefficients framework assuming (γ_i, β'_i) are independently distributed with constant mean $(\bar{\gamma}, \bar{\beta}')$ and finite covariance matrix (e.g. Hsiao (2014, Chapter 6), Hsiao and Pesaran (2008)) provide a convenient framework to reconcile the issue of “heterogeneity” in aggregation while providing reasonable estimates of average relation. As a matter of fact, Hsiao, Shen and Fujiki (2005) using a random coefficient framework for the analysis of Japanese disaggregate data (prefecture panel data) find that there is a stable average relation between money demand and income. The estimated short-run income elasticity for M1 and M2 is 0.88 and 0.47, respectively. The long-run income elasticity is 2.56 for M1 and 1.01 for M2. These results appear to be consistent with economic theory and the broadly observed facts about Japan. The average growth rate for M2 in the 1980s is about 9.34%. The inflation rate is 1.98%. The real M2 growth rate is 7.36%. The real growth

rate of GDP during this period is 4.13%. Taking account the impact of 5-year bond rate fell from 9.332% at 1980.I to 5.767 at 1989.IV, the results are indeed very close to the estimated long-run income elasticities based on disaggregate data analysis.

If “heterogeneity” is indeed present in micro units, then shall we predict the aggregate outcome based on the summation of estimated micro relations or shall we predict the aggregate outcomes based on the estimated aggregate relations? Unfortunately, there is not much work on this specific issue. In choosing between whether to predict aggregate variables using aggregate (H_a) or disaggregate equations (H_d), Griliches and Grunfeld (1960) suggest using the criterion of:

$$\text{Choose } H_d \text{ if } \varepsilon_d' \varepsilon_d < \varepsilon_a' \varepsilon_a, \text{ otherwise choose } H_a \quad (7.5)$$

where ε_d and ε_a are the estimates of the errors in predicting aggregate outcomes under H_d and H_a , respectively. However, if cross-sectional units are cross-correlated, simple aggregation of individual unit forecasts may not be the optimal way of predicting aggregate (or average) outcomes. For a discussion of optimal way of combining individual forecasts and related literature, see e.g. Hsiao and Wan (2014).

8. Concluding Remarks

This paper reviews the panel methodology to control the impact of unobserved heterogeneities among cross-sectional units to obtain valid inference for a dynamic system of stable relations, from both reduced form and structural form perspectives. However, many challenging methodological issues remain such as: (i) statistical properties of multi-dimensional statistics (Bai and Silverstein (2004), Phillips and Moon (1999, 2000a,b)); (ii) nonlinear system; (iii) dynamic system with unobserved interactive effects (dynamic factor models) (e.g. Bai and Ng (2004), Phillips and Sul (2003) Stock and Watson (2002, 2005)); (iv) fixed versus random effects inference with either N or T or both are large, etc. In addition, macro-econometric modeling depends on the existence of stable relations among aggregate variables. Summing the micro variables to yield aggregate variables does not

lead to the distortion of fundamental economic relations only if certain “homogeneity” conditions among micro units hold. Panel data focus on individual outcomes. Information on disaggregate outcomes provide the possibility of investigating the source of discrepancies between the aggregate analysis and disaggregate analysis and find ways to reconcile the discrepancies (e.g. Hsiao, shen and Fujiki (2005)).

REFERENCES

- Akashi, K., and N. Kunitomo (2012). “Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation”, *Journal of Econometrics*, 166, 167-183.
- _____(2014). “The Limited Information Maximum Likelihood Approach to Dynamic Structural Equation Models”, *Annales of the Institute of Mathematical Statistics* (forthcoming).
- Alvarez, J., and M. Arellano (2003). “The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators,” *Econometrica*, 71, 1121-1159.
- Anderson, T.W. and C. Hsiao (1981), “Estimation of Dynamic Models with Error Components,” *Journal of the American Statistical Association*, 76, 598-606.
- _____(1982). “Formulation and Estimation of Dynamic Models Using Panel Data,” *Journal of Econometrics*, 18, 47-82.
- Arellano, M., and S. Bond (1991). “Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Employment Equations,” *Review of Economic Studies*, 58, 277-297.
- Bai, J. and S. Ng (2004). “A Panic on Unit Root Tests and Cointegration,” *Econometrica*, 72, 1127-1177.
- _____(2010). “Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation,” *Econometric Theory*, 26, 1088-1114.
- Bai, Z.D. and J.W. Silverstein (2004). “CLT for Linear Statistics of Large-Dimensional Sample Covariance Matrices,” *Annals of Probability*, 32, 553-605.
- Baltagi, B.H., and C. Kao (2000). “Nonstationary Panels, Cointegration in Panels and Dynamic panels, A Survey in Nonstationary panels, Panel Cointegration and Dynamic Panels,” *Advances in Econometrics*, vol. 15, ed. by B. Baltagi, Amsterdam: JAI Press, 7-52.
- Banerjee, A. (1999). “Panel Data Unit Roots and Cointegration: An Overview,” *Oxford Bulletin of Economics and Statistics*, 61, 607-629.

- Bernard, A., and C. Jones (1996). "Productivity Across Industries and Countries: Time Series Theory and Evidence," *Review of Economics and Statistics*, 78, 135-146.
- Binder, M., C. Hsiao and M.H. Pesaran (2005). "Estimation and Inference in Short Panel Vector Autoregression with Unit Roots and Cointegration," *Econometric theory*, 21, 795-837.
- Bodkin, R.G., L.R. Klein and K. Marwah (1991). "A History of Macroeconometric Model-Building," Hants, England: Edward Elgar.
- Box, GEP and G.M. Jenkins (1970). *Time Series Analysis, Forecasting and Control*. Holden Day: San Francisco.
- Breitung, J., and S. Das (2008). "Panel Unit Roots Under Cross Sectional Dependence," *Statistica Neerlandica*, 59, 414-433.
- Breitung, J., and M.H. Pesaran (2008). "Unit Roots and Cointegration in Panels," *The Econometrics of Panel Data*, edited by L. Matyas and P. Sevestre, Berlin: Springer-Verlag, 279-322.
- Chang, Y. (2002). "Nonlinear IV Unit Roots Tests in Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 110, 261-292.
- Choi, In (2001). "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20, 249-272.
- _____ (2002). "Combination Unit Root Tests for Cross-Sectionally Correlated Panels," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research, Essays in Honor of P.C.B. Phillips*, Cambridge: Cambridge University Press.
- _____ (2006). "Nonstationary panels," in *Palgrave Handbooks of Econometrics*, vol I, ed. by K. Patterson and T.C. Mills, New York: Palgrave Macmillan, 511-539.
- Choi, I., and T.K. Chue (2007). "Subsampling Hypothesis Tests for Nonstationary Panels with Applications to Exchange Rates and Stock Prices," *Journal of applied Econometrics*, 22, 223-264.
- Dickey, D.A., and W.A. Fuller (1979). "Distribution of the Estimators for Autoregressive

Time Series with a Unit Root". *Journal of the American Statistical Association*, 74: 427-431.

_____(1981). "Likelihood ratio statistics for autoregressive Time series with a Unit Root". *Econometrica*, 49: 1057-1072.

Engle, R.F., and C.W.J. Granger (1987). "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

Frankel, J.A., and A.K. Rose (1996). "A Panel Project on Purchasing Power Parity: Mean Revision Between and Within Countries," *Journal of International Economics*, 40, 209-244.

Griliches, Z. and Y. Grunfeld (1960). "Is Aggregation Really Bad?", *Review of Economics and Statistics*, 42, 1-13.

Groen, J.J.J., and F. Kleiergen (2003). "Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models," *Journal of Business and Economic Statistics*, 21, 295-318.

Hadri, K. (2000). "Testing for Stationarity in Heterogeneous Panel Data," *Econometrics Journal*, 3, 148-161.

Hadri, K., and R. Larsson (2005). "Testing for Stationarity in Heterogeneous panel Data Where the Time Dimension is Fixed," *Econometric Journal*, 8, 55-69.

Han, C., and P.C.B. Phillips (2013). "First Difference Maximum Likelihood and Dynamic Panel Estimation," *Journal of Econometrics*, 175, 35-45.

Hsiao, C. (1979a). "Causality Tests in Econometrics," *Journal of Economic Dynamics and Control*, 1, 321-46.

_____(1979b). "Autoregressive Modelling of Canadian Money and Income Data," *Journal of the American Statistical Association*, 74, 553-560.

_____(1982). "Autoregressive Modelling and Causal ordering of Economic Variables," *Journal of Economic Dynamics and Control*, 4, 243-259.

_____(1983). "Identification, in *Handbook of Econometrics*, vol. I, edited by Z.

Griliches and M. Intriligator, pp. 223-83. Amsterdam: North-Holland.

_____(2007). "Panel Data Analysis — Advantages and Challenges." *Test*, 16, 1-22.

_____(2014). *Analysis of Panel Data*, 3rd edition, Cambridge: Cambridge University Press (forthcoming).

Hsiao, C. and M.H. Pesaran (2008). "Random Coefficients Models", *The Econometrics of Panel Data*, 3rd edition, edited by L. Matayas and P. Sevestre, Berlin: Springer, 187-216.

Hsiao, C. and S.K. Wan (2014). "Is There an Optimal Forecast Combination?," *Journal of Econometrics*, 178, 294-309.

Hsiao, C., and J. Zhang (2013). "MLE vs GMM for Dynamic Panel Data Models When Both N and T are Large," *mimeo*.

Hsiao, C. and Q.K. Zhou (2014). "Statistical Inference for Panel Dynamic simultaneous Equations Models," *Journal of Econometrics* (forthcoming).

Hsiao, C., H.S. Ching and S. Wan (2012). "A Panel Data Approach for Program Evaluation - Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 27, 705-740.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (2002). "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering short Time Periods," *Journal of Econometrics*, 109, 107-150.

Hsiao, C., Y. Shen and H. Fujiki (2005). "Aggregate vs. Disaggregate Data Analysis - A Paradox in the Estimation of Money Demand Function of Japan Under the Low Interest Rate Policy," *Journal of applied Econometrics*, 20, (2005), 579-601.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Kao, C. (1999). "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90, 1-44.

- Kao, C., and M.H. Chiang (2000). "On the estimation and Inference of a Cointegrated Regression in panel Data," in *Advances in Econometrics*, vol. 15, ed. by B. Baltagi, Amsterdam: JAI Press, 161-168.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). "Testing Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159-178.
- Levin, A., C. Lin, and J. Chu (2002). "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics*, 108, 1-24.
- Lewbel, A., (1994). "Aggregation and Simple Dynamics". *American Economic Review*, 84: 905-918.
- Maddala, G.S. and S. Wu (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, " *Oxford Bulletin of Economics and Statistics*, 61, 631-652.
- Magnus, J.R., and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised ed., New York: John Wiley and Sons.
- McCoskey, S., and C. Kao (1998). "A Residual-Based Test of the Null of Cointegration in Panel Data," *Econometric Reviews*, 17, 57-84.
- Moon, H.R. and B. Perron (2004), "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics* 122, 81-126.
- Nerlove, M. (2000). "An Essay on the History of Panel Data Econometrics," paper presented at 2000 Panel Data Conference in Geneva.
- Park, J. (1992). "Canonical Cointegrating Regressions," *Econometrica*, 60, 119-143.
- Pedroni, P. (1994). "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," *Econometric theory*, 20, 597-625.
- Pesaran, M.H. (2004). "General Diagnostic Tests for Cross-Section Dependence in Panels," *mimeo*.

- Pesaran, M.H. (2007). "A simple Panel Unit Root Test in the Presence of Cross-Section Dependence," *Journal of Applied Econometrics*, 22, 265-312.
- Pesaran, M.H., T. Schuermann and S.M. Weiner (2004). "Modelling Regional Interdependencies Using a Global Error-Correlation Macroeconometrics Models," *Journal of Business and Economic Statistics*, 22, 129-162.
- Pesaran, J.H., Y. Shin, and R.J. Smith (2000). "Structural Analysis of Vector Error Correction Models with Exogenous $I(1)$ Variables," *Journal of Econometrics*, 2000, 97, 293-343.
- Pesaran, M.H., L.V. Smith and T. Yamagata (2012). "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *mimeo*.
- Phillips, P.C.B. (1991). "Optimal Inference in Cointegrating Systems," *Econometrica*, 59, 283-306.
- _____(1986). "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311-340.
- Phillips, P.C.B., and S.N. Durlauf (1986). "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-496.
- Phillips, P.C.B., and B.E. Hansen (1990). "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies*, 57, 99-125.
- Phillips, P.C.B., and H.R. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057-1111.
- _____(2000a). "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Reviews*, 19, 263-286.
- _____(2000b). "Estimation of Autoregressive Roots Near Unity Using Panel Data," *Econometric theory*, 16, 927-997.
- Phillips, P.C.B., and M. Loretan (1991). "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 58, 407-436.
- Phillips, P.C.B. and D. Sul (2003). "Dynamic Panel Estimation and Homogeneity Testing

- Under Cross-Section Dependence,” *Econometrics Journal*.
- Saikkonen, P. (1991). “Asymptotically Efficient Estimation of Cointegration Regressions,” *Econometric Theory*, 7, 1-21.
- Sargan, J.D., and A. Bhargava (1983). “Testing for Residuals from Least Squares Regression being Generated by Gaussian Random Walk,” *Econometrica*, 51, 153-174.
- Sims, C. (1980). “Macroeconomic and Reality,” *Econometrica*, 48, 1-48.
- Stock, J.H. and M.W. Watson (2002). “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20, 147-162.
- _____(2005). “Implications of Dynamic Factor Models for VAR Analysis,” *mimeo*.
- Stoker, T.M. (1993). Empirical Approaches to the Problem of Aggregation over Individuals”. *Journal of Economic Literature*, 31: 1827-1874.
- Theil, H., (1954). *Linear Aggregation of Economic Relations*. North-Holland: Amsterdam.
- Westerlund, J. (2005). “New Simple tests for Panel Cointegration,” *Econometric Reviews*, 24, 297-31.
- Westerlund, J. and R. Larsson (2012). “Testing for a Unit Root in Random Coefficient Panel Data Model,” *Journal of Econometrics*, 167, 254-273.
- Zaffaroni, P., (2001). Contemporaneous Aggregation of Linear Dynamic Models in Large Economies”. Mimeo, Banca d’Italia.
- Zellner, A. (1962). “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,” *Journal of the American Statistical Association*, 57, 348-68.