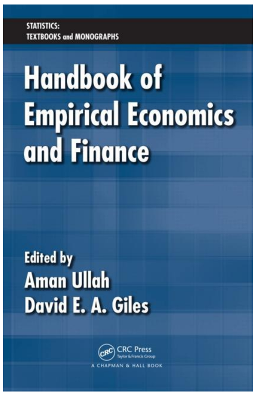

Handbook of Empirical Economics and Finance

Ullah Aman, E. A. Giles David

Dynamic Panel Data Models

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b10440-14
Hsiao Cheng
Published online on: 20 Dec 2010

13

Dynamic Panel Data Models

Cheng Hsiao

CONTENTS

13.1 Introduction

Panel data, by blending inter-individual differences and intra-individual dynamics, have greater capacity for capturing the complexity of human behavior than data sets with only a temporal or a cross-sectional dimension (e.g., Hsiao 2003, 2007). However, typical panels focus on individual outcomes. Factors affecting individual outcomes are numerous. It is rare that the conditional density of the outcomes, y_{it}, conditional on certain variables, $\underset{\sim}{x}_{it}$, is independently, identically distributed across individual i and over time, t. To capture the effects of those omitted factors, empirical researchers often assume that, in addition to the effects of observed $\underset{\sim}{x}_{it}$, there exist unobserved

P1: BINAYA KUMAR DASH
November 3, 2010 16:25 C7035 C7035˙C013

individual-specific effects α_i and time-specific effects λ_t. These unobserved individual-specific and/or time-specific effects, α_i and λ_t, are supposed to capture the impacts of those omitted variables that vary across individuals but stay constant over time and the impact of those variables that vary over time but are the same for all individuals at a given time. They can be either treated as fixed constants or random variables, respectively called fixed effects (FE) or random effects (RE) model. The advantage of the FE modeling is that there is no need to postulate the relationship between the unobserved effects and the conditioning variables. The disadvantage is that it introduces the classical "incidental parameter" problems if either the time series dimension T or cross-sectional dimension, N, is finite (e.g., Neyman and Scott 1948). The advantage of the random effects modeling is that the number of unknown parameters stay constant as N and/or T increases. The disadvantage is that the relationships between the effects and the observed conditional variables have to be postulated, say, the conditional distribution of the effects given the observed factors (e.g., Hsiao 2007).

The unobserved heterogeneity across individuals and over time that are not captured by the included conditional variables could either be modeled additively or multiplicatively. Furthermore, many people believe that "all interesting economic behavior is inherently dynamic, dynamic models are the only relevant models" (e.g., Nerlove 2002). However, the estimation of dynamic models with specific effects is a great deal more difficult than the estimation of nondynamic models because the estimation of structural parameters (those parameters that are the same across i and over t) is not independent of the estimation of incidental parameters. For dynamic models there is also an issue of how to model "initial observations."

We set up the basic models in Section 13.2. Since for models involving incidental parameters the conditions for law of large numbers and central limit theorems to hold are violated, estimators based on the likelihood principle or methods of moments are no longer consistent. Section 13.3 shows the inconsistency of the maximum likelihood estimator (MLE) or covariance estimator (CV) of structural parameters in the presence of incidental parameters. Section 13.4 discusses the issues of initial values.

A general principle to obtain consistent estimators for structural parameters for models involving incidental parameters is to transform the original models into models that no longer involve incidental parameters; in Sections 13.5 and 13.6 we illustrate the implementation of this principle for the likelihood and method of moments approach by considering a simple dynamic panel data model with additive individual-specific effects. Section 13.7 discusses the estimation of dynamic models with both individual- and time-specific additive effects. Section 13.8 discusses the estimation with multiplicative individual- and time-specific effects. Section 13.9 proposes a test of additive versus multiplicative effects. Concluding remarks are in Section 13.10.

13.2 The Basic Models

We consider a dynamic model of the form[1]

$$y_{it} = \rho y_{i,t-1} + \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + v_{it}, \quad |\rho| < 1, i = 1, \ldots, N, t = 1, \ldots, T, \tag{13.1}$$

and the initial values y_{io} are observable. For ease of exposition, we assume $\underset{\sim}{x}_{it}$ is a $K \times 1$ vector of strictly exogenous variables and the error term either takes the form

$$v_{it} = \alpha_i + \lambda_t + \epsilon_{it} \tag{13.2}$$

or

$$v_{it} = \alpha_i \lambda_t + \epsilon_{it}, \tag{13.3}$$

where ϵ_{it} is independently, identically distributed with mean 0 and variance σ_ϵ^2, and the individual- and time-specific effects α_i and λ_t can be either fixed or random. When α_i and λ_t are fixed constants, we impose the normalization condition $\sum_{i=1}^N \alpha_i = 0$, $\sum_{t=1}^T \lambda_t = 0$ and assume $\lim \frac{1}{N} \sum_{i=1}^N \alpha_i^2$ and $\lim \frac{1}{T} \sum_{t=1}^T \lambda_t^2$ are finite positive constants. When α_i and λ_t are random, we assume that

$$\begin{aligned}
E\alpha_i &= E\lambda_t = E\epsilon_{it} = 0, \\
E\alpha_i \underset{\sim}{x}_{it} &= E\lambda_t \underset{\sim}{x}_{is} = E\underset{\sim}{x}_{it}\epsilon_{it} = \underset{\sim}{0}, \\
E\alpha_i\lambda_t &= E\lambda_t\epsilon_{is} = E\alpha_i\epsilon_{it} = 0, \\
E\alpha_i\alpha_j &= \begin{cases} \sigma_\alpha^2, & \text{if} \quad i = j, \\ 0, & \text{otherwise.} \end{cases} \\
E\lambda_t\lambda_s &= \begin{cases} \sigma_\lambda^2, & \text{if} \quad t = s, \\ 0, & \text{otherwise.} \end{cases} \\
E\epsilon_{it}\epsilon_{js} &= \begin{cases} \sigma_\epsilon^2, & \text{if} \quad i = j \quad \text{and} \quad t = s, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{13.4}$$

The presence of unknown α_i introduces serial correlation that does not die out as T increases. The presence of λ_t introduces correlation across individuals that does not die out as N increases.

[1] When T is finite, there is no need to restrict $|\rho| < 1$ to obtain the asymptotic normality results. However, for ease of exposition we shall assume $|\rho| < 1$.

13.3 The Maximum Likelihood Estimator (MLE) (or Covariance Estimators (CV)) in the Presence of Incidental Parameters

Under the assumption that ϵ_{it} is independent normal and fixed y_{i0} the MLE of the FE model Equation 13.1 and Equation 13.2 is equal to

$$\begin{pmatrix} \hat{\rho} \\ \hat{\underset{\sim}{\beta}} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} \begin{pmatrix} y_{i,t-1}^{*2}, & y_{i,t-1}^{*} \underset{\sim}{x}_{it}^{*'} \\ \underset{\sim}{x}_{it}^{*} y_{i,t-1}^{*}, & \underset{\sim}{x}_{it}^{*} \underset{\sim}{x}_{it}^{*'} \end{pmatrix} \right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} \begin{pmatrix} y_{i,t-1}^{*} \\ \underset{\sim}{x}_{it}^{*} \end{pmatrix} y_{it}^{*} \right], \tag{13.5}$$

$$\hat{\alpha}_i = \bar{y}_i - \hat{\rho} \bar{y}_{i,-1} - \hat{\underset{\sim}{\beta}}' \bar{\underset{\sim}{x}}_i, \quad i = 1, \dots, N, \tag{13.6}$$

$$\hat{\lambda}_t = \bar{y}_t - \hat{\rho} \bar{y}_{t-1} - \hat{\underset{\sim}{\beta}}' \bar{\underset{\sim}{x}}_t, \quad t = 1, \dots, T, \tag{13.7}$$

where $y_{it}^* = (y_{it} - \bar{y}_i - \bar{y}_t + \bar{y})$, $\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}$, $\bar{y}_t = \frac{1}{N} \sum_{i=1}^{N} y_{it}$, $\bar{y} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$, and similarly for $\bar{y}_{t-1}$, $\bar{y}_{i,-1}$, $\bar{\underset{\sim}{x}}_i$, $\bar{\underset{\sim}{x}}_t$, $\underset{\sim}{x}_{it}^*$, v_{it}^*, $\bar{v}_i$, $\bar{v}_t$, and $\bar{v}$. The FE MLE of $(\rho, \underset{\sim}{\beta})$ is also called the covariance estimator because it is equivalent to first applying covariance transformation to sweep out α_i and λ_t,

$$y_{it}^* = \rho y_{i,t-1}^* + \underset{\sim}{\beta}' \underset{\sim}{x}_{it}^* + v_{it}^*, \tag{13.8}$$

then apply the least squares estimator to Equation 13.8. When T is finite, there are only finite number of y_{it} that contain information about α_i and α_i increases with N, the MLE is inconsistent no matter how large N is because α_i becomes incidental parameter. To illustrate this, there is no loss of generality to just consider the simple case of $\underset{\sim}{\beta} = 0$, so Equation 13.1 becomes

$$y_{it} = \rho y_{i,t-1} + v_{it}. \tag{13.9}$$

The MLE of ρ under the assumption that y_{i0} are fixed is equal to

$$\hat{\rho}_{cv} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-1}^* y_{it}^*}{\sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-1}^{*2}} \tag{13.10}$$

The probability limit of $\hat{\rho}_{cv}$ is equal to (Hahn and Moon 2006; Hsiao and Tahmiscioglu 2008)

$$\text{plim}_{N\to\infty}(\hat{\rho}_{cv} - \rho) = -\frac{1+\rho}{T-1} \left(1 - \frac{1}{T} \frac{1-\rho^T}{1-\rho}\right) \left\{1 - \frac{2\rho}{(1-\rho)(T-1)} \left[1 - \frac{1-\rho^T}{T(1-\rho)}\right]\right\}^{-1}. \tag{13.11}$$

This estimator is biased to the order of $(1/T)$ and the bias is identical independent of whether α_i and λ_t are fixed or random and is identical to the case when λ_t are 0 for all t. (e.g., Anderson and Hsiao 1981, 1982; Hahn and Kuersteiner 2002; Hahn and Moon 2006; Hsiao and Tahmiscioglu 2008). When $T \longrightarrow \infty$, the MLE of the FE model is consistent. However, if both N and T go to infinity and $\lim\left(\frac{N}{T}\right) = c > 0$, Hahn and Moon (2006) have shown that $\sqrt{NT}(\hat{\rho}_{cv} - \rho)$ is asymptotically normally distributed with mean $-\sqrt{c}(1+\rho)$ and variance $1-\rho^2$. In other words, the usual t-statistic based on $\hat{\rho}_{cv}$ could be subject to severe size distortion.

13.4 Issues of Initial Observations

One way to get around incidental parameters problem is to assume α_i and λ_t random and satisfying Equation 13.4, then the system

$$\underset{\sim}{y}_i = \underset{\sim}{y}_{i,-1}\rho + X_i\underset{\sim}{\beta} + \underset{\sim}{v}_i, \quad i = 1, \ldots, N, \tag{13.12}$$

where $\underset{\sim}{y}_i' = (y_{i1}, \ldots, y_{iT})$, $\underset{\sim}{y}_{i,-1}' = (y_{i0}, \ldots, y_{i,T-1})$, $\underset{\sim}{v}_i' = (v_{i1}, \ldots, v_{iT})$ and X_i is the $T \times K$ matrix of $(\underset{\sim}{x}_{it}')$, has

$$\begin{aligned} E\underset{\sim}{v}_i &= \underset{\sim}{0}, \\ E\underset{\sim}{v}_i\underset{\sim}{v}_i' &= \sigma_\epsilon^2 I_T + \sigma_\alpha^2 \underset{\sim}{e}_T\underset{\sim}{e}_T' + \sigma_\lambda^2 I_T, \\ E\underset{\sim}{v}_i\underset{\sim}{v}_j' &= \sigma_\lambda^2 I_T \end{aligned} \tag{13.13}$$

where I_T is T rowed identity matrix and $\underset{\sim}{e}_T$ is a $T \times 1$ vector of 1's. If $(\epsilon_{it}, \alpha_i, \lambda_t)$ are normally distributed, and y_{i0} are fixed constants, the likelihood function is

$$2\pi^{-\frac{NT}{2}}|\Omega|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\underset{\sim}{y} - \underset{\sim}{y}_{-1}\rho - X\underset{\sim}{\beta})'\Omega^{-1}(\underset{\sim}{y} - \underset{\sim}{y}_{-1}\rho - X\underset{\sim}{\beta})\right\} \tag{13.14}$$

where $\underset{\sim}{y} = (\underset{\sim}{y}_1', \ldots, \underset{\sim}{y}_N')'$, $\underset{\sim}{y}_{-1} = (\underset{\sim}{y}_{1,-1}', \ldots, \underset{\sim}{y}_{N,-1}')'$, $X = (X_1', \ldots, X_N')'$,

$$\Omega = \sigma_\epsilon^2 I_{NT} + \sigma_\alpha^2 I_N \otimes \underset{\sim}{e}_T\underset{\sim}{e}_T' + \sigma_\lambda^2 \underset{\sim}{e}_N\underset{\sim}{e}_N' \otimes I_T \tag{13.15}$$

and $\otimes$ denotes the kroecker product. The likelihood function no longer involves incidental parameters and the MLE is consistent and asymptotically normally distributed either N or T or both tend to infity. Given σ_ϵ^2, σ_α^2 and σ_λ^2, the MLE is identical to the generalized least squares estimator (GLS)

$$\begin{pmatrix}\tilde{\rho} \\ \tilde{\underset{\sim}{\beta}}\end{pmatrix} = \left[\begin{pmatrix}\underset{\sim}{y}_{-1}' \\ X'\end{pmatrix}\Omega^{-1}(\underset{\sim}{y}_{-1}, X)\right]^{-1}\left[\begin{pmatrix}\underset{\sim}{y}_{-1}' \\ X'\end{pmatrix}\Omega^{-1}\underset{\sim}{y}\right]. \tag{13.16}$$

However, most panels contain only finite T time series observations. The starting dates of data collection need not correspond to the starting dates of the data generating process. There is no reason to believe that the data generating process of y_{i0} to be different from the data generating process of y_{it}. If y_{i0} and y_{it} are generated from the same process, then $Ey_{i0}v_{it} = Ey_{i0}\alpha_i = 0$ implied by fixed y_{i0} assumption cannot hold.

For ease of notation, we shall assume in this section that $\lambda_t \equiv 0 \quad \forall t$. Continuous substitution of Equation 13.1 yields

$$y_{it} = \underset{\sim}{\beta}' \sum_{j=0}^{t-1} \underset{\sim}{x}_{i,t-j}\rho^j + \rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}\alpha_i + \sum_{j=0}^{t-1} \epsilon_{i,t-j}\rho^j, \tag{13.17}$$

and

$$y_{i0} = \theta_{i0} + v_{i0}, \tag{13.18}$$

where $\theta_{i0} = \underset{\sim}{\beta}' \sum_{j=0}^{m} \underset{\sim}{x}_{i,-j}\rho^j$, $v_{i0} = \frac{1-\rho^m}{1-\rho}\alpha_i + \sum_{j=0}^{m} \epsilon_{i,-j}\rho^j$, assuming the process started at period $-m$. Then

$$E v_{i0} v_{it} = \frac{1-\rho^m}{1-\rho}\sigma_\alpha^2 = c^* \neq 0. \tag{13.19}$$

Therefore, conditional on y_{i0} (or v_{i0}),

$$\underset{\sim}{y}_i = \underset{\sim}{y}_{i,-1}\rho + x_i'\underset{\sim}{\beta} + \underset{\sim}{e}_T(y_{i0} - \theta_{i0})c^* + \underset{\sim}{v}_i^*, \; i = 1, \ldots, N, \tag{13.20}$$

where $\underset{\sim}{v}_i^* = \underset{\sim}{v}_i - \underset{\sim}{e}_T(y_{i0} - \theta_{i0})c^*$. When T is large, the correlation between y_{it} and y_{i0} will approach zero as can be seen from Equation 13.17. When $|\rho| < 1$, asymptotically there is no difference between Equations 13.12 and 13.20, thus between treating y_{i0} fixed or y_{i0} random. However, in finite T, $y_{i,t-1}$ and y_{i0} are correlated from Equation 13.17. Regressing y_{it} on $y_{i,t-1}$ and $\underset{\sim}{x}_{it}$ is subject to omitted variable $((y_{i0} - \theta_{i0}))$ bias no matter how large N is.

To obtain consistent estimators of ρ and $\underset{\sim}{\beta}$, one should either apply GLS to Equation 13.20 (namely, the conditional system Equation 13.20 conditional on y_{i0}), or to complete the system by maximizing the joint likelihood function of $(y_{i0}, y_{i1}, \ldots, y_{iT})$,

$$\begin{aligned} y_{i0} &= \theta_{i0} + v_{i0}, \\ \underset{\sim}{y}_i &= \underset{\sim}{y}_{i,-1}\rho + X\underset{\sim}{\beta} + \underset{\sim}{v}_i, \quad i = 1, \ldots, N. \end{aligned} \tag{13.21}$$

However, θ_{i0} depends on $\underset{\sim}{x}_{i,-j}$ which are unobservable. Treating θ_{i0} as unknown parameters again will subject the system Equation 13.21 to incidental parameters when T is finite and N is large.

To get around the incidental parameters issues, Bhargava and Sargan (1983) show that if $\underset{\sim}{x}_{it}$ is generated by a homogeneous process

$$\underset{\sim}{x}_{it} = \underset{\sim}{a} + \sum_{j=0} B_j \underset{\sim}{\eta}_{i,t-j}, \sum |B_j| < \infty, \tag{13.22}$$

where $\eta_{i,t-j}$ are i.i.d. random variables with mean zero and constant variance Σ_η, then[2]

$$E(\underset{\sim}{\theta}_{i0} \mid \underset{\sim}{x}_i) = \underset{\sim}{\pi}'\underset{\sim}{x}_i, \; i = 1, \ldots, N, \tag{13.23}$$

[2] For ease of notation, we have merged the intercept term into $\underset{\sim}{x}_i$.

where $\underset{\sim}{x}_i = (\underset{\sim}{x}'_{i1}, \ldots, \underset{\sim}{x}'_{iT})$. Substituting Equation 13.23 into Equation 13.21 yields

$$
\begin{aligned}
y_{i0} &= \underset{\sim}{\pi}'\underset{\sim}{x}_i + v^*_{i0},\\
\underset{\sim}{y}_i &= \underset{\sim}{y}_{i,-1}\rho + X_i\underset{\sim}{\beta} + \underset{\sim}{v}_i, \quad i = 1, \ldots, N.
\end{aligned} \tag{13.24}
$$

System Equation 13.24 no longer involves incidental parameters. Therefore, the MLE or GLS of Equation 13.24,[3]

$$
\hat{\underset{\sim}{\delta}} = \left(\sum_{i=1}^{N} \tilde{X}'_i \tilde{V}^{-1}\tilde{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \tilde{X}'_i \tilde{V}^{-1}\tilde{\underset{\sim}{y}}_i\right), \tag{13.25}
$$

is consistent and asymptotically normally distributed either N or T or both tend to infinity with covariance matrix

$$
\text{Cov}\,(\hat{\underset{\sim}{\delta}}_{\text{GLS}}) = \left(\sum_{i=1}^{N} \tilde{X}'_i \tilde{V}^{-1}\tilde{X}_i\right)^{-1} \tag{13.26}
$$

where $\hat{\underset{\sim}{\delta}} = (\underset{\sim}{\pi}', \rho, \underset{\sim}{\beta}')$, $\tilde{\underset{\sim}{y}}'_i = (y_{i0}, \underset{\sim}{y}'_i)$, and

$$
\tilde{X}_i = \begin{pmatrix}
\underset{\sim}{x}'_i & 0 & \underset{\sim}{0}' \\
\underset{\sim}{0} & y_{i0} & \underset{\sim}{x}'_{i1} \\
\underset{\sim}{0} & y_{i1} & \underset{\sim}{x}'_{i2} \\
\cdots & . & \cdots \\
\cdots & y_{i,T-1} & \underset{\sim}{x}'_{iT}
\end{pmatrix}. \tag{13.27}
$$

13.5 Method of Moments Estimator for Dynamic Models with Individual-Specific Effects Only

We illustrate the basic idea of generalized methods of moments and the likelihood principle for dynamic model with individual-specific effects only (i.e., $\lambda_t \equiv 0, \forall t$) in this section and the next, then discuss the estimator of models involving both additive α_i and λ_t in Section 13.7.

Taking the first difference of Equation 13.1 under the assumption of $\lambda_t = 0$ yields

$$
\begin{aligned}
\Delta y_{it} &= \rho\Delta y_{i,t-1} + \underset{\sim}{\beta}'\Delta\underset{\sim}{x}_{it} + \Delta\epsilon_{it},\\
i &= 1, \ldots, N,\\
t &= 2, \ldots, T,
\end{aligned} \tag{13.28}
$$

[3] Alternatively, one may apply the conditional MLE or GLS to Equation 13.20 (e.g., Blundell and Bond 1998).

where $\Delta = (1-L)$, L denotes the lag operator so $\Delta y_{it} = y_{it} - y_{i,t-1}$. Equation 13.28 no longer involves α_i. However, $E(\Delta y_{i,t-1}\Delta\epsilon_{it}) \neq 0$. Regressing Δy_{it} on $(\Delta y_{i,t-1}, \Delta \underset{\sim}{x}'_{it})$ yields inconsistent estimators for ρ and $\underset{\sim}{\beta}$. On the other hand, $E(y_{i,t-2}\Delta\epsilon_{it}) = 0$. Therefore, $y_{i,t-2}$ can be used as instrument for $\Delta y_{i,t-1}$. However, $y_{i,t-2}$ is not the only instrument for $\Delta y_{i,t-1}$. As noted by Amemiya and MaCurdy (1986), Ahn and Schmidt (1995), Arellano and Bond (1991), Arellano and Bover (1995), Breusch, Mizon, and Schmidt (1989), etc. that all $y_{i,t-2-j}$, $j = 0, 1, \ldots, t-2$, and all $\underset{\sim}{x}_{it}$ satisfy the condition $E(y_{i,t-2-j}\Delta\epsilon_{it}) = 0$, $E(\underset{\sim}{x}_i \Delta\underset{\sim}{\epsilon}_{it}) = \underset{\sim}{0}$. Let $\underset{\sim}{q}_{it} = (y_{i0}, y_{i1}, \ldots, y_{i,t-2}, \underset{\sim}{x}'_i)'$, we have the moment conditions

$$E(\underset{\sim}{q}_{it}\Delta\underset{\sim}{\epsilon}_{it}) = 0, \quad t = 2, \ldots, T. \tag{13.29}$$

Stacking the $(T-1)$ first difference equation of Equation 13.28 in matrix form, we have

$$\Delta\underset{\sim}{y}_i = \Delta\underset{\sim}{y}_{i,-1}\rho + \Delta X_i\underset{\sim}{\beta} + \Delta\underset{\sim}{\epsilon}_i, \quad i = 1, \ldots, N, \tag{13.30}$$

where $\Delta\underset{\sim}{y}_i$, $\Delta\underset{\sim}{y}_{i,-1}$ and $\Delta\underset{\sim}{\epsilon}_i$ are $(T-1)\times 1$ vectors of $(\Delta y_{i2}, \ldots, \Delta y_{iT})'$, $(\Delta y_{i1}, \ldots, \Delta y_{i,T-1})'$ and $(\Delta\epsilon_{i2}, \ldots, \Delta\epsilon_{iT})'$, respectively, and ΔX_i is the $(T-1)\times K$ stacked matrix $(\Delta\underset{\sim}{x}_{i2}, \ldots, \Delta\underset{\sim}{x}_{iT})'$. Let

$$W_i = \begin{bmatrix} \underset{\sim}{q}_{i2} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{q}_{i3} & \cdots & \underset{\sim}{0} \\ & \vdots & & . \\ & . & \cdots & . \\ \underset{\sim}{0} & \cdot & \underset{\sim}{0}' & \underset{\sim}{q}_{iT} \end{bmatrix} \tag{13.31}$$

be the $(T-1)[(T-1)K + \frac{T}{2}] \times (T-1)$ block diagonal matrix. Then, we have the orthogonality conditions,

$$E W_i \Delta\underset{\sim}{\epsilon}_i = \underset{\sim}{0}. \tag{13.32}$$

Under the assumption that $(\underset{\sim}{y}'_i, \underset{\sim}{x}'_i)$ are independently, identically distributed across i, we may approximate the population moments (Equation 13.32) by the sample moments $\frac{1}{N}\sum_{i=1}^{N} W_i(\Delta\underset{\sim}{y}_i - \Delta\underset{\sim}{y}_{i,-1}\rho - \Delta x_i\underset{\sim}{\beta})$. Since there are in general more moments conditions than the number of unknowns, an efficient moment estimator is to apply the generalized lease squares principle to Equation 13.32 [Generalized methods of moments (GMM)],

$$\underset{\rho,\,\underset{\sim}{\beta}}{\text{Min}} \left(\frac{1}{N}\sum_{i=1}^{N} \Delta\underset{\sim}{\epsilon}'_i W_i\right) \Psi^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} W_i \Delta\underset{\sim}{\epsilon}_i\right), \tag{13.33}$$

where $\Psi = E[\frac{1}{N^2}\sum_{i=1}^{N} W_i \Delta \underset{\sim}{\epsilon}_i \Delta \underset{\sim}{\epsilon}_i' W_i]$. Under the assumption that $\underset{\sim}{\epsilon}_i$ is i.i.d., Ψ may be approximated by $\frac{\sigma_\epsilon^2}{N^2}\sum_{i=1}^{N} W_i A W_i'$, where

$$\underset{(T-1)\times(T-1)}{A} = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdot & 0 \\ -1 & 2 & -1 & \cdots & \cdot & 0 \\ 0 & -1 & 2 & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & -1 \\ 0 & \cdot & \cdot & \cdots & -1 & 2 \end{bmatrix}. \tag{13.34}$$

Thus, the Arellano and Bover (1995) GMM estimator takes the form

$$\hat{\underset{\sim}{\theta}}_{\text{GMM}} = \begin{pmatrix} \hat{\rho} \\ \hat{\underset{\sim}{\beta}} \end{pmatrix}_{\text{GMM}} = \left\{ \left[\sum_{i=1}^{N} \begin{pmatrix} \Delta \underset{\sim}{y}_{i,-1}' \\ \Delta X_i' \end{pmatrix} W_i' \right] \left[\sum_{i=1}^{N} W_i A W_i' \right]^{-1} \left[\sum_{i=1}^{N} W_i (\Delta \underset{\sim}{y}_{i,-1}, \Delta X_i) \right] \right\}^{-1} \cdot \left\{ \left[\sum_{i=1}^{N} \begin{pmatrix} \Delta \underset{\sim}{y}_{i,-1}' \\ \Delta X_i' \end{pmatrix} W_i' \right] \left[\sum_{i=1}^{N} W_i A W_i' \right]^{-1} \left[\sum_{i=1}^{N} W_i \Delta \underset{\sim}{y}_i \right] \right\}, \tag{13.35}$$

with asymptotic covariance matrix

$$\text{Cov}\,(\hat{\underset{\sim}{\theta}}_{\text{GMM}}) = \left\{ \left[\sum_{i=1}^{N} \begin{pmatrix} \Delta \underset{\sim}{y}_{i,-1}' \\ \Delta X_i' \end{pmatrix} W_i' \right] \left[\sum_{i=1}^{N} W_i A W_i' \right]^{-1} \left[\sum_{i=1}^{N} W_i (\Delta \underset{\sim}{y}_i, \Delta X_i) \right] \right\}^{-1}. \tag{13.36}$$

Remark 13.1 The GMM estimator is consistent and asymptotically normally distributed whether α_i is treated as a fixed constant or a random variable because the first difference of Equation 13.1 eliminates α_i from the transformed model (Equation 13.28). However, GMM cannot estimate the coefficients of time-invariant variables, say gender, because first differencing also eliminates such variables in Equation 13.28 but the likelihood approach can if α_i is indeed random and uncorrelated with $\underset{\sim}{x}_i$.

Remark 13.2 When α_i is random and satisfies Equation 13.3, the likelihood approach uses the level equation (Equation 13.1) while the GMM approach uses the first difference equation (Equation 13.28). In general, the variation across individuals is much larger than the variation over time of an individual. Moreover, the likelihood approach uses T equations of (Equation 13.1) but the GMM uses $(T-1)$ equations of (Equation 13.28). Therefore, the likelihood approach is more efficient than the GMM approach (for detail, see Hsiao, Pesaran, and Tahmiscioglu 2002).

Remark 13.3 In implementing the likelihood approach we invoke the normality assumption. However, Equation 13.25 remains consistent and asymptotically normally distributed even v_{it} is not normally distributed. One may view estimators of the type of Equation 13.25 as a quasi-MLE (QMLE).

Remark 13.4 Although we make no assumption about the initial value distribution of y_{i0} in implementing the GMM approach, the assumption that $(\underset{\sim}{y}_i', \underset{\sim}{x}_i')$ are i.i.d. across i actually implies Equation 13.22 which is invoked to get around the incidental parameters problem in the likelihood approach. In other words, the conditions for implementing the likelihood approach are no more restrictive than the GMM approach. Moreover, as shown by Hayakawa (2009), the efficiency of GMM actually depends on the distribution of y_{i0} and σ_α^2.

Remark 13.5 The likelihood approach uses the moment conditions $E\left[\tilde{X}_i' V^{-1}\binom{v_{i0}^*}{\underset{\sim}{v}_i}\right] = \underset{\sim}{0}$, which stay fixed as N and T increases. The number of moment conditions for GMM increases at order T^2 as T increases. In finite sample, the procedure of equating sample moments to population moments can lead to severe bias in GMM as demonstrated in a Monte Carlo by Ziliak (1997). Moreover, if ρ is close to one, the correlations between Δy_{it} and $y_{i,t-j}$ for $j \geq 2$ could be weak and lead to weak instrumental variables problem as demonstrated in the Monte Carlos by Binder, Hsiao, and Pesaran (2005) and Hsiao, Pesaran, and Tahmiscioglu (2002).

Remark 13.6 The moment conditions, Equation 13.31 assumes $\underset{\sim}{x}_{it}$ are strictly exogenous. However, there could be feedback relations between y_{it} and $\underset{\sim}{x}_{i,t+j}$ as in Cheng and Kwan (2000). If $\underset{\sim}{x}_{it}$ is only weakly exogenous, it does not affect the likelihood approach. But for GMM, instead of defining $\underset{\sim}{q}_{it}$ as $(y_{it}, \ldots, y_{i,t-2}, \underset{\sim}{x}_i)$, we have to redefine $\underset{\sim}{q}_{it}$ as $(y_{it}, \ldots, y_{i,t-2}, \underset{\sim}{x}_{it}', \ldots, \underset{\sim}{x}_{i1}')'$, then Equation 13.29 still holds and GMM can be applied with the redefined $\underset{\sim}{q}_{it}$.

13.6 Likelihood Approach for the Dynamic Fixed Individual-Specific Effects Model

When α_i is treated as fixed constants, we can estimate ρ and $\underset{\sim}{\beta}$ by the GMM method discussed in Section 13.5. A similar likelihood approach can also be implemented on the system (Equation 13.28), which no longer contains the fixed α_i. However, just like the RE case, there is the problem of initial values if T is finite. If the data generating process of Δy_{i1} is no different from the data generating process of Δy_{it}, then

$$\Delta y_{i1} = \theta_{i1} + v_{i1}, \tag{13.37}$$

where $\theta_{i1} = \underset{\sim}{\beta}' \sum_{j=0}^{m-1} \Delta \underset{\sim}{x}_{i,1-j} \rho^j$, $v_{i1} = \sum_{j=0}^{m-1} \Delta \epsilon_{i,1-j} \rho^j$, if the process started at period $-m$. Since $\Delta \underset{\sim}{x}_{i,1-j}$ are unknown, so are θ_{i1}. Treating θ_{i1} as an unknown parameter again introduces incidental parameters. To get around this issue, the expected value of θ_{i1} conditional on $\Delta \underset{\sim}{x}_i$ has to be a function of constant parameters,

$$E(\theta_{i1} \mid \Delta \underset{\sim}{x}_i) = \underset{\sim}{\pi}' \Delta \underset{\sim}{x}_i, \quad i = 1, \ldots, N, \tag{13.38}$$

where $\Delta \underset{\sim}{x}_i' = (\Delta \underset{\sim}{x}_{i2}', \ldots, \Delta \underset{\sim}{x}_{iT}')$. Hsiao, Pesaran, and Tahmiscioglu (2002) have shown that if $\underset{\sim}{x}_{it}$ is generated by

$$\underset{\sim}{x}_{it} = \underset{\sim}{\mu}_i + \underset{\sim}{g} + \sum_{j=0}^{\infty} B_j \underset{\sim}{\xi}_{i,t-j}, \quad \sum |B_j| < \infty, \tag{13.39}$$

then Equation 13.38 holds.

Given Equation 13.38, we may write the system of T equations in the form,

$$\begin{aligned} \Delta y_{i1} &= \underset{\sim}{\pi}' \Delta \underset{\sim}{x}_i + v_{i1}^*, \\ \Delta \underset{\sim}{y}_i &= \Delta \underset{\sim}{y}_{i,-1} \rho + \Delta X_i \underset{\sim}{\beta} + \Delta \underset{\sim}{\epsilon}_i, \qquad i = 1, \ldots, N, \end{aligned} \tag{13.40}$$

where $v_{i1}^* = v_{i1} + (\theta_{i1} - E\theta_{i1})$. By construction, $E(v_{i1}^* \mid \Delta \underset{\sim}{x}_i) = 0$, $E v_{i1}^{*2} = \sigma_{v^*}^2$, $E(v_{i1}^* \Delta \epsilon_{i2}) = -\sigma_\epsilon^2$, and $E(v_{i1}^* \Delta \epsilon_{it}) = 0$, for $t = 3, \ldots, T$. Let the $(T \times 1)$ vector $\Delta \underset{\sim}{\epsilon}_i^{*'} = (v_{i1}^*, \Delta \underset{\sim}{\epsilon}_i')$. The covariance matrix of $\Delta \underset{\sim}{\epsilon}_i^*$ is

$$E \Delta \underset{\sim}{\epsilon}_i^* \Delta \underset{\sim}{\epsilon}_i^{*'} = \sigma_\epsilon^2 \begin{bmatrix} h & -1 & 0 & \cdots & \cdots & \cdot \\ -1 & 2 & -1 & \cdots & \cdot & \cdot \\ 0 & -1 & 2 & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdots & -1 & 2 \end{bmatrix} = \Omega^*, \tag{13.41}$$

where $h = \frac{\sigma_{v^*}^2}{\sigma_\epsilon^2}$. Assuming $\Delta \underset{\sim}{\epsilon}_i^*$ is independently normally distributed with covariance matrix Ω^*, then the likelihood function of $\Delta \underset{\sim}{y}_i^* = (\Delta y_{i1}, \Delta \underset{\sim}{y}_i')'$, $i = 1, \ldots, N$, is in the form

$$(2\pi)^{-\frac{NT}{2}} |\Omega^*|^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \Delta \underset{\sim}{\epsilon}_i^{*'} \Omega^{*-1} \Delta \underset{\sim}{\epsilon}_i^* \right\}, \tag{13.42}$$

where $\Delta \underset{\sim}{\epsilon}_i^{*'} = [\Delta y_{i1} - \underset{\sim}{\pi}' \Delta \underset{\sim}{x}_i, \Delta y_{i2} - \Delta y_{i1} \rho - \Delta \underset{\sim}{x}_{i2}' \underset{\sim}{\beta}, \ldots, \Delta y_{iT} - \Delta y_{i,T-1} \rho - \Delta \underset{\sim}{x}_{iT}' \underset{\sim}{\beta}]$. The likelihood function (Equation 13.42) is a function of fixed number of parameters, hence, the MLE is consistent and asymptotically normally distributed either N or T or both tend to infinity. Therefore, standard t or F tests can be applied.[4]

[4] For further discussion of hypothese testing involving dynamic panel data models, see Harris, Matyas, and Sevestre (2008, Section 8.6).

Conditional on h, the MLE of $\underset{\sim}{\delta} = (\underset{\sim}{\pi}', \rho, \underset{\sim}{\beta}')'$ is identical to the GLS,

$$\hat{\underset{\sim}{\delta}}_{\text{GLS}} = \left(\sum_{i=1}^{N} \Delta \tilde{X}_i' \Omega^{*-1} \Delta \tilde{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \Delta \tilde{X}_i' \Omega^{*-1} \Delta \underset{\sim}{y}_i\right), \tag{13.43}$$

where

$$\Delta \tilde{X}_i = \begin{pmatrix} \Delta \underset{\sim}{x}_i' & 0 & \underset{\sim}{0}' \\ \underset{\sim}{0} & \Delta \underset{\sim}{y}_{i-1} & \Delta X_i \end{pmatrix}. \tag{13.44}$$

When h is unknown, one can use a two-step procedure. In the first step, we regress Δy_{i1} on $\Delta \underset{\sim}{x}_i$ to obtain $\hat{\sigma}^2_{v^*}$ and apply GMM to obtain $\hat{\sigma}^2_{\epsilon}$. In the second step, we substitute estimated $\hat{h}$ for h in Equation 13.43. However, the feasible GLS is not as efficient as GLS (for detail, see Hsiao, Pesaran, and Tahmiscoglu 2002).

13.7 Models with Both Individual- and Time-Specific Additive Effects

When time-specific effects also appear in v_{it} as in Equation 13.2, the estimators ignoring the presence of λ_t like those discussed in Sections 13.13 to 13.6 are no longer consistent when T is finite. For notational ease and without loss of generality, we illustrate the fundamental issues of dynamic model with both individual- and time-specific additive effects model by restricting $\underset{\sim}{\beta} = \underset{\sim}{0}$ in Equation 13.1, thus the model becomes

$$y_{it} = \rho y_{i,t-1} + v_{it}, \tag{13.45}$$

$$v_{it} = \alpha_i + \lambda_t + \epsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T,\ y_{i0} \text{ observable.} \tag{13.46}$$

The panel data estimators discussed in Sections 13.5 and 13.6 assume no presence of λ_t (i.e., $\lambda_t = 0 \forall t$). When λ_t are indeed present, those estimators are not consistent if T is finite when $N \to \infty$. For instance, the consistency of GMM (Equation 13.33) is based on the assumption that $\frac{1}{N}\sum_{i=1}^{N} y_{i,t-j}\Delta v_{it}$ converges to the population moments (Equation 13.32). However, if λ_t are also present as in Equation 13.46, this condition is likely to be violated. To see this, taking first difference of Equation 13.45 yields

$$\begin{aligned} \Delta y_{it} &= \rho \Delta y_{i,t-1} + \Delta v_{it} \\ &= \rho \Delta y_{i,t-1} + \Delta \lambda_t + \Delta \epsilon_{it}, \\ i &= 1, \ldots, N, \\ t &= 2, \ldots, T. \end{aligned} \tag{13.47}$$

Although

$$E(y_{i,t-j}\Delta v_{it}) = 0 \text{ for } j = 2, \ldots, t, \tag{13.48}$$

the sample moment, as $N \longrightarrow \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,t-j}\Delta v_{it} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t-j}\Delta\lambda_t + \frac{1}{N}\sum_{i=1}^{N} y_{i,t-j}\Delta\epsilon_{it} \tag{13.49}$$

converges to $\bar{y}_{t-j}\Delta\lambda_t$, which in general is not equal to zero, in particular, if y_{it} has mean different from zero,[5] where $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$.

To obtain consistent estimators of ρ, we need to take explicit account of the presence of λ_t in addition to α_i. If α_i and λ_t are random and satisfy Equation 13.4, because $Ey_{i0}v_{it} \neq 0$, we either have to write Equation 13.45 conditional on y_{i0} or to complete the system (Equation 13.45) by deriving the marginal distribution of y_{i0}. By continuous substitutions, we have

$$\begin{aligned} y_{i0} &= \frac{1-\rho^m}{1-\rho}\alpha_i + \sum_{j=0}^{m-1}\lambda_{-j}\rho^j + \sum_{j=0}^{m-1}\epsilon_{i,-j}\rho^j \\ &= v_{i0}, \end{aligned} \tag{13.50}$$

assuming the process started at period $-m$.

Under Equation 13.4, $Ey_{i0} = Ev_{i0} = 0$, Var $(y_{i0}) = \sigma_0^2$, $E(v_{i0}v_{it}) = \frac{1-\rho^m}{1-\rho}\sigma_\alpha^2 = c$, $Ev_{it}v_{jt} = d$. Stacking the $T+1$ time series observations for the ith individual into a vector, $\underset{\sim}{y}_i = (y_{i0}, \ldots, y_{iT})'$ and $\underset{\sim}{y}_{i,-1} = (0, y_{i0}, \ldots, y_{i,T-1})'$, $\underset{\sim}{v}_i = (v_{i0}, \ldots, v_{iT})'$. Let $\underset{\sim}{y} = (\underset{\sim}{y}'_1, \ldots, \underset{\sim}{y}'_N)'$, $\underset{\sim}{y}_{-1} = (\underset{\sim}{y}'_{1,-1}, \ldots, \underset{\sim}{y}'_{N,-1})$, $\underset{\sim}{v} = (\underset{\sim}{v}'_1, \ldots, \underset{\sim}{v}'_N)'$, then

$$\begin{aligned} \underset{\sim}{y} &= \underset{\sim}{y}_{-1}\rho + \underset{\sim}{v}, \\ E\underset{\sim}{v} &= \underset{\sim}{0}, \\ E\underset{\sim}{v}\underset{\sim}{v}' &= \sigma_\epsilon^2 I_N \otimes \begin{pmatrix} \omega & \underset{\sim}{0}' \\ \underset{\sim}{0} & I_T \end{pmatrix} + \sigma_\alpha^2 I_N \otimes \begin{pmatrix} 0 & c^*\underset{\sim}{e}'_T \\ c^*\underset{\sim}{e}_T & \underset{\sim}{e}_T\underset{\sim}{e}'_T \end{pmatrix} \\ &\quad + \sigma_\lambda^2 \underset{\sim}{e}_N\underset{\sim}{e}_N \otimes \begin{pmatrix} d^* & \underset{\sim}{0}' \\ \underset{\sim}{0} & I_T \end{pmatrix}, \end{aligned} \tag{13.51, 13.52}$$

$$\omega = \frac{\sigma_0^2 - d}{\sigma_\epsilon^2},\ d^* = \frac{d}{\sigma_\lambda^2},\ c^* = \frac{c}{\sigma_\alpha^2}, \tag{13.53}$$

where ⊗ denotes the kronecker product. The system (Equation 13.51) has a fixed number of unknowns $(\rho, \sigma_\epsilon^2, \sigma_\alpha^2, \sigma_\lambda^2, \sigma_0^2, c, d)$ as N and T increase. Therefore, the MLE (or quasi-MLE or GLS) of Equation 13.51 is consistent and asymptotically normally distributed.

---

[5] For instance, if y_{it} is also a function of exogenous variables as Equation 13.1.

When α_i and λ_t are fixed constants, we note that first differencing only eliminates α_i from the specification. The time-specific effects, $\Delta\lambda_t$, remain at Equation 13.47. To further eliminate $\Delta\lambda_t$, we note that the cross-sectional mean $\Delta y_t = \frac{1}{N}\sum_{i=1}^{N}\Delta y_{it}$ is equal to

$$\Delta y_t = \rho\Delta y_{t-1} + \Delta\lambda_t + \Delta\epsilon_t, \tag{13.54}$$

where $\Delta\epsilon_t = \frac{1}{N}\sum_{i=1}^{N}\Delta\epsilon_{it}$. Taking deviation of Equation 13.47 from Equation 13.54 yields

$$\begin{aligned}\Delta y^*_{it} &= \rho\Delta y^*_{i,t-1} + \Delta\epsilon^*_{it},\\ i &= 1,\dots,N,\\ t &= 2,\dots,T,\end{aligned} \tag{13.55}$$

where $\Delta y^*_{it} = (\Delta y_{it} - \Delta y_t)$ and $\Delta\epsilon^*_{it} = (\Delta\epsilon_{it} - \Delta\epsilon_t)$. The system (Equation 13.55) no longer involves α_i and λ_t.

Since

$$E[y_{i,t-j}\Delta\epsilon^*_{it}] = 0 \text{ for } \begin{array}{l} j = 2,\dots,t,\\ t = 2,\dots,T,\end{array} \tag{13.56}$$

the $\frac{1}{2}T(T-1)$ orthogonality conditions can be represented as

$$E(W_i\Delta\underset{\sim}{\tilde{\epsilon}}^*_i) = \underset{\sim}{0}, \tag{13.57}$$

where $\Delta\underset{\sim}{\tilde{\epsilon}}^*_i = (\Delta\epsilon^*_{i2},\dots,\Delta\epsilon^*_{iT})'$,

$$W_i = \begin{pmatrix} \underset{\sim}{q}_{i2} & \underset{\sim}{0} & \cdots & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{q}_{i3} & & \\ \cdot & \cdot & \ddots & \\ \vdots & \vdots & & \\ \underset{\sim}{0} & \underset{\sim}{0} & & \underset{\sim}{q}_{iT}\end{pmatrix}, \quad i = 1,\dots,N,$$

and $\underset{\sim}{q}_{it} = (y_{i0}, y_{i1},\dots,y_{i,t-2})'$, $t = 2, 3,\dots,T$. Following Arellano and Bond (1991), we can propose a generalized method of moments (GMM) estimator,[6]

$$\begin{aligned}\tilde{\rho}_{\text{GMM}} = &\left\{\left[\frac{1}{N}\sum_{i=1}^{N}\Delta\underset{\sim}{\tilde{y}}^{*'}_{i,-1}W_i'\right]\hat{\Psi}^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}W_i\Delta\underset{\sim}{\tilde{y}}^{*}_{i,-1}\right]\right\}^{-1}\\ &\left\{\left[\frac{1}{N}\sum_{i=1}^{N}\Delta\underset{\sim}{\tilde{y}}^{*'}_{i,-1}\underset{\sim}{W}_i'\right]\hat{\Psi}^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}W_i\Delta\underset{\sim}{\tilde{y}}^{*}_{i}\right]\right\},\end{aligned} \tag{13.58}$$

[6] For ease of exposition, we have only considered the GMM that makes use of orthogonality conditions. For additional moments conditions such as homoscedasticity or initial observations see, e.g., Ahn and Schmidt (1995), Blundell and Bond (1998).

where $\Delta \underset{\sim}{\tilde{y}}_i^* = (\Delta y_{i2}^*, \ldots, \Delta y_{iT}^*)'$, $\Delta \underset{\sim}{\tilde{y}}_{i-1}^* = (\Delta y_{i1}^*, \ldots, \Delta y_{i,T-1}^*)'$, and

$$\hat{\Psi} = \frac{1}{N^2}\left[\sum_{i=1}^{N} W_i \hat{\underset{\sim}{\epsilon}}_i^*\right]\left[\sum_{i=1}^{N} W_i \hat{\underset{\sim}{\epsilon}}_i^*\right]' \tag{13.59}$$

and $\Delta \hat{\underset{\sim}{\epsilon}}_i^* = \Delta \underset{\sim}{\tilde{y}}_i^* - \Delta \underset{\sim}{\tilde{y}}_{i,-1}^* \tilde{\rho}$, and $\tilde{\rho}$ denotes some initial consistent estimator of ρ, say a simple instrumental variable estimator.

The asymptotic covariance matrix of $\tilde{\rho}_{\text{GMM}}$ can be approximated by

$$\text{asy. cov}\,(\tilde{\rho}_{\text{GMM}}) = \left\{\left[\sum_{i=1}^{N} \Delta \underset{\sim}{\tilde{y}}_{i,-1}^{*'} W_i\right] \hat{\Psi}^{-1}\left[\sum_{i=1}^{N} W_i \Delta \underset{\sim}{\tilde{y}}_{i,-1}^*\right]\right\}^{-1}. \tag{13.60}$$

To implement the likelihood approach, we need to complete the system (Equation 13.55) by deriving the marginal distribution of Δy_{i1}^* through continuous substitution,

$$\begin{aligned}\Delta y_{i1}^* &= \sum_{j=0}^{m-1} \Delta \epsilon_{i,1-j}^* \rho^j \\ &= \Delta \tilde{\epsilon}_{i1}^*, \qquad i = 1, \ldots, N.\end{aligned} \tag{13.61}$$

Let $\Delta \underset{\sim}{y}_i^* = (\Delta y_{i1}^*, \ldots, \Delta y_{iT}^*)$, $\Delta \underset{\sim}{y}_i^* = (0, \ldots, \Delta y_{i,T-1}^*)$, $\Delta \tilde{\underset{\sim}{\epsilon}}_i^{*'} = (\Delta \tilde{\epsilon}_{i1}^*, \ldots, \Delta \epsilon_{iT}^*)$, the system

$$\Delta \underset{\sim}{y}_i^* = \Delta \underset{\sim}{y}_{i,-1}^* \rho + \Delta \tilde{\underset{\sim}{\epsilon}}_i^*, \tag{13.62}$$

does not involve α_i and λ_t. The MLE conditional on $\omega = \frac{\text{Var}\,(\Delta y_{i1}^*)}{\sigma_\epsilon^2}$ is identical to the GLS

$$\hat{\rho}_{\text{GLS}} = \left[\sum_{i=1}^{N} \Delta \underset{\sim}{y}_{i,-1}^{*'} \tilde{A}^{-1} \Delta \underset{\sim}{y}_{i,-1}^*\right]^{-1}\left[\sum_{i=1}^{N} \Delta \underset{\sim}{y}_{i,-1}^{*'} \tilde{A}^{-1} \Delta \underset{\sim}{y}_i^*\right]. \tag{13.63}$$

where

$$\tilde{A} = \begin{bmatrix} \omega & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & \cdot & \cdot \\ 0 & -1 & 2 & -1 & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & & -1 & 2 \end{bmatrix}. \tag{13.64}$$

The GLS is consistent and asymptotically normally distributed with covariance matrix equal to

$$\text{Var}(\hat{\rho}_{\text{GLS}}) = \sigma_\epsilon^2 \left[\sum_{i=1}^{N} \Delta \underset{\sim}{y}_{i,-1}^{*'} \tilde{A}^{-1} \Delta \underset{\sim}{y}_{i,-1}^*\right]^{-1}. \tag{13.65}$$

Remark 13.7 The GLS with $\Delta\underset{\sim}{\lambda}$ present is basically of the same form as the GLS without the time-specific effects (i.e., $\Delta\underset{\sim}{\lambda} = \underset{\sim}{0}$) (Hsiao, Pesaran, and Tahmiscioglu 2002), (Equation 13.25). However, there is an important difference between the two. The estimator (Equation 13.63) uses $\Delta y^*_{i,t-1}$ as the regressor for the equation Δy^*_{it} (Equation 13.62), not uses $\Delta y_{i,t-1}$ as the regressor for the equation Δy_{it} (Equation 13.47). If there are indeed common shocks that affect all the cross-sectional units, then the estimator Equation 13.25 is inconsistent while Equation 13.63 is consistent (for detail, see Hsiao and Tahmiscioglu 2008). Note also that even though when there are no time-specific effects, Equation 13.63 remains consistent, although it will not be as efficient as Equation 13.25.

Remark 13.8 The estimator (Equation 13.63) and the estimator Equation 13.58 remain consistent and asymptotically normally distributed when the effects are random because the transformation (Equation 13.54) effectively removes the individual- and time-specific effects from the specification. However, if the effects are indeed random,then the MLE or GLS of Equation 13.51 is more efficient.

Remark 13.9 The GLS (Equation 13.63) assumes known ω. If ω is unknown, one may substitute it by a consistent estimator $\hat{\omega}$, then apply the feasible GLS. However, there is an important difference between the GLS and the feasible GLS in a dynamic setting. The feasible GLS is not asymptotically equivalent to the GLS when T is finite. However, if both N and $T \rightarrow \infty$ and $\lim\left(\frac{N}{T}\right) = c > 0$, then the FGLS will be asymptotically equivalent to the GLS. (Hsiao and Tahmiscioglu 2008).

Remark 13.10 The MLE or GLS of Equation 13.63 can also be derived by treating $\Delta\lambda_t$ as fixed parameters in the system (Equation 13.47). Through continuous substitution, we have

$$\Delta y_{i1} = \lambda^*_1 + \Delta\tilde{\epsilon}_{i1}, \tag{13.66}$$

where $\lambda^*_1 = \sum_{j=0}^{m} \rho^j \Delta\lambda_{1-j}$ and $\Delta\tilde{\epsilon}_{i1} = \sum_{j=0}^{m} \rho^j \Delta\epsilon_{i,1-j}$. Let $\Delta\underset{\sim}{y}'_i = (\Delta y_{i1}, \ldots, \Delta y_{iT})$, $\Delta\underset{\sim}{y}'_{i,-1} = (0, \Delta y_{i1}, \ldots, \Delta y_{i,T-1})$, $\Delta\underset{\sim}{\epsilon}'_i = (\Delta\tilde{\epsilon}_{i1}, \ldots, \Delta\epsilon_{iT})$, and $\Delta\underset{\sim}{\lambda}' = (\lambda^*_1, \Delta\lambda_2, \ldots, \Delta\lambda_T)$, we may write

$$\underset{NT\times 1}{\Delta\underset{\sim}{y}} = \begin{pmatrix} \Delta\underset{\sim}{y}_1 \\ \vdots \\ \Delta\underset{\sim}{y}_N \end{pmatrix} = \begin{pmatrix} \Delta\underset{\sim}{y}_{1,-1} \\ \vdots \\ \Delta\underset{\sim}{y}_{N,-1} \end{pmatrix} \rho + (\underset{\sim}{e}_N \otimes I_T)\Delta\underset{\sim}{\lambda} + \begin{pmatrix} \Delta\underset{\sim}{\epsilon}_1 \\ \vdots \\ \Delta\underset{\sim}{\epsilon}_N \end{pmatrix}$$
$$= \Delta\underset{\sim}{y}_{-1}\rho + (\underset{\sim}{e}_N \otimes I_T)\Delta\underset{\sim}{\lambda} + \Delta\underset{\sim}{\epsilon}, \tag{13.67}$$

If ϵ_{it} is i.i.d. normal with mean 0 and variance σ^2_ϵ, then $\Delta\underset{\sim}{\epsilon}'_i$ is independently normally distributed across i with mean $\underset{\sim}{0}$ and covariance matrix $\sigma^2_\epsilon \tilde{A}$, and $\omega = \frac{\text{Var}\,(\Delta\tilde{\epsilon}_{i1})}{\sigma^2_\epsilon}$.

The log-likelihood function of $\Delta \underset{\sim}{y}$ takes the form

$$\log L = -\frac{NT}{2}\log\sigma_\epsilon^2 - \frac{N}{2}\log|\tilde{A}| - \frac{1}{2\sigma_\epsilon^2}[\Delta\underset{\sim}{y} - \Delta\underset{\sim}{y}_{-1}\rho - (e_N \otimes I_T)\Delta\underset{\sim}{\lambda}]' \\ (I_N \otimes \tilde{A}^{-1})[\Delta\underset{\sim}{y} - \Delta\underset{\sim}{y}_{-1}\rho - (\underset{\sim}{e}_N \otimes I_T)\Delta\underset{\sim}{\lambda}]. \quad (13.68)$$

Taking partial derivative of Equation 13.68 with respect to $\Delta\underset{\sim}{\lambda}$ and solving for $\Delta\underset{\sim}{\lambda}$ yields

$$\Delta\hat{\underset{\sim}{\lambda}} = (N^{-1}\underset{\sim}{e}_N' \otimes I_T)(\Delta\underset{\sim}{y} - \Delta\underset{\sim}{y}_{-1}\rho). \quad (13.69)$$

Substituting Equation 13.69 into Equation 13.68 yields the concentrated log-likelihood function.

$$\log L_c = -\frac{NT}{2}\log\sigma_\epsilon^2 - \frac{N}{2}\log|\tilde{A}| \\ - \frac{1}{2\sigma_\epsilon^2}(\Delta\underset{\sim}{y}^* - \Delta\underset{\sim}{y}_{-1}^*\rho)'(I_N \otimes \tilde{A}^{-1})(\Delta\underset{\sim}{y}^* - \Delta\underset{\sim}{y}_{-1}^*\rho). \quad (13.70)$$

Maximizing Equation 13.69 conditional on ω yields Equation 13.63.

Remark 13.11 When ρ approaches to 1 and σ_α^2 is large relative to σ_ϵ^2, the GMM estimator of the form (Equation 13.68) suffers from the weak instrumental variables issues and performs poorly (e.g., Binder, Hsiao, and Pesaran 2005). On the other hand, the performance of the likelihood or GLS estimator (Equation 13.63) is not affected by these problems.

Remark 13.12 Hahn and Moon (2006) propose a bias corrected estimator as

$$\tilde{\rho}_b = \tilde{\rho}_{cv}^* + \frac{1}{T}(1 + \tilde{\rho}_{cv}^*). \quad (13.71)$$

They show that when $N/T \to c$, as both N and T tend to infinity where $0 < c < \infty$,

$$\sqrt{NT}(\tilde{\rho}_b - \rho) \Longrightarrow N(0, 1 - \rho^2). \quad (13.72)$$

The limited Monte Carlo studies conducted by Hsiao and Tahmiscioglu (2008) to investigate the finite sample properties of the feasible GLS (FGLS), GMM, bias corrected (BC) estimator of Hahn and Moon (2006) have shown that in terms of bias and root mean square errors, FGLS dominates. However, the BC rapidly improves as T increase. In terms of the closeness of actual size to the nominal size, again FGLS dominates and rapidly approaches the nominal size when N or T increases. The GMM also has actual sizes close to nominal sizes except for the cases when ρ is close to unity (here $\rho = 0.8$). The BC has significant size distortion, presumably because of the correction of bias being based on $\hat{\rho}_{cv}^*$ and the use of asymptotic covariance matrix which is significantly downward biased in finite sample.

Remark 13.13 Hsiao and Tahmiscioglu (2008) also compared the FGLS and GMM with and without the correction of time-specific effects in the presence of both individual- and time-specific effects or in the presence of individual-specific effects only. It is interesting to note that when both individual- and time-specific effects are present, the biases and root mean squares errors are large for estimators assuming no time-specific effects. On the other hand, even in the case of no time-specific effects in the true data generating process, there is hardly any efficiency loss for the FGLS or GMM that makes the correction of presumed presence of time-specific effects. Therefore, if an investigator is not sure if the assumption of cross-sectional independence is valid or not, it might be advisable to use estimators that take account both individual- and time-specific effects.

13.8 Estimation of Multiplicative Models

In this section we consider the estimation of Equation 13.1, where v_{it} is assumed to be of the form

$$v_{it} = \alpha_i \lambda_t + \epsilon_{it}. \tag{13.73}$$

When α_i is independently distributed across i with mean 0 and variance σ_α^2 and λ_t is independently distributed over t with mean 0 and variance σ_λ^2, $Ev_{it} = 0$, $Ev_{it}^2 = \sigma_\epsilon^2 + \sigma_\alpha^2\sigma_\lambda^2 = \sigma_v^2$, and $Ev_{it}v_{is} = 0$ for $t \neq s$, $Ev_{it}v_{js} = 0$ for $i \neq j$. In other words, Equation 13.1 has error terms that are uncorrelated over time and across individuals, with constant variance σ_v^2. Hence the least squares estimator is consistent and asymptotically normally distributed either N or T or both tend to infinity.

When α_i and λ_t are treated as fixed constants, the MLE are inconsistent if T is finite for the same basic reason as the additive model (Equation 13.2). Ahn, Lee, and Schmidt (2001), Bai (2007), Kiefer (1980), etc., have proposed a nonlinear GMM and iterative LS estimators for the static model with multiplicative effects. Their nonlinear GMM approach can be similarly generalized to obtain a consistent estimator of ρ (e.g., Hsiao 2008).

Let $\theta_t = \lambda_t/\lambda_{t-1}$, then

$$(y_{it} - \theta_t y_{i,t-1}) = \rho(y_{i,t-1} - \theta_t y_{i,t-2}) + (\epsilon_{it} - \theta_t \epsilon_{i,t-1}), \ t = 2, \ldots, T. \tag{13.74}$$

It follows that

$$E[y_{i,t-j}(\epsilon_{it} - \theta_t \epsilon_{i,t-1})] = 0, \ \text{for } j = 2, \ldots, t. \tag{13.75}$$

Let

$$\underset{\frac{T(T-1)}{2}\times(T-1)}{W_i} = \begin{bmatrix} \underset{\sim}{q}_{i2} & \underset{\sim}{0} & \cdot & \cdot & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{q}_{i3} & \cdot & \cdot & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \underset{\sim}{0} & \cdot & \cdot & \cdot & \underset{\sim}{q}_{iT} \end{bmatrix},$$

$$\underset{(T-1)\times(T-1)}{\Theta} = \begin{bmatrix} \theta_2 & 0 & \cdot & \cdot & 0 \\ 0 & \theta_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \theta_T \end{bmatrix},$$

$$\underset{\sim}{q}'_{it} = (y_{i0}, \ldots, y_{i,t-2}), \quad t = 2, \ldots, T,$$

$$\underset{\sim}{\epsilon}_i = (\epsilon_{i2}, \ldots, \epsilon_{iT})', \underset{\sim}{\epsilon}_{i,-1} = (\epsilon_{i1}, \ldots, \epsilon_{i,T-1})'.$$

Then a GMM estimator of ρ and Θ can be obtained from the moment conditions

$$E[W_i(\underset{\sim}{\epsilon}_i - \Theta\underset{\sim}{\epsilon}_{i,-1})] = \underset{\sim}{0}. \tag{13.76}$$

The nonlinear GMM estimators of ρ and Θ amount to applying nonlinear three-stage least squares to the system

$$\underset{\sim}{y}_i = [\rho I_{T-1} + \Theta]\underset{\sim}{y}_{i,-1} - \rho\Theta\underset{\sim}{y}_{i,-2} + \underset{\sim}{\epsilon}_i - \Theta\underset{\sim}{\epsilon}_{i,-1},\ i = 1, \ldots, N, \tag{13.77}$$

using W_i as instruments, where $\underset{\sim}{y}_i = (y_{i2}, \ldots, y_{iT})'$, $\underset{\sim}{y}_{i,-1} = (y_{i1}, \ldots, y_{i,T-1})'$, and $\underset{\sim}{y}_{i,-2} = (y_{i0}, \ldots, y_{i,T-2})'$.

The nonlinear GMM estimators of ρ and θ_t are consistent and asymptotically normally distributed as $N \to \infty$. From the θ_t, we can solve for λ_t through the normalization rule $\lambda_1 = 1$ or $\sum_{t=1}^{T}\lambda_t^2 = 1$. From ρ and λ_t, we obtain

$$\hat{\alpha}_i = \frac{1}{\sum_{t=1}^{T}\hat{\lambda}_t^2}\left[\sum_{t=1}^{T}\hat{\lambda}_t y_{it} - \hat{\rho}\sum_{t=1}^{T}\hat{\lambda}_t y_{i,t-1}\right], i = 1, \ldots, N. \tag{13.78}$$

The estimator (Equation 13.78) is consistent if $T \to \infty$.

The implementation of nonlinear GMM is quite complicated, Pesaran (2006, 2007) notes that

$$\bar{y}_t = \rho\bar{y}_{t-1} + \bar{\alpha}\lambda_t + \bar{\epsilon}_t, \tag{13.79}$$

where

$$\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it},\ \bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i,\ \bar{\epsilon}_t = \frac{1}{N}\sum_{i=1}^{N}\epsilon_{it}.$$

When $N \rightarrow \infty$, $\bar{\epsilon}_t \longrightarrow 0$. Assuming $\bar{\alpha} \neq 0$, substituting $\lambda_t = \bar{\alpha}^{-1}(\bar{y}_t - \rho \bar{y}_{t-1})$ into Equation 13.45 yields,

$$y_{it} = \rho y_{i,t-1} + \gamma_{1i}\bar{y}_t + \gamma_{2i}\bar{y}_{t-1} + \epsilon_{it} \tag{13.80}$$

Therefore, Pesaran (2006, 2007) suggests estimating the cross-sectional mean augment regression (Equation 13.80) and shows that as both N and $T \rightarrow \infty$, the least squares estimator of Equation 13.80 yields consistent and asymptotically normally distributed $\hat{\rho}$.

13.9 Test of Additive versus Multiplicative Model

Multiplicative model implies departure from additivity in their effects on outcomes. It is shown by Bai (2007) that the additive model is embedded into the model of multiple common factors with heterogeneous response by letting

$$\underset{\sim}{\alpha}_i = \begin{bmatrix} \alpha_i \\ 1 \end{bmatrix}, \underset{\sim}{\lambda}_t = \begin{bmatrix} 1 \\ \lambda_t \end{bmatrix},$$

then Equation 13.2 becomes

$$v_{it} = \underset{\sim}{\alpha}_i' \underset{\sim}{\lambda}_t + \epsilon_{it}. \tag{13.81}$$

When $N \longrightarrow \infty$, one may solve $\underset{\sim}{\lambda}_t$ from Equation 13.79 that yields

$$\hat{\underset{\sim}{\lambda}}_t = (\bar{\underset{\sim}{\alpha}}\bar{\underset{\sim}{\alpha}}')^{-}\bar{\underset{\sim}{\alpha}}(v_t - \rho \bar{y}_{t-1}), \tag{13.82}$$

where $(\bar{\underset{\sim}{\alpha}}\bar{\underset{\sim}{\alpha}}')^{-}$ denotes the generalized inverse of $(\bar{\underset{\sim}{\alpha}}\bar{\underset{\sim}{\alpha}}')$. Substituting Equation 13.82 into Equation 13.45 again yields Equation 13.80. Therefore, the Pesaran cross-sectional mean augmented regression of Equation 13.80 is consistent whether the unobserved heterogeneity is additive or multiplicative, but Equation 13.80 is inefficient if the unobserved heterogeneities are additive compared to Equation 13.58 or Equation 13.63. However, if the underlying model is multiplicative, Equation 13.80 is consistent, but not Equation 13.58 or Equation 13.63. Therefore, a Hausman type specification test can be proposed to test the null:

H_0: Equation 13.2 holds

versus

H_1: Equation 13.2 does not hold

by considering the test statistic

$$\frac{\hat{\rho}_A - \hat{\rho}_m}{\sqrt{\mathrm{Var}\,(\hat{\rho}_m) - \mathrm{Var}(\hat{\rho}_A)}} \sim N(0, 1), \tag{13.83}$$

where $\hat{\rho}_A$ denotes the efficient estimator of Equation 13.1 under the additive assumption (Equation 13.2) and $\hat{\rho}_m$ is the estimator (Equation 13.1) under the multiplicative assumption (Equation 13.73).

13.10 Concluding Remarks

In this chapter we review three fundamental issues of modeling dynamic panel data in the presence of unobserved heterogeneity across individuals and over time—the fixed effects of modeling unobserved individual- and time-specific heterogeneity versus random effects; additive versus multiplicative effects and the likelihood versus methods of moments approach.

We have not discussed issues of modeling multivariate dynamic panel models (e.g., Binder, Hsiao, and Pesaran (2005), panel unit root tests (e.g., Breitung and Pesaran 2008; Moon and Perron 2004; Phillips and Sul 2003); parameter heterogeneity (e.g., Hsiao and Pesaran 2008), etc. However, in principle, those issues can also be put in these perspectives.

The advantage of the fixed effects specification is that there is no need to specify the relations between the unobserved effects and observed conditional (or explanatory) variables. The disadvantages are that (1) unless both cross-sectional dimension and time dimension of panels are large, the fixed effects specification introduces incidental parameters issues on the individual-specific effects, α_i, if the time dimension is fixed and on the time-specific effects, λ_t if the cross-sectional dimension is small; (2) the impact of time-invariant but individual-specific variables such as gender or socio-demographic background variables with the presence of additive individual-specific effects and the impact of time-specific but individual invariant such as price and some macro-variables with the presence of additive time-specific effects are unidentified; and (3) the fixed effects inference only makes use of within-group variation. The between group information is ignored.

The advantages of random-effects specification are (1) there are no incidental parameter issues; (2) the impacts of observed individual-specific but time-invariant and individual-invariant but time-varying variables can be identified; (3) both the within-group and between group information are used for inference. Since the between group variation in general is much larger than the within group variation, the RE specification can lead to much more efficient use of sample information. The disadvantage is that the relationship between the unobserved effects and observed conditional variables need to be specified. In short, the advantages of random effects specification are the disadvantage of fixed effects specification and the advantages of fixed effects specification are the disadvantages of random effects specification.

Statistical inference procedures for additive effects models are simpler than the multiplicative effects models. However, if the data generating process calls for a multiplicative effects specification, statistical inference procedures based on additive effects specification will be misleading. On the other hand, if the effects are additive, statistical procedures based on multiplicative effects will also be misleading. In this chapter, we have proposed a testing procedure for additive versus multiplicative effects.

Inference procedures based on the likelihood and moments approaches are reviewed. The likelihood approach uses a fixed number of moment conditions. The moment conditions used in the moments approach increase at the order of square of time series dimension of the panel. In finite sample the moments approach is likely to generate larger bias than the likelihood approach as shown in the Monte Carlo by Binder, Hsiao and Pesaran (2005), Hsiao and Tahmiscioglu (2008), Hsiao, Pesaran, and Tahmiscioglu (2002), Ziliak (1997), etc. Moreover, if the observed outcomes in the time dimension is persistent (when the coefficient of lagged variables, ρ, is close to one) or if the variance of individual-specific effects is large relative to overall variance, the moments approach either breaks down or suffers from the weak instrumental variables issue, but the performance of the likelihood approach is not affected.

13.11 Acknowledgment

I would like to thank a referee for helpful comments.

References

Ahn, S. C., and P. Schmidt. 1995. Efficient Estimation of Models for Dynamic Panel Data. *Journal of Econometrics* 68:5–27.

Ahn, S. G., Y. H. Lee, and P. Schmidt. 2001. GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects. *Journal of Econometrics* 101:219–255.

Amemiya, T., and W. A. Fuller. 1967. A Comparative Study of Alternative Estimators in a Distributed-Lag Model. *Econometrica* 35:509–529.

Amemiya, T., and T. E. MaCurdy. 1986. Instrumental-Variable Estimation of an Error-Components Model. *Econometrica* 54, 869–880.

Anderson, T. W. and C. Hsiao. 1981. Estimation of Dynamic Models with Error Components. *Journal of American Statistical Association* 76:598–606.

Anderson, T. W., and C. Hsiao. 1982. Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics* 18:47–82.

Arellano, M., and S. R. 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.

Arellano, M., and O. Bover. 1995. Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics* 68:29–51.

Bai, J. 2009. Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77, 1229–1279.

Bhargava, A., and D. Sargan. 1983. Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods. *Econometrica* 51:1635–1659.

Binder, M., C. Hsiao, and M. H. Pesaran. 2005. Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration. *Econometric Theory* 21:795–837.

Blundell, R., and S. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87:115–143.

Breitung, J., and M. H. Pesaran. 2008. Unit Roots and Cointegration in Panels. In *The Econometrics of Panel Data*. Berlin: Springer. pp. 279–322.

Breusch, T., G. E. Mizon, and P. Schmidt. 1989. Efficient Estimation Using Panel Data. *Econometrica* 57:695–700.

Cheng, L. K., and Y. K. Kwan. 2000. The Location of Foreign Direct Investment in Chinese Regions – Further Analysis of Labor Quality. In *The Role of Foreign Direct Investment in East Asian Economic Development*. T. Ito and A.O. Krueger (eds). Chicago: Chicago University Press. pp. 213–238.

Hahn, J., and G. Kuersteiner. 2002. Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large. *Econometrica* 70:1639–1659.

Hahn, J., and H. R. Moon. 2006. Reducing Bias of MLE in a Dynamic Panel Model. *Econometric Theory* 22:499–512.

Harris, M. N., L. Matyas, and P. Sevestre. 2008. Dynamic Models for Short Panels. In *The Econometrics of Panel Data*. 3rd ed. L. Matyas and P. Sevestre (eds). Berlin: Springer. pp. 249–278.

Hayakawa, K. 2009. On the Effect of Mean-Nonstationary Initial Conditions in Dynamic Panel Data Models. *Journal of Econometrics* (forthcoming).

Hsiao, C. 2003. *Analysis of Panel Data*. 2nd ed. Econometric Society Monograph 36, New York: Cambridge University Press.

Hsiao, C. 2007. Panel Data Analysis – Advantages and Challenges. *Test* 16:1–22.

Hsiao, C. 2008. Dynamic Panel Data Models with Interactive Effects. *mimeo*.

Hsiao, C., and M. H. Pesaran. 2008. Random Coefficients Models. *The Econometrics of Panel Data*, 3rd ed. L. Matayas and P. Sevestre (eds). Berlin: Springer. pp. 187–216.

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., and A. K. Tahmiscioglu. 2008. Estimation of Dynamic Panel Data Models with Both Individual- and Time-Specific Effects. *Statistical Planning and Statistics Inference* 138:2698–2721.

Kiefer, N. 1980. Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance. *Journal of Econometrics* 14:195–202.

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. 2002. *Essays in Panel Data Econometrics*. Cambridge: Cambridge University Press.

Neyman, J., and E. Scott. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16:1–32.

Pesaran, M. H. 2006. Estimation and Inference in Large Heterogeneous Panels with Cross-Section Dependence. *Econometrica* 74:967–1012.

Pesaran, M. H. 2007. A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence. *Journal of Applied Econometrics* 22:265–312.

Phillips, P.C.B., and D. Sul. 2003. Dynamic Panel Estimation and Homogeneity Testing Under Cross-Section Dependence. *Econometrics Journal* 6: 217–259.

Ziliak, J. P. 1997. Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators, *Journal of Business and Economic Statistics* 15:419–431.