CHAPTER 10

Interactive Effects Models

10.1 INTRODUCTION

An economic model is a simplification of reality. A model hopes to capture the fundamental causal relations of a few variables that are considered important to an investigator. It is a common practice in economic modeling to relegate the impact of omitted variables as random error terms. However, if the impact of omitted variables are correlated with the included conditioning variables, inference on the causal relations of the included variables could be misleading. The impacts of omitted variables in panel data are typically driven by three types of variables: (i) variables that stay constant over time but vary across individuals such as gender, race, family background variables, etc.; (ii) variables that are the same across individuals but vary over time, such as the price of a product, interest rate, etc.; and (iii) variables that vary across individuals and over time, such as age, work experience, etc. With panel data, it is possible to separate the impacts due to individual varying but time-invariant variables and the impacts due to individual-invariant but time-varying variables from the impacts of individual-time varying variables. We have discussed issues of statistical inference for a linear regression model where the unobserved heterogeneity over i and t, v_{it}, is put in the additive form as the sum of individual-specific effects, α_i, time-specific effect, λ_t, and individual-time varying effects, u_{it},

$$v_{it} = \alpha_i + \lambda_t + u_{it}, \tag{10.1.1}$$

in Chapters 2 and 3. In this chapter, we consider inference issues when the individual-specific and time-specific effects are put in multiplicative,

$$v_{it} = \boldsymbol{b}_i' \boldsymbol{f}_t + u_{it}, \tag{10.1.2}$$

where $\boldsymbol{b}_i$ is $(r \times 1)$ time-invariant but varying over i, and $\boldsymbol{f}_t$ is $r \times 1$ individual-invariant but varying over t. We assume $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}_t \boldsymbol{f}_t' \to \sum_f$ and $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{b}_i' = \sum_b$, where $\sum_f$ and $\sum_b$ are positive definite matrices with rank r, $r \geq 1$.

The multicative form encompasses additive form as a special case. Let $r = 2$, $\boldsymbol{b}_i' = (\alpha_i, 1)$, and $\boldsymbol{f}_t = (1, \lambda_t)'$; then (10.1.2) becomes (10.1.1). When the unobserved individual-specific and/or time-specific effects are put in additive form, there is no way to separately identify the impact of each of those time-invariant but individual varying variables and/or time-varying but individual-invariant variables due to perfect multicollinearity. Only the joint impacts of those individual varying but time-invariant variables and/or time-varying but individual-invariant variables can be identified. When they are put in the multiplicative form, it allows the identification of the "r-dimensional common shocks at time t that affect all individuals," while the impact of each element of the common shocks $\boldsymbol{f}_t$ to the ith individual $\boldsymbol{b}_i$ can be different due to differences in "individual's innate ability" arising from differences in natural endowment, distinct social background, etc., or alternatively identifying for the ith individual the impact of each time-invariant variables to the common time-varying shocks.

However, it is the product of $\boldsymbol{b}_i'\boldsymbol{f}_t$ that affects v_{it}, not the individual elements of $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$. Without prior knowledge, it is not possible to identify $\boldsymbol{b}_i'$ and $\boldsymbol{f}_t$ because

$$\boldsymbol{b}_i'\boldsymbol{f}_t = \boldsymbol{b}_i^{*'}\boldsymbol{f}_t^* \tag{10.1.3}$$

for any $\boldsymbol{b}_i^* = C\boldsymbol{b}_i$ *and* $\boldsymbol{f}_t^* = C^{'-1}\boldsymbol{f}_t$ for any $r \times r$ nonsingular constant matrix C. To uniquely identify $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$, Anderson and Rubin (1956) suggest the following normalization conditions:

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t\boldsymbol{f}_t' = I_r \tag{10.1.4}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{b}_i\boldsymbol{b}_i' = D, \tag{10.1.5}$$

while D is an $r \times r$ nonsingular diagonal matrix. Condition (10.1.4) imposes $\frac{r(r+1)}{2}$ restrictions on C. Condition (10.1.5) further imposes $\frac{r(r-1)}{2}$ restrictions on C.[1]

The individual-specific effects, $\boldsymbol{b}_i$, and time-specific effects, $\boldsymbol{f}_t$, are unobserved. Just like the additive effects models, they can be either assumed random or fixed. In this chapter, we shall first consider treating $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ as unknown constants, then consider treating $\boldsymbol{b}_i$ or $\boldsymbol{f}_t$ as random. The advantages of treating $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ as fixed constants are: (i) there is no need to specify the data generating process of $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$; and (ii) it allows the observed variables to be correlated with $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$. The disadvantage is that the unknown $\boldsymbol{b}_i$ increases with the cross-sectional dimension N, and the unknown $\boldsymbol{f}_t$ increases with time dimension T. The advantages of treating $\boldsymbol{b}_i$ or $\boldsymbol{f}_i$ as random are that it simplifies the computation as well as improves the efficiency of the estimates of the coefficients of included observed covariates. The disadvantage is that a specific data generating process for $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ needs to be assumed.

In Section 10.2, we discuss the estimation of the coefficients of observed covariates, $\boldsymbol{x}_{it}$ that are strictly exogenous with respect to the idiosyncratic errors, u_{it}. Section 10.3 considers the estimation when $\boldsymbol{x}_{it}$ includes lag dependent variables when $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are treated as unknown constants, respectively. Section 10.4 consider models of either treating $\boldsymbol{b}_i$ random $\boldsymbol{f}_t$ fixed or treating $\boldsymbol{f}_t$ random and $\boldsymbol{b}_i$ fixed, Section 10.5 considers quantile estimation of interactive effects models. Section 10.6 considers issues of detecting the dimension of common factors.

[1] Condition (10.1.4) restricts C only to be an orthogonal matrix $C\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t\boldsymbol{f}_t'\right]C' = I_r$. Condition (10.1.5) further restricts C to be an identify matrix. Even in this case uniqueness is only up to a sign change. For instance, $-\boldsymbol{f}_t$ and $-\boldsymbol{b}_i$ also satisfy the restrictions. However, the covariance between v_{it} and v_{jt} remains the same, $E(v_{it}v_{jt}) = \boldsymbol{b}_i'\boldsymbol{b}_j = \boldsymbol{b}_i^{*'}\boldsymbol{b}_j^* = \boldsymbol{b}_i'CC'\boldsymbol{b}_j$ for any $r \times r$ orthonormal matrix.

10.2 FIXED-EFFECTS LINEAR STATIC MODELS

Consider the model of the form

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{10.2.1}$$

$$v_{it} = \boldsymbol{b}_i'\boldsymbol{f}_t + u_{it}, \tag{10.2.2}$$

where $\boldsymbol{x}_{it}$ are a $K \times 1$ vector of observed variables, $\boldsymbol{\beta}$ is a $K \times 1$ vector of constants, $\boldsymbol{f}_t$ is an $r \times 1$ vector of time-specific effects that varies over t, $\boldsymbol{b}_i$ is an $r \times 1$ vector of individual-specific effects that varies over i, and u_{it} denotes the error that varies over i and t with $E(u_{it}|\boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{x}_{it}) = 0$. We assume $\boldsymbol{x}_{it}$ are strictly exogenous with respect to u_{is}, i.e., $E(u_{is}|\boldsymbol{x}_{it}) = 0$ for all t and s.

When $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are treated as unknown constants, we may consider y_{it} driven by the K observed variables, $\boldsymbol{x}_{it}$, and r unobserved variables $\boldsymbol{f}_t$, with coefficients $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, respectively. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})', X_i = (\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT}')', F = (f_1, \ldots, f_T)'$ and $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})$, We rewrite model (10.2.1) and (10.2.2) in vector form,

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + F\boldsymbol{b}_i + \boldsymbol{u}_i, \quad i = 1, \ldots, N. \tag{10.2.3}$$

10.2.1 Bai (2009) Least Squares Estimator

Under the assumption that u_{it} is independently identically distributed over i and t with constant variance σ_u^2, Bai (2009a) suggests minimizing

$$S = \sum_{i=1}^{N}(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - F\boldsymbol{b}_i)'(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - F\boldsymbol{b}_i) \tag{10.2.4}$$

with respect to $\boldsymbol{\beta}$, F and $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)$, subject to (10.1.4) and (10.1.5). Taking partial derivatives of S with respect to $\boldsymbol{b}_i$,

$$\frac{\partial S}{\partial \boldsymbol{b}_i} = -2(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - F\boldsymbol{b}_i)'F = \boldsymbol{0}. \tag{10.2.5}$$

Substituting the solution of $\boldsymbol{b}_i$ from (10.2.5) to (10.2.4), then solving the resulting concentrated S yields the least squares estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i'MX_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'M\boldsymbol{y}_i\right) \tag{10.2.6}$$

where $M = I_T - F(F'F)^{-1}F'$. Conditional on $\boldsymbol{\beta}$, the residual v_{it} is a pure factor structure (10.2.2). The least squares estimator of F is $\sqrt{T}$ times the r eigenvectors corresponding to the r largest eigenvalues of the matrix (Anderson 1985),

$$\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{y}_i - X_i\boldsymbol{\beta})(\boldsymbol{y}_i - X_i\boldsymbol{\beta})', \tag{10.2.7}$$

i.e.

$$\left[\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{y}_i - X_i\boldsymbol{\beta})(\boldsymbol{y}_i - X_i\boldsymbol{\beta})'\right]\hat{F} = \hat{F}\Lambda \tag{10.2.8}$$

where Λ is a diagonal matrix that consists of the r largest eigenvalues of (10.2.7). Conditional on $(\hat{\boldsymbol{\beta}}, \hat{F})$, the least squares estimator of B is

$$\hat{B}' = T^{-1}\left[\hat{F}'(\boldsymbol{y}_1 - X_1\hat{\boldsymbol{\beta}}), \ldots, \hat{F}'(\boldsymbol{y}_N - X_N\hat{\boldsymbol{\beta}})\right]. \tag{10.2.9}$$

Therefore, Bai (2009a) suggests iterating between (10.2.6), (10.2.8), and (10.2.9) until convergence to obtain the least squares estimator of $\boldsymbol{\beta}$.

When both N and T are large, the least squares estimator (10.2.6) is consistent and asymptotically normally distributed with the covariance matrix

$$\sigma_u^2\left(\sum_{i=1}^{N} X_i' M X_i\right)^{-1} \tag{10.2.10}$$

if u_{it} is independently identically distributed with mean 0 and constant variance σ_u^2. However, if u_{it} is heteroscedastic and weakly cross-sectionally or serially correlated,[2] when $\frac{T}{N} \to c \neq 0$ as $N, T \to \infty, \sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically biased of the form

$$\left(\frac{T}{N}\right)^{\frac{1}{2}} C + \left(\frac{N}{T}\right)^{\frac{1}{2}} D, \tag{10.2.11}$$

where C denotes the bias induced by heteroscadasticity and cross-sectional correlation, and D denotes the bias induced by serial correlation and heteroscadasticity of u_{it}. Bai (2009a) has provided the formulas for constructing the bias-corrected estimator.

For models with both additive and interactive effects such as

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \boldsymbol{b}_i'\boldsymbol{f}_t + u_{it}, \tag{10.2.12}$$

one may view one of the r-dimensional shocks as known a priori, $f_{1t} = 1$ for all t; hence, there is no need to estimate f_{1t}. To estimate (10.2.12), just like the standard fixed-effects estimator (Chapter 2), we can first take the deviation of individual observations from its time series mean, $\tilde{y}_{it} = y_{it} - \bar{y}_i, \tilde{x}_{it} = \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i$, to get rid of α_i from (10.2.12). However, now in addition to the normalization conditions, $F'F = I_r$ and $B'B$ diagonal, we also need to impose the restriction $\sum_{i=1}^N \alpha_i = 0$, $\sum_{i=1}^N \boldsymbol{b}_i = \boldsymbol{0}$, $\sum_{t=1}^T \boldsymbol{f}_t = \boldsymbol{0}$, to obtain a unique solution of $(\boldsymbol{\beta}, \alpha_i, \boldsymbol{b}_i, \boldsymbol{f}_t)$, (Bai 2009a, p. 1253). The least squares estimater of (10.2.12) is now

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{M}\tilde{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{M}\tilde{\boldsymbol{y}}_i\right) \tag{10.2.13}$$

where $\tilde{\boldsymbol{y}}_i, \tilde{X}_i$ denote the stacked T times series observations of $\tilde{y}_{it}$, and $\tilde{\boldsymbol{x}}_{it}, \tilde{M} = I - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$, and $\hat{F}$ is the $T \times r$ matrix consisting of the first r eigenvectors (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix

$$\frac{1}{NT}\sum_{i=1}^{N}(\tilde{\boldsymbol{y}}_i - \tilde{X}_i\hat{\boldsymbol{\beta}})(\tilde{\boldsymbol{y}}_i - \tilde{X}_i\hat{\boldsymbol{\beta}})'. \tag{10.2.14}$$

[2] Loosely speaking, weakly cross-correlation can be viewed as the number of cross-sectional u_{jt} that are correlated with u_{it}, $n(i)$, $(n(i)/N) \to 0$ as $N \to \infty$ (for further discussion, see Section 11.1). Weakly time correlation can be viewed as $\sum_{s=0}^{\infty} |\sigma_i(s)| < M < \infty$ where $\sigma_i(s) = E(u_{it}u_{is})$. (For rigorous definition, see Anderson 1971; Bai 2009).

Starting from some initial estimates of $\boldsymbol{\beta}$ and F, the least squares estimator can be obtained by iterating between (10.2.13) and (10.2.14) until the solution converges. After convergent solutions of $\hat{\boldsymbol{\beta}}$ and $\hat{F}$ are obtained, one can obtain $\hat{\alpha}_i$ and $\hat{B}'$ by

$$\hat{\alpha}_i = \bar{y}_i - \bar{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}}, \tag{10.2.15}$$

$$\hat{B}' = T^{-1}[\hat{F}'(\tilde{\boldsymbol{y}}_1 - \tilde{X}_1 \hat{\boldsymbol{\beta}}), \ldots, \hat{F}'(\tilde{\boldsymbol{y}}_N - \tilde{X}_N \hat{\boldsymbol{\beta}})]. \tag{10.2.16}$$

Although the Bai (2009) least squares estimator is consistent and asymptotically normally distributed, simultaneously solving for the first-order conditions (10.2.6), (10.2.8), and (10.2.9) can be computationally complicated (e.g., Hsiao 2018). On the other hand, the recursively iterating between (10.2.6) and (10.2.7) until convergence is computationally feasible. However, the recursive estimation procedure can be contaminated by the inconsistent $\tilde{\beta}^{(j)}$. If $\left|\tilde{\beta}^{(j)} - \beta\right| = O(1)$, then rank $(\frac{1}{T}\hat{F}^{(j)\prime}F)$ could be less than r. In the extreme case $r = 0$, then $\tilde{\beta}^{(j+1)} = \tilde{\beta}^{(j)}$. In general, we cannot rule out the possibility that $\left|\tilde{\beta}^{(j)} - \beta\right| = O(1)$ if the initial estimator is inconsistent; for instance, if the least squares estimator of y_{it} on $\boldsymbol{x}_{it}$ is used as the initial estimator in the case that $\boldsymbol{x}_{it}$ and $\boldsymbol{b}_i' \boldsymbol{f}_t$ are correlated (see Jiang et al. 2020 for details).

10.2.2 Quasi-Difference Estimator

Ahn, Lee, and Schmidt (2001, 2013) proposed a nonlinear GMM method to estimate a linear panel data model with interactive effects. For ease of exposition, suppose $r = 1$. Let $\theta_t = \frac{f_t}{f_1}$, then

$$(y_{it} - \theta_t y_{i1}) = \boldsymbol{x}_{it}' \boldsymbol{\beta} - \boldsymbol{x}_{i,t-1}' \boldsymbol{\beta} \theta_t + (u_{it} - \theta_t u_{i1}), \; t = 2, \ldots, T. \tag{10.2.17}$$

It follows that

$$E[\boldsymbol{x}_i (u_{it} - \theta_t u_{i1})] = \boldsymbol{0}. \tag{10.2.18}$$

Let $W_i = I_{T-1} \otimes \boldsymbol{x}_i$: $\boldsymbol{x}_i = (\boldsymbol{x}_i', \ldots, \boldsymbol{x}_{it}')$,

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

$$\tilde{\boldsymbol{u}}_i = (u_{i2}, \ldots, u_{iT})', \; \tilde{\boldsymbol{u}}_{i1} = (u_{i1}, \ldots, u_{i1})'.$$

Then a GMM estimator of $\boldsymbol{\beta}$ and $\ominus$ can be obtained from the moment conditions,

$$E[W_i(\tilde{\boldsymbol{u}}_i - \ominus \tilde{\boldsymbol{u}}_{i1})] = \boldsymbol{0}. \tag{10.2.19}$$

The nonlinear GMM estimator is consistent and asymptotically normally distributed when $N \to \infty$ under fixed T; even u_{it} is serially correlated and heteroscedastic. However, the computation can be very cumbersome when $r > 1$. The successive quasi-differencing yield highly nonlinear moment conditions. For instance, if $r = 2$, we will have to let $\theta_t = \frac{f_t}{f_1}$, and $\delta_t = f_{2t} - f_{22}\theta_t$ for $t = 3, \ldots, T$, in order to take the quasi-difference of $(y_{it} - \theta_t y_{i,t-1})$ equation one more time to eliminate the factor error. When $r > 1$, one will have to take r such successive differences to get rid of $\boldsymbol{b}_i' \boldsymbol{f}_t$, yielding a highly nonlinear moment conditions. When the moment conditions are highly nonlinear, multiple solutions may occur. Additional conditions have to be imposed to ensure that the solution converges to the true value.

10.2.3 Group Mean Augmented (Common Correlated Effects (CCE)) Approach

The feasible least square approach of iteratively estimating (10.2.6) and (10.2.8) works only if both N and T are large. Nevertheless, when $N \longrightarrow \infty, \bar{u}_t = \frac{1}{N}\sum_{i=1}^{N} u_{it} \longrightarrow 0$, model (10.2.1) and (10.2.2) (or (10.2.3)) imply that

$$\bar{\boldsymbol{b}}'\boldsymbol{f}_t \simeq \bar{y}_t - \overline{\boldsymbol{x}}'_t\boldsymbol{\beta}, \tag{10.2.20}$$

where $\bar{\boldsymbol{b}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{b}_i$, $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$, and $\bar{\boldsymbol{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}$. If $\boldsymbol{b}_i'\boldsymbol{f}_t = c_i\bar{\boldsymbol{b}}'\boldsymbol{f}_t$, for all t, or if $\boldsymbol{f}_t$ can be approximated by linear combinations of $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$, instead of estimating $\hat{\boldsymbol{f}}$, Pesaran (2006) suggests a simple approach to filter out the impact of $\boldsymbol{b}_i'\boldsymbol{f}_t$ by augmenting (10.2.12) by $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$,

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \bar{y}_t c_i + \bar{\boldsymbol{x}}_t'\boldsymbol{d}_i + e_{it}. \tag{10.2.21}$$

The pooled estimator

$$\hat{\boldsymbol{\beta}}^* = \left(\sum_{i=1}^{N} w_i X_i' M^* X_i\right)^{-1}\left(\sum_{i=1}^{N} w_i X_i' M^* \boldsymbol{y}_i\right) \tag{10.2.22}$$

is consistent and asymptotically normally distributed when $N \rightarrow \infty$ and T either is fixed or $\rightarrow \infty$, where $w_i = \frac{\sigma_i^2}{\sum_{j=1}^{N}\sigma_j^2}, \sigma_j^2 = \text{Var}(u_{jt}), M^* = (I - H(H'H)^{-1}H'), H = (\boldsymbol{e}, \bar{\boldsymbol{y}}, \bar{X})$ with $\boldsymbol{e}, a, T \times 1$ vector of 1's, $\bar{\boldsymbol{y}}$ and $\bar{X}$ are $T \times 1$ and $T \times K$ stacked $\bar{y}_t$ and $\bar{\boldsymbol{x}}'_t$, respectively.[3] Pesaran (2006) called (10.2.22) the common correlated effects pooled estimator (CCEP). The limited Monte Carlo studies conducted by Westerlund and Larsson (2012) appear to show that the Pesaran (2006) CCEP estimator of $\boldsymbol{\beta}$ (10.2.22) is less biased than the Bai (2009a) iterated least squares estimator (10.2.6).

Kapetanios, Pesaran, and Yamagata (2011) further show that the cross-section-average-based method is robust to a wide variety of data generating processes. For instance, for the error process generated by a multi-factor error structure (10.1.1), whether the unobservable common factors $\boldsymbol{f}_t$ follow $I(0)$ or unit root processes, the asymptotic properties of (10.2.22) remain similar.

The advantage of Pesaran's (2006) cross-sectional mean-augmented approach to take account of the cross-sectional dependence is its simplicity, and it requires only N large. However, there are restrictions on its application. The method works when $\boldsymbol{b}_i'\boldsymbol{f}_t = c_i\bar{\boldsymbol{b}}'\boldsymbol{f}_t$ for all t, or if $\boldsymbol{f}_t$ can be considered as a linear combinations of $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$. It is hard to ensure $\boldsymbol{b}_i'\boldsymbol{f}_t = c_i\bar{\boldsymbol{b}}'\boldsymbol{f}_t$ for all t if $r > 1$ for a constant c_i. For instance, consider the case that $r = 2, \boldsymbol{b}_i' = (1,1), \bar{\boldsymbol{b}}' = (2,0), \boldsymbol{f}_t' = (1,1)$; then $\boldsymbol{b}_i'\boldsymbol{f}_t = \bar{\boldsymbol{b}}'\boldsymbol{f}_t = 2$. However, if $\boldsymbol{f}_s' = (2,0)$, then $\boldsymbol{b}_i'\boldsymbol{f}_s = 2$, while $\bar{\boldsymbol{b}}'\boldsymbol{f}_s = 4$. If $\boldsymbol{b}_i'\boldsymbol{f}_t$ is not equal to $c_i\bar{\boldsymbol{b}}'\boldsymbol{f}_t$ for all t, the residual of (10.2.21), e_{it}, is correlated with $\boldsymbol{x}_{it}$. The CCE estimate (10.2.22) is inconsistent if $\boldsymbol{b}_i'\boldsymbol{f}_t = c_{it}\bar{\boldsymbol{b}}'\boldsymbol{f}_t$. Additional conditions are needed to approximate $\boldsymbol{b}_i'\boldsymbol{f}_t$. For instance, Pesaran (2006) assumes that

$$\boldsymbol{x}_{it} = \Gamma_i\boldsymbol{f}_t + \boldsymbol{\epsilon}_{it}, \tag{10.2.23}$$

$$E(\boldsymbol{\epsilon}_{it}u_{it}) = \boldsymbol{0}. \tag{10.2.24}$$

[3] When $\sigma_i^2 = \sigma^2$ for all $i, w_i = \frac{1}{N}$.

Then

$$\begin{aligned} z_{it} = \begin{pmatrix} y_{it} \\ \boldsymbol{x}_{it} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\beta}'\Gamma_i + \boldsymbol{b}_i' \\ \Gamma_i \end{pmatrix} \boldsymbol{f}_t + \begin{pmatrix} \boldsymbol{\beta}'\boldsymbol{\epsilon}_{it} + u_{it} \\ \boldsymbol{\epsilon}_{it} \end{pmatrix} \\ &= C_i \boldsymbol{f}_t + \boldsymbol{e}_{it} \end{aligned} \tag{10.2.25}$$

It follows that

$$\bar{z}_t = \frac{1}{N}\sum_{i=1}^{N} z_{it} = \bar{C}\boldsymbol{f}_t + \bar{\boldsymbol{e}}_t \tag{10.2.26}$$

where $\bar{C} = \frac{1}{N}\sum_{i=1}^{N} C_i, \bar{\boldsymbol{e}}_t = \begin{pmatrix} \boldsymbol{\beta}'\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it}\right) + \bar{u}_t \\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it} \end{pmatrix}$. If $r \leq K+1$, and $\bar{C}$ is of rank r, and $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it} \longrightarrow \boldsymbol{0}$ (or $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it} \longrightarrow$ a constant vector) as $N \longrightarrow \infty$, then

$$\boldsymbol{f}_t = (\bar{C}'\bar{C})^{-1}\bar{C}'\bar{z}_t. \tag{10.2.27}$$

Then the model (10.2.3) is formally identical to (10.2.21) where $(c_i, \boldsymbol{d}_i') = \boldsymbol{b}_i'(\bar{C}\bar{C})^{-1}C'$. However, under (10.2.23), (10.2.24), and the additional assumption that

$$\text{Cov}\,(\Gamma_i, \boldsymbol{b}_i) = \boldsymbol{0}, \tag{10.2.28}$$

one can obtain a consistent estimator of $\boldsymbol{\beta}$ by simply adding time dummies to (10.2.1). The least squares dummy-variable estimator of $\boldsymbol{\beta}$ is equivalent to the within (time) estimator of

$$(y_{it} - \bar{y}_t) = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)'\boldsymbol{\beta} + (v_{it} - \bar{v}_t) \tag{10.2.29}$$

(see Section 2.2), where

$$v_{it} - \bar{v}_t = (\boldsymbol{b}_i - \bar{\boldsymbol{b}})'\boldsymbol{f}_t + (u_{it} - \bar{u}_t),$$

$$\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t = (\Gamma_i - \bar{\Gamma})\boldsymbol{f}_t + (\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_t),$$

$$\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}, \bar{\boldsymbol{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}, \bar{\boldsymbol{b}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{b}_i, \bar{\Gamma} = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i, \bar{u}_t = \frac{1}{N}\sum_{i=1}^{N} u_{it},$$

and $\bar{\boldsymbol{\epsilon}}_t = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it}$. Under (10.2.28),

$$\text{Cov}\,(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t, v_{it} - \bar{v}_t) = E\left\{(\Gamma_i - \bar{\Gamma})(\boldsymbol{b}_i - \bar{\boldsymbol{b}})'\right\}\text{Cov}\,(\boldsymbol{f}_t)\,\text{Cov}\,(\boldsymbol{\epsilon}_{it}, u_{it}) = \boldsymbol{0}. \tag{10.2.30}$$

Therefore, as $N \to \infty$, the least squares estimator of (10.2.29),

$$\hat{\boldsymbol{\beta}}_{cv} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)(y_{it} - \bar{y}_t)\right], \tag{10.2.31}$$

is consistent and asymptotically normally distributed with the covariance matrix

$$\text{Cov}\,(\hat{\boldsymbol{\beta}}_{cv}) = \sigma_u^2\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_t)'\right]^{-1}. \tag{10.2.32}$$

However, if $\boldsymbol{x}_{it}$ takes the form

$$\boldsymbol{x}_{it} = \Gamma_i \boldsymbol{f}_t + \boldsymbol{\phi}_i'\tilde{\boldsymbol{f}}_t + \boldsymbol{\varepsilon}_{it}, \tag{10.2.33}$$

where $\tilde{\boldsymbol{f}}_t$ denotes the m common factors that are orthogonal to $\boldsymbol{f}_t$, and the factor loading $\boldsymbol{\phi}_i$ satisfies $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\phi}_i'\boldsymbol{\phi}_i \to \sum_{\phi}$, a nonsingular matrix of rank m when $N \to \infty$, then the average of $\boldsymbol{e}_{it}$,

$$\boldsymbol{e}_{it} = \boldsymbol{\phi}_i'\tilde{\boldsymbol{f}}_t + \boldsymbol{\varepsilon}_{it} \tag{10.2.34}$$

$\bar{e}_t = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{e}_{it}$ will not converge to zero as $N \to \infty$ even under the assumption that

$$E(\boldsymbol{\varepsilon}_{it} \mid \boldsymbol{f}_t, \tilde{\boldsymbol{f}}_t) = \mathbf{0}. \tag{10.2.35}$$

Then neither the Pesaran (2006) CCE, nor the within estimator of (10.2.31) is consistent. Kapetanios, Serlenga, and Shin (2020) proposed a Lagrangian Multiplier Test for the adequacy of the fixed-effects estimator in the presence of factor structure cross-sectional dependence in panel data by testing the independence of the residuals of the covariance transformed model and the regressors.

10.2.4 Transformed Estimator

We note that if $N > T > (K + r)$, then there are uncountably many $(T \times 1)$ vectors $\boldsymbol{w}$ that lie on the null space of $F = (\boldsymbol{f}_1', \ldots, \boldsymbol{f}_T')'$. Multiplying any such $\boldsymbol{w}$ to (10.2.3) eliminates $F\boldsymbol{b}_i$,

$$\boldsymbol{w}'\boldsymbol{y}_i = \boldsymbol{w}'X_i\boldsymbol{\beta} + \boldsymbol{w}'\boldsymbol{u}_i,\ i = 1, \ldots, N, \tag{10.2.36}$$

because $\boldsymbol{w}'F = 0$. The least squares estimator (10.2.36) for $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i'\boldsymbol{w}\boldsymbol{w}'X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\boldsymbol{w}\boldsymbol{w}'\boldsymbol{y}_i\right), \tag{10.2.37}$$

is consistent and asymptotically normally distributed as $N \to \infty$ under fairly general condition (Hsiao, Shi, and Zhou 2021).

However, F is unknown, so $\boldsymbol{w}$ cannot be identified. When $N > T > (K + r)$, Hsiao, Shi, and Zhou (2021) suggest minimizing

$$\boldsymbol{w}'\left[\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - X_i\boldsymbol{\beta})(\boldsymbol{y}_i - X_i\boldsymbol{\beta})'\right]\boldsymbol{w} \tag{10.2.38}$$

subject to

$$\boldsymbol{w}'\boldsymbol{w} = 1, \tag{10.2.39}$$

can yield a consistent and asymptotic normal estimator,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i'\hat{\boldsymbol{w}}\hat{\boldsymbol{w}}'X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\hat{\boldsymbol{w}}\hat{\boldsymbol{w}}'\boldsymbol{y}_i\right), \tag{10.2.40}$$

when $\frac{T}{N} \to a < \infty$ as $N \to \infty$, where $\hat{\boldsymbol{w}}$ is the eigenvector corresponding to the smallest root of the determinant equation

$$\left|\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}})' - \delta\ I_T\right| = 0. \tag{10.2.41}$$

The impact of $\boldsymbol{b}_i'\boldsymbol{f}_t$ is symmetric in $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$. When $T > N > (K + r)$, one can apply the transformed estimator by first stacking N cross-sectional units at time t to express (10.2.1) in the form

$$\boldsymbol{y}_t = X_t\boldsymbol{\beta} + B\boldsymbol{f}_t + \boldsymbol{u}_t, \quad t = 1, \ldots, T, \tag{10.2.42}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $X_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$, $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)'$, and $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})'$. Then there exists an $N \times 1$ vector $\tilde{\boldsymbol{w}}$ lying on the null space of B. Multiplying such a $\tilde{\boldsymbol{w}}$ to (10.2.42) eliminates the term $B\boldsymbol{f}_t$,

$$\tilde{\boldsymbol{w}}'\boldsymbol{y}_t = \tilde{\boldsymbol{w}}'X_t\boldsymbol{\beta} + \tilde{\boldsymbol{w}}'\boldsymbol{u}_t \quad t = 1, \ldots, T. \tag{10.2.43}$$

Therefore, a consistent and asymptotically normally distributed $\tilde{\boldsymbol{\beta}}$ can be derived by minimizing the objective function

$$\tilde{\boldsymbol{w}}'\left(\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_t - X_t\boldsymbol{\beta})(\boldsymbol{y}_t - X_t\boldsymbol{\beta})'\right)\tilde{\boldsymbol{w}} \tag{10.2.44}$$

subject to

$$\tilde{\boldsymbol{w}}'\tilde{\boldsymbol{w}} = 1. \tag{10.2.45}$$

When $\frac{N}{T} \to c < \infty$ as $T \to \infty$, just like stacking each individual's time series observations to (10.2.3), the $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{\beta}}$ can be obtained by iterating between the

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T} X_t'\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}'X_t\right)^{-1}\left(\sum_{t=1}^{T} X_t'\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}'\boldsymbol{y}_t\right), \tag{10.2.46}$$

and the eigenvector corresponding to the smallest root

$$\left|\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_t - X_t\tilde{\boldsymbol{\beta}})(\boldsymbol{y}_t - X_t\tilde{\boldsymbol{\beta}})' - \hat{\delta}^* I_N\right| = 0 \tag{10.2.47}$$

until the solution converges.

The advantages of the transformed estimator are: (i) it requires only either N or T to go to infinity to achieve consistency and asymptotic normality; (ii) it is not affected by structural breaks in the factor loading matrix B, since $\hat{\boldsymbol{w}}'F\boldsymbol{b}_i = 0$ whatever the factor loading matrix B is; (iii) there is no need to know the dimension of common factors of r; and (iv) the asymptotic distribution is not affected by heteroscedasticity in u_{it}. The disadvantage is that it yields no information on B and F, although they can be obtained by applying the principal component analysis (Anderson and Rubin 1956) to the covariance matrix of $\boldsymbol{v}_i$ or $\boldsymbol{v}_t$ once a consistent estimate of $\boldsymbol{\beta}$ is obtained.

10.3 FIXED-EFFECTS LINEAR DYNAMIC MODELS

When $\boldsymbol{x}_{it}$ contains lag dependent variables, say $y_{i,t-1}$, just like the additive effects models, the asymptotic properties no longer stay the same.

10.3.1 The Least Squares Estimation or the Quasi-Difference Estimator

When $\boldsymbol{x}_{it}$ contains predetermined variables, the least squares estimator is the QMLE treating initial values as fixed constants. Just like the dynamic panel models with individual effects only, if the data generating process (DGP) of y_{i0} is no different from the DGP of y_{it}, the QMLE of $\boldsymbol{\beta}$ treating initial values as fixed constants is inconsistent if T is fixed, no matter how large N is. To see this, consider the simple case that $\boldsymbol{x}_{it} = y_{i,t-1}$ and $r = 1$; then

$$y_{it} = y_{i,t-1}\beta + v_{it}, \qquad |\ \beta\ | < 1. \tag{10.3.1}$$

For ease of notation, we also assume y_{i0} is available. The least squares estimator of (10.3.1) under the normalization condition,

$$\sum_{t=1}^{T} f_t^2 = T,$$

is

$$\hat{\beta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \tilde{y}_{i,-1}f_t)^2\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \tilde{y}_{i,-1}f_t)(y_{it} - \tilde{y}_i f_t)\right], \tag{10.3.2}$$

where $\tilde{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}f_t$ and $\tilde{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}f_t$. We note that by continuous substitution,

$$y_{it} = b_i(f_t + \beta f_{t-1} + \cdots + \beta^{t-1}f_1) + \beta^t y_{io} + \sum_{s=0}^{t-1} u_{i,t-s}\beta^s, \tag{10.3.3}$$

$$\begin{aligned}\sum_{t=1}^{T} y_{it}f_t = b_i&\left(\sum_{t=1}^{T} f_t^2 + \beta\sum_{t=2}^{T} f_t f_{t-1} + \beta^2\sum_{t=3}^{T} f_t f_{t-2} + \cdots + \beta^{T-1}f_T f_1\right)\\ &+ y_{io}\left(\sum_{t=1}^{T}\beta^t f_t\right) + \sum_{t=2}^{T} f_t\left(\sum_{s=0}^{t-1} u_{i,t-s}\beta^s\right),\end{aligned} \tag{10.3.4}$$

$$\begin{aligned}\sum_{t=1}^{T} y_{i,t-1}f_t = b_i&\left(\sum_{t=2}^{T} f_t f_{t-1} + \beta\sum_{t=3}^{T} f_t f_{t-2} + \ldots + \beta^{T-2}f_T f_1\right)\\ &+ y_{io}\left(\sum_{t=1}^{T}\beta^{t-1}f_t\right) + \sum_{t=1}^{T} f_t\left(\sum_{s=0}^{t-2} u_{i,t-s-1}\beta^s\right),\end{aligned} \tag{10.3.5}$$

$$\tilde{y}_i - \beta\tilde{y}_{i,-1} = b_i + \frac{1}{T}\sum_{t=1}^{T} f_t u_{it}. \tag{10.3.6}$$

Substituting (10.3.6) into (10.3.2) yields

$$\begin{aligned}\hat{\beta} = \beta + &\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \tilde{y}_{i,-1}f_t)^2\right]^{-1}\\ &\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \tilde{y}_{i,-1}f_t)(u_{it} - \frac{1}{T}\sum_{t=1}^{T} f_t u_{it})\right]\end{aligned} \tag{10.3.7}$$

We note that the numerator of (10.3.7) converges to

$$
\begin{aligned}
&\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} y_{i,t-1}\left(\frac{1}{T}\sum_{s=1}^{T} f_s u_{is}\right) f_t \\
&= \left(-\frac{\sigma_u^2}{T^2}\right)\left\{\sum_{t=2}^{T} f_t f_{t-1} + \beta\sum_{t=3}^{T} f_t f_{t-2} + \beta^2\sum_{t=4}^{T} f_t f_{t-3} + \cdots + \beta^{T-2} f_T f_1\right\} \\
&= O\left(\frac{1}{T}\right). \qquad (10.3.8)
\end{aligned}
$$

The least squares estimator (10.3.2) is identical to the within estimator if $f_t = 1$ for all t. The within estimator of $\boldsymbol{\beta}$ (Chapter 3) is inconsistent when T is fixed no matter how large N is if $\boldsymbol{x}_{it}$ contains lagged dependent variables. The order of bias is $\frac{1}{T}$.

Formula (10.3.8) shows that conditional on $\boldsymbol{f}_t$, the least squares estimator is consistent if $T \to \infty$. However, to get a consistent estimator of $\boldsymbol{f}_t$, we need $N \to \infty$. If both N and T are large and $\frac{N}{T} \to a \neq 0 < \infty$, Moon and Weidner (2017) show that the least squares estimator is consistent, but $\sqrt{NT}(\hat{\boldsymbol{\beta}}_{LS} - \hat{\boldsymbol{\beta}})$ is asymptotically biased of order $\sqrt{N/T}$. Monte Carlo studies conducted by Hsiao and Zhang (2015) show that the validity of statistical inference depends critically on whether an estimator is asymptotically unbiased or not. Moon and Weidner (2017) suggest a bias corrected estimator,

$$
\hat{\boldsymbol{\beta}}^*_{LS} = \hat{\boldsymbol{\beta}}_{LS} + W^{-1}(T^{-1}\hat{C}_1 + N^{-1}\hat{C}_2 + T^{-1}\hat{C}_3), \qquad (10.3.10)
$$

that satisfies

$$
\sqrt{NT}(\hat{\boldsymbol{\beta}}^*_{LS} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, W^{-1}\Omega W^{-1}), \qquad (10.3.11)
$$

where $W = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}, \hat{\Omega} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{u}_{it})^2\boldsymbol{x}_{it}\boldsymbol{x}'_{it}$, the K-vectors of $\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$ are

$$
\hat{C}_{1,k} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} K\left(\frac{s-t}{M}\right)[P_{\hat{f}}]_{ts}\hat{u}_{it}X_{k,is}, \qquad (10.3.12)
$$

$$
\hat{C}_{2,k} = \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{u}_{it})^2[M_{\hat{b}}X_k\hat{F}(\hat{F}'\hat{F})^{-1}(\hat{B}'\hat{B})^{-1}\hat{B}']_{ii}, \qquad (10.3.13)
$$

$$
\hat{C}_{3,k} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{u}_{it})^2[M_{\hat{f}}X'_k\hat{B}(\hat{B}'\hat{B})^{-1}(\hat{F}'\hat{F})^{-1}\hat{F}']_{ii}, \qquad (10.3.14)
$$

$\hat{u}_{it} = y_{it} - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{b}}'_i\hat{\boldsymbol{f}}_t, M_A = I - P_A, P_A = A(A'A)^{-1}A, \hat{X}_k = M_{\hat{\lambda}}X_K M_{\hat{f}}, X_K$ is an $N \times T$ matrix of the kth explanatory variable, x_{kit}, and $[A]_{ts}$ denote (t,s)th element of the matrix A, and $K(\nu)$ is a truncation kernel defined by $K(\nu) = 1$ for $|\nu| \leq 1$ and 0 otherwise. W and Ω are the plim of $\hat{W}$ and $\hat{\Omega}$ when $(N,T) \to \infty$.

Similar issues arise for the quasi-difference estimator of Ahn, Lee, and Schmidt (2013). It can similarly show that the quasi-difference estimator is inconsistent when T is fixed no matter how large N is. If both N and T are large, the quasi-difference estimator is consistent, but (Hsiao and Zhang 2015)

$$Asy.E\left[\sqrt{NT}(\hat{\beta}_{\text{QMLE}} - \beta)\right] = O\left(\sqrt{\frac{N}{T}}\right). \tag{10.3.15}$$

10.3.2 Pesaran (2006) CCE

Adding cross-sectional means of y_{it} and $\boldsymbol{x}_{it}$ into model (10.2.1) to approximate the unobserved $\boldsymbol{f}_t$ as suggested by Pesaran (2006) creates correlations between $(\bar{y}_t, \bar{\boldsymbol{x}}_t)$ and v_{it} if $\boldsymbol{x}_{it}$ is not strictly exogenous. For instance, suppose

$$y_{it} = \beta_1 y_{i,t-1} + \boldsymbol{\beta}_2' \boldsymbol{x}_{it} + \boldsymbol{b}_i' \boldsymbol{f}_t + u_{it}, \mid \beta_1 \mid < 1. \tag{10.3.16}$$

By continuous substitution,

$$y_{it} = \boldsymbol{\beta}_2' \sum_{s=0} \boldsymbol{x}_{i,t-s} \beta_1^s + \boldsymbol{b}_i' \left(\sum_{s=0} \boldsymbol{f}_{t-s} \beta_1^s \right) + \sum_{s=0} u_{i,t-s} \beta_1^s. \tag{10.3.17}$$

Then,

$$\begin{aligned}
\text{Cov}\,(\bar{y}_t, v_{it}) = {} & \frac{1}{N} \sum_{j \neq i} \boldsymbol{b}_i' \left(\sum_{s=0} \boldsymbol{f}_{t-s} \beta_1^s \right) \boldsymbol{b}_j' \boldsymbol{f}_t \\
& + \frac{1}{N} \sum_{j \neq i} \boldsymbol{b}_i' \left(\sum_{s=0} \boldsymbol{f}_{t-s} \beta_1^s \right) \boldsymbol{b}_j' \boldsymbol{f}_t \\
& + \frac{\sigma_u^2}{N} \frac{1}{1 - \beta_1} \\
= {} & O(1).
\end{aligned} \tag{10.3.18}$$

Furthermore, if $x_{it} = y_{i,t-1}$, then $\bar{x}_t = \bar{y}_{t-1}$ is a function of f_{t-1}. Thus, adding $\bar{x}_t$ into (10.2.1) introduce lagged f_t into the error term. In other words, the Pesaran CCE (Equation 10.2.22) is not feasible if $\boldsymbol{x}_{it}$ contains lagged dependent variables.

To get around the issue of correlations between $\bar{z}_t = (\bar{y}_t, \bar{\boldsymbol{x}}_t)$ and v_{it}, Chudik and Pesaran (2015) suggest augmenting model (10.2.1) by $\bar{z}_t, \bar{z}_{t-1}, \ldots, \bar{z}_{t-p}$. The rationale is that for models of the type

$$z_{it} = A_i z_{i,t-1} + \boldsymbol{b}_i' \boldsymbol{f}_t + \boldsymbol{e}_{it}^*, \tag{10.3.19}$$

where the innovation term $\boldsymbol{e}_{it}^*$ is independent over i and t. If $\bar{z}_t = \frac{1}{N} \sum_{i=1}^N z_{it}$ can be approximated by $\sum_{s=1}^p \tilde{A}_s \bar{z}_{t-s}$, then $\boldsymbol{f}_t$ can be estimated by

$$\hat{\boldsymbol{f}}_t \simeq (\tilde{B}' \tilde{B})^{-1} \left[\bar{z}_t - \tilde{A}_1 \bar{z}_{t-1} - \cdots - \tilde{A}_p \bar{z}_{t-p} \right],$$

where $\tilde{B} = \frac{1}{N} \sum_{i=1}^N [A_i, \boldsymbol{b}_i']$. However, for model (10.3.19),

$$\bar{z}_t = \frac{1}{N} \left(\sum_{i=1}^N A_i z_{i,t-1} \right) + \left(\frac{1}{N} \sum_{i=1}^N \boldsymbol{b}_i' \right) \boldsymbol{f}_t + \frac{1}{N} \sum_{i=1}^N \boldsymbol{e}_{it}^*, \tag{10.3.20}$$

strong conditions are needed to ensure that approximating $\frac{1}{N}\left(\sum_{i=1}^N A_i z_{i,t-1} \right)$ by $\sum_{s=1}^p \tilde{A}_s \tilde{z}_{t-s}$ is of $o(1)$.

10.3.3 Transformed Estimator

The regressor $\boldsymbol{w}'X_i$ of model (10.2.36) is no longer uncorrelated with $\boldsymbol{w}'\boldsymbol{u}_i$ if $\boldsymbol{x}_{it}$ contains lag dependent variables. Hence, (10.2.37) is no longer consistent. However, if $T > N > r$, transforming (10.2.42) into (10.2.43) has the regressor $\tilde{\boldsymbol{w}}'X_t$ uncorrelated with $\tilde{\boldsymbol{w}}'\boldsymbol{u}_t$ as long as $\boldsymbol{u}_t$ is independently distributed over t, whether $\boldsymbol{x}_{it}$ is strictly exogenous with respect to u_{it} (i.e., $E\boldsymbol{x}_{it}u_{is} = 0$ for all s) or contains lag dependent variables. Hence, the transformed estimator (10.2.44) is consistent and asymptotically normally distributed centered at the true value even when $\boldsymbol{x}_{it}$ contains lag dependent variables.

10.3.4 Monte Carlo Studies

Under the assumption that u_{it} are independently identically distributed over i and t, the limited Monte Carlo studies conducted by Bai (2009a) and Hsiao (2018) show that the Bai (2009a) least squares estimator performs well, as does the Pesaran (2006) CCE if the rank condition for $\bar{C}$ holds for the static model. However, if the regressors contain lag dependent variables, both the least squares and the CCE have significant size distortion. The Moon–Weidner (2017) bias corrected estimater is able to correct the size distortion if both N and T are large, although the bias corrected estimator still has significant size distortion in finite N and T.

When u_{it} are heteroscedastic or weakly cross-sectional dependent or the regressors contain lag dependent variables, Hsiao, Shi, and Zhou (2021) have conducted Monte Carlos on the finite sample performance of the transformed estimator, the average of the transformed estimators, and the Bai (2009a) least squares estimator[4]. The simulation designs for the following three data generating processes are reproduced in Tables 10.1–10.3:

DGP1 (model with two factors):

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \lambda_{1,i}f_{1,t} + \lambda_{2,i}f_{2,t} + u_{it}, \tag{10.3.21}$$

where $\beta_1 = 1$ and $\beta_2 = 2$, and $u_{it} \sim IIDN(0, 4)$. The covariate x_{it} is generated as in Bai (2009),

$$x_{k,it} = 1 + \lambda_{1,i} + \lambda_{2,i} + f_{1,t} + f_{2,t} + \lambda_{1,i}f_{1,t} + \lambda_{2,i}f_{2,t} + \eta_{k,it}, k = 1, 2, \tag{10.3.22}$$

where $\eta_{k,it} \sim IIDN(0, 1)$.

DGP2 (weakly cross-sectionally correlated errors): Hsiao, Shi, and Zhou (2021) assume y_{it} and $x_{k,it}$ are generated as in DGP1, except that now u_{it} is generated as

$$u_{it} = 0.7\, u_{i-1,t} + \varepsilon_{it} \tag{10.3.23}$$

where $\varepsilon_{it} \sim IIDN(0, 2)$. Following Bai (2009a), they discarded the first 100 individuals to set $i = 1, \ldots, N$.

DGP3 (dynamic model with i.i.d. common factors):

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \lambda_{1,i}f_{1,t} + \lambda_{2,i}f_{2,t} + \alpha_i + u_{it}, \tag{10.3.24}$$

where $\rho = 0.5$ and $\beta = 1$, respectively; $\alpha_i \sim IIDN(0, 1)$; $u_{it} \sim IIDN(0, \sigma^2_{ui})$, with σ^2_{ui} being random draws from $(1 + 0.5\ \chi^2(2))$; and x_{it} is generated the same as (10.3.22).

These DGPs assume $\lambda_{i,j}$ and $f_{j,t}$ are all random draws from $IIDN(0, 1)$, for $j = 1$ or 2. The number of replication is set to be 1,000, and $N = 10, 20, 50$ and $T = 100, 200, 500$. The mean of the estimates, bias, root mean square error (RMSE), IQR (inter-quantile range, 75%–25% percentile) and empirical size (computed as the empirical rejection frequency using 5% nominal critical value) are reproduced in Tables 10.1–10.3.

[4] The transformation vector $\boldsymbol{w}$ or $\tilde{\boldsymbol{w}}$ is not unique. For detail, see Hsiao, Shi and Zhou (2021).

Table 10.1. *Simulation results for β_1 and β_2 of DGP 1*

β_1		Transformed Estimator				Average Estimator1				Average Estimator2				PCA1				PCA2			
N	T	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size
	100	0.0596	0.2528	0.3222	5.4%	0.0596	0.2528	0.3222	5.4%	0.0207	0.109	0.1371	6.1%	0.0604	0.2524	0.3211	5.5%	0.133	0.1461	0.0767	59.3%
10	200	0.0234	0.1779	0.2344	6.3%	0.0234	0.1779	0.2344	6.3%	0.0061	0.0693	0.0877	5.6%	0.0241	0.1774	0.234	6.4%	0.1298	0.1377	0.0586	81.1%
	500	0.004	0.1005	0.1253	5.3%	0.004	0.1005	0.1253	5.3%	0.0004	0.0408	0.0539	4.9%	0.004	0.1004	0.1253	5.3%	0.1316	0.1365	0.0491	96.4%
	100	0.0462	0.2755	0.3622	5.4%	0.012	0.0521	0.063	5.9%	0.0066	0.0782	0.1033	5.4%	0.0328	0.0746	0.0801	6.5%	0.1299	0.1365	0.0586	86.3%
20	200	0.0281	0.1811	0.2278	5.8%	0.0088	0.0368	0.0497	5.5%	0.0014	0.0493	0.0668	5.5%	0.0221	0.0535	0.0573	8.0%	0.1296	0.1335	0.0435	97.9%
	500	0.0055	0.0966	0.1222	5.6%	0.0022	0.0224	0.0292	5.5%	−0.0004	0.0304	0.041	4.5%	0.0085	0.0371	0.0274	5.6%	0.1303	0.1324	0.0332	100.0%
	100	0.0764	0.222	0.2779	6.8%	0.0037	0.0291	0.0385	5.0%	0.004	0.0496	0.0683	5.5%	0.0171	0.0509	0.0404	8.0%	0.1325	0.1352	0.0374	99.9%
50	200	0.0566	0.2233	0.2963	6.0%	0.002	0.0209	0.0295	4.8%	−0.0001	0.0312	0.0419	5.3%	0.0132	0.04	0.0282	10.4%	0.1313	0.1329	0.0283	100.0%
	500	0.0177	0.1132	0.1439	6.2%	0.0006	0.0125	0.0176	5.1%	0.0003	0.018	0.0248	5.0%	0.0013	0.0134	0.0142	2.3%	0.1324	0.1332	0.0203	100.0%
	100	0.0882	0.1903	0.2361	6.8%	0.0009	0.0202	0.0278	4.7%	0.0006	0.0292	0.0389	4.4%	0.0067	0.031	0.0251	5.9%	0.1325	0.1341	0.0281	100.0%
100	200	0.0622	0.1547	0.1861	7.9%	0.0009	0.0138	0.0184	5.0%	0.0006	0.0179	0.0243	5.1%	0.0024	0.0165	0.0166	2.2%	0.1332	0.134	0.0202	100.0%
	500	0.0194	0.1413	0.1825	5.6%	0.0003	0.0085	0.0114	4.8%	0.0003	0.0105	0.0145	5.2%	0.0008	0.0108	0.0101	0.7%	0.132	0.1324	0.0143	100.0%
β_2																					
	100	0.0465	0.2425	0.3069	5.9%	0.0465	0.2425	0.3069	5.9%	0.0209	0.1076	0.141	5.6%	0.0474	0.2416	0.3058	5.9%	0.1317	0.1457	0.0862	52.4%
10	200	0.0228	0.1807	0.2205	6.5%	0.0228	0.1807	0.2205	6.5%	0.0082	0.0733	0.0912	5.9%	0.0234	0.1803	0.2208	6.5%	0.1304	0.1382	0.0605	80.7%
	500	0.0067	0.096	0.1317	4.5%	0.0067	0.096	0.1317	4.5%	0.0021	0.0404	0.056	5.3%	0.0067	0.096	0.1317	4.5%	0.133	0.1376	0.044	97.6%
	100	0.0493	0.2743	0.3628	5.5%	0.0146	0.0509	0.0658	6.6%	0.0034	0.0789	0.1049	5.3%	0.0354	0.075	0.078	7.7%	0.1312	0.1377	0.0559	87.3%
20	200	0.0285	0.183	0.2377	4.5%	0.009	0.0373	0.0475	6.3%	0.0023	0.0505	0.0668	5.2%	0.0225	0.0564	0.0575	8.6%	0.1303	0.134	0.0411	99.0%
	500	0.0044	0.0997	0.1324	5.3%	0.003	0.0227	0.0297	5.1%	0.0007	0.0308	0.0416	4.9%	0.0091	0.0368	0.0266	5.9%	0.1309	0.133	0.0306	99.9%
	100	0.0722	0.2161	0.2574	6.6%	0.0027	0.0286	0.0384	5.1%	0.0031	0.0495	0.0674	4.6%	0.0164	0.0507	0.0426	7.5%	0.1314	0.1343	0.0366	99.8%
50	200	0.0486	0.2235	0.2934	6.2%	0.0014	0.0206	0.0277	5.2%	−0.0007	0.0328	0.0441	5.5%	0.0128	0.0394	0.0276	10.3%	0.1311	0.1327	0.0295	100.0%
	500	0.0202	0.1167	0.1525	5.9%	0.0012	0.013	0.0183	4.3%	0.0005	0.0182	0.0239	5.1%	0.0018	0.0139	0.0147	1.9%	0.1325	0.1334	0.0204	100.0%
	100	0.0736	0.184	0.2178	6.8%	0.0008	0.0195	0.0264	5.2%	0.0006	0.0291	0.0368	5.9%	0.0067	0.0298	0.0225	5.8%	0.1325	0.1339	0.0269	100.0%
100	200	0.0665	0.1592	0.201	7.2%	−0.0001	0.0151	0.0211	5.2%	−0.0002	0.0188	0.0253	5.4%	0.0014	0.0176	0.0167	1.9%	0.1322	0.133	0.02	100.0%
	500	0.0241	0.1381	0.1714	5.9%	0.0005	0.0089	0.0121	5.3%	0.0007	0.011	0.0153	4.1%	0.0009	0.0108	0.0103	1.0%	0.1324	0.1329	0.0145	100.0%

Table 10.2. *Simulation results for β_1 and β_2 of DGP 2*

β_1		Transformed Estimator				Average Estimator1				Average Estimator2				PCA1				PCA2			
N	T	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size
	100	0.0456	0.2466	0.31	6.1%	0.0456	0.2466	0.31	6.1%	0.0199	0.112	0.1433	4.7%	0.0458	0.2462	0.31	6.2%	0.1331	0.1458	0.0766	60.4%
10	200	0.0126	0.1735	0.2336	4.7%	0.0126	0.1735	0.2336	4.7%	0.0022	0.0686	0.0897	5.1%	0.0128	0.1734	0.233	4.7%	0.1314	0.1389	0.061	82.7%
	500	0.0031	0.0958	0.1224	4.8%	0.0031	0.0958	0.1224	4.8%	0.0015	0.042	0.0587	4.6%	0.0036	0.0959	0.122	5.0%	0.1315	0.1359	0.0454	97.3%
	100	0.0509	0.273	0.3515	5.7%	0.0115	0.0538	0.0705	6.1%	0.0055	0.0826	0.1076	5.8%	0.0362	0.079	0.0849	5.9%	0.1289	0.1354	0.057	87.0%
20	200	0.0297	0.18	0.2329	5.6%	0.0082	0.0367	0.0479	5.6%	0.0007	0.0515	0.0683	4.8%	0.0245	0.061	0.064	6.0%	0.129	0.1328	0.0416	98.8%
	500	0.0051	0.1041	0.1363	4.5%	0.0028	0.0227	0.0294	5.5%	0.001	0.0299	0.0402	5.6%	0.0084	0.037	0.0272	5.6%	0.1308	0.1329	0.0281	100.0%
	100	0.0865	0.2251	0.2852	7.5%	0.0048	0.0305	0.0387	4.9%	0.0022	0.0512	0.0644	6.1%	0.0204	0.0517	0.0463	10.7%	0.1323	0.1351	0.035	99.9%
50	200	0.0495	0.2185	0.2935	5.7%	0.0008	0.0199	0.0261	4.8%	−0.0013	0.0317	0.0432	5.3%	0.0122	0.0394	0.0282	9.4%	0.1321	0.1337	0.0277	100.0%
	500	0.015	0.1111	0.1562	5.4%	0.0008	0.013	0.0172	5.3%	0.0012	0.0182	0.023	5.8%	0.0015	0.0164	0.0151	0.9%	0.1319	0.1328	0.0216	100.0%
	100	0.0726	0.189	0.2424	6.6%	0.0015	0.0194	0.0256	4.7%	0.0016	0.0284	0.0393	4.8%	0.0063	0.0313	0.0232	4.8%	0.1315	0.1329	0.0255	100.0%
100	200	0.0709	0.1632	0.2064	7.5%	0.0007	0.0146	0.0195	5.2%	0.0006	0.0188	0.0248	4.7%	0.003	0.0198	0.0174	2.2%	0.1332	0.1341	0.0215	100.0%
	500	0.0243	0.1394	0.1853	5.0%	0.0002	0.0084	0.0115	4.8%	0.0001	0.0105	0.0146	5.2%	0.0009	0.0115	0.0104	0.6%	0.1323	0.1327	0.0151	100.0%
β_2																					
	100	0.0401	0.2501	0.3229	5.1%	0.0401	0.2501	0.3229	5.1%	0.0141	0.1081	0.1345	4.6%	0.0402	0.2498	0.3229	5.2%	0.1294	0.1427	0.0777	58.6%
10	200	0.0264	0.177	0.2333	5.1%	0.0264	0.177	0.2333	5.1%	0.0057	0.0745	0.0931	5.4%	0.0268	0.1766	0.232	5.1%	0.1326	0.141	0.0632	79.1%
	500	0.0058	0.1007	0.1273	5.4%	0.0058	0.1007	0.1273	5.4%	0.0015	0.0413	0.054	5.0%	0.0059	0.1002	0.1273	5.9%	0.1315	0.1364	0.0468	95.9%
	100	0.0526	0.2717	0.3574	5.4%	0.0132	0.0533	0.0703	5.9%	0.0069	0.0842	0.1118	4.4%	0.038	0.0783	0.082	7.6%	0.1308	0.1381	0.0601	83.7%
20	200	0.0438	0.1887	0.2351	6.6%	0.011	0.0394	0.0522	6.5%	0.0048	0.0513	0.0691	4.4%	0.0272	0.0646	0.0664	5.9%	0.1312	0.1348	0.0439	99.4%
	500	0.0061	0.1026	0.136	5.0%	0.0022	0.0218	0.0298	6.2%	−0.0003	0.0305	0.0414	4.1%	0.0078	0.0352	0.0272	4.7%	0.1293	0.1313	0.0324	100.0%
	100	0.0805	0.2174	0.254	7.8%	0.0037	0.0294	0.0416	4.7%	0.0025	0.0517	0.0703	4.4%	0.0195	0.0499	0.044	10.5%	0.1321	0.1349	0.0386	100.0%
50	200	0.0503	0.2199	0.2856	6.3%	0.0009	0.0212	0.0278	6.0%	−0.0017	0.0315	0.043	5.6%	0.0122	0.0413	0.0292	9.2%	0.1316	0.1332	0.0271	100.0%
	500	0.0156	0.1197	0.1595	5.3%	0.0001	0.0124	0.0167	4.5%	−0.0004	0.0178	0.0237	4.9%	0.0009	0.0158	0.0144	1.0%	0.1314	0.1322	0.0203	100.0%
	100	0.0598	0.1821	0.2285	5.9%	0.0008	0.0193	0.0259	4.4%	0.0012	0.0275	0.0358	4.9%	0.0057	0.0314	0.0242	4.8%	0.132	0.1334	0.0255	100.0%
100	200	0.0702	0.164	0.2194	7.3%	0.0006	0.0142	0.0192	4.4%	0.0008	0.018	0.024	4.7%	0.0029	0.0198	0.0174	2.3%	0.1327	0.1335	0.0201	100.0%
	500	0.0164	0.1396	0.1777	4.6%	0.0003	0.0084	0.0119	4.9%	0.0008	0.0098	0.0139	5.1%	0.0011	0.0113	0.01	0.9%	0.1326	0.133	0.0144	100.0%

Table 10.3. *Simulation results for ρ and β of DGP 4*

ρ		Transformed Estimator				Average Estimator1				Average Estimator2				PCA1				PCA2			
N	T	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size	Bias	RMSE	IQR	Size
10	100	0.0119	0.0703	0.0841	6.2%	0.0119	0.0703	0.0841	6.2%	−0.0037	0.0249	0.0309	4.4%	0.0077	0.0652	0.0797	5.9%	0.1854	0.1895	0.0503	99.4%
	200	0.0166	0.0632	0.0697	7.8%	0.0166	0.0632	0.0697	7.8%	−0.0012	0.0183	0.0236	4.9%	0.006	0.0457	0.0473	4.8%	0.1855	0.1896	0.0535	99.1%
	500	0.0157	0.0524	0.0577	9.0%	0.0157	0.0524	0.0577	9.0%	0.0003	0.0117	0.0153	4.7%	0.0022	0.0266	0.0271	2.6%	0.1859	0.1898	0.0525	99.5%
20	100	−0.0027	0.0492	0.0617	4.2%	0.0119	0.0165	0.0157	18.2%	−0.0065	0.0172	0.0207	7.2%	0.1902	0.1922	0.0368	99.9%	0.1904	0.1923	0.0366	100.0%
	200	−0.0026	0.0351	0.0457	4.5%	0.014	0.0162	0.011	41.2%	−0.0036	0.0114	0.0147	6.2%	0.1893	0.1911	0.0341	100.0%	0.1893	0.1911	0.0341	100.0%
	500	−0.0002	0.0234	0.0296	4.1%	0.0162	0.017	0.0072	87.9%	−0.001	0.0071	0.0098	4.2%	0.1916	0.1933	0.0343	100.0%	0.1916	0.1933	0.0343	100.0%
50	100	−0.0038	0.048	0.0674	4.8%	0.0002	0.007	0.01	4.8%	−0.008	0.0126	0.0131	13.4%	0.1848	0.1856	0.0221	100.0%	0.1848	0.1856	0.0221	100.0%
	200	−0.0038	0.0327	0.0461	5.0%	0.0034	0.0057	0.0067	11.2%	−0.0039	0.007	0.0079	11.1%	0.1853	0.186	0.0225	100.0%	0.1853	0.186	0.0225	100.0%
	500	−0.0013	0.0204	0.0282	4.9%	0.0056	0.0064	0.0043	42.6%	−0.0012	0.004	0.0054	5.3%	0.1855	0.1862	0.0226	100.0%	0.1855	0.1862	0.0226	100.0%
100	100	−0.01	0.0529	0.07	6.5%	−0.0048	0.0075	0.0078	13.3%	−0.0106	0.0143	0.0125	19.7%	0.1903	0.1907	0.0162	100.0%	0.1903	0.1907	0.0162	100.0%
	200	−0.0031	0.0354	0.0478	4.2%	−0.0006	0.0037	0.0051	4.9%	−0.0045	0.0065	0.0062	14.8%	0.1896	0.1899	0.0146	100.0%	0.1896	0.1899	0.0146	100.0%
	500	−0.0009	0.0208	0.0289	4.8%	0.0017	0.0029	0.0032	12.5%	−0.0017	0.0031	0.0036	9.0%	0.19	0.1903	0.0141	100.0%	0.19	0.1903	0.0141	100.0%
β																					
10	100	0.0077	0.0942	0.1229	4.3%	0.0077	0.0942	0.1229	4.3%	0.0072	0.0403	0.0548	5.0%	0.0153	0.0881	0.1114	6.1%	−0.1363	0.1492	0.0844	61.6%
	200	0.0006	0.0709	0.0936	4.8%	0.0006	0.0709	0.0936	4.8%	0.004	0.0288	0.0387	5.3%	0.0041	0.0557	0.0736	5.6%	−0.1379	0.1481	0.0787	71.4%
	500	0.0044	0.0542	0.0718	5.1%	0.0044	0.0542	0.0718	5.1%	0.0016	0.0189	0.0268	4.9%	0.0016	0.0337	0.0449	5.0%	−0.1406	0.1499	0.074	76.0%
20	100	0.0118	0.0897	0.1178	5.1%	0.0071	0.0211	0.0276	6.8%	0.0069	0.0279	0.0363	5.5%	−0.1436	0.1498	0.0561	91.9%	−0.1437	0.1498	0.0559	92.2%
	200	0.0062	0.0625	0.0833	5.7%	0.0029	0.0152	0.0192	5.4%	0.0034	0.0198	0.0246	5.7%	−0.1429	0.1479	0.0557	95.9%	−0.1429	0.1479	0.0557	95.9%
	500	0.002	0.0407	0.0586	4.7%	0.0006	0.0098	0.0128	4.6%	0.0009	0.0129	0.0178	4.9%	−0.1469	0.151	0.0464	98.6%	−0.1469	0.151	0.0464	98.6%
50	100	0.0088	0.0849	0.1147	5.1%	0.0083	0.0149	0.0163	10.2%	0.0078	0.0171	0.0197	8.3%	−0.1203	0.1228	0.0317	99.6%	−0.1203	0.1228	0.0317	99.6%
	200	0.005	0.0594	0.0801	5.0%	0.0038	0.0093	0.0115	7.5%	0.0034	0.0111	0.0142	5.7%	−0.1216	0.1235	0.0293	100.0%	−0.1216	0.1235	0.0293	100.0%
	500	0.0013	0.037	0.0493	4.8%	0.0019	0.0057	0.0072	6.3%	0.0011	0.0066	0.0088	5.8%	−0.1219	0.1236	0.0288	100.0%	−0.1219	0.1236	0.0288	100.0%
100	100	0.0085	0.084	0.1166	4.6%	0.0073	0.0119	0.0124	12.8%	0.0075	0.0147	0.0162	9.8%	−0.1187	0.1201	0.0242	100.0%	−0.1187	0.1201	0.0242	100.0%
	200	0.0054	0.0591	0.0838	4.1%	0.0041	0.0076	0.0086	10.9%	0.0037	0.0084	0.0103	8.6%	−0.118	0.119	0.0204	100.0%	−0.118	0.119	0.0204	100.0%
	500	0.0001	0.0376	0.0507	4.9%	0.0023	0.0047	0.0055	8.7%	0.0016	0.0048	0.0063	5.9%	−0.1182	0.1192	0.0206	100.0%	−0.1182	0.1192	0.0206	100.0%

In general: (i) The transformed estimator performs well for finite or large N. The bias is almost negligible and the empirical size is quite close to the nominal size, whether the errors are heteroscedastic or cross-correlated or whether x contains lag dependent variables. (ii) Taking the average of $(N - \hat{r})$ orthogonal transformed estimates reduces the IQR. However, its actual size in some cases is not as close to the nominal size as the transformed estimator if the unknown r is determined through the Bai and Ng information criterion. On the other hand, using the maximum $p > r$, the actual size of the average estimator is close to the nominal size. (iii) The Bai (2009) least squares estimator performes less well when N is small. However, its performance improves as (N, T) become large. The actual size could be very different from the nominal size, probably due to the presence of heteroscedasticity in the idiosyncratic errors. (iv) When a lag dependent variable appears, there are also significant bias and size distortion, which is expected as shown by Moon and Weidner (2017) that the least squares estimator is consistent but asymptotically biased. (iv) Among the two Bai (2009a) recursive iterative estimation procedures to obtain the slope coefficients, PCA1 that iterates between (10.2.13) and (10.2.14) performs better than PCA2 that iterates between (10.2.14) and

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} X_i' X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' M_F \boldsymbol{y}_i \right). \tag{10.3.25}$$

10.4 MODELS WITH MIXED FIXED AND RANDOM EFFECTS

When $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are treated as random variables, specific assumptions for their data generation process need to be assumed. For instance, Sarafidis, Yamagata, and Robertson (2009) assume that $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ are independently distributed over i and t, respectively. Under the assumption that $\boldsymbol{b}_i$ are independently distributed over i with constant covariance matrix Σ_b, and $\boldsymbol{f}_t$ are independently distributed over t with constant covariance matrix Σ_f and are uncorrelated with x_{it} then

$$\begin{aligned} \mathrm{Cov}(v_{it}, v_{js}) &= \sigma_u^2 + tr(\Sigma_f \Sigma_b), && \text{if } i = j \text{ and } t = s, \\ &= 0, && \text{otherwise.} \end{aligned} \tag{10.4.1}$$

In other words, there is neither cross-sectional dependence nor time dependence for the error term. However, σ_u^2 and $tr(\Sigma_f \Sigma_b)$ cannot be separately identifiable from the constant variance $\sigma_v^2 = \sigma_u^2 + tr(\Sigma_f \Sigma_b)$. Thus, unless there is a strong prior information about the data generating process of $\boldsymbol{b}_i, \boldsymbol{f}_t$, or u_{it}, it might be simpler to ignore the component structure of v_{it} and just consider obtaining efficient estimates of $\boldsymbol{\beta}$ from the relations between x_{it} and v_{it}. Therefore, we shall consider only the case that either $\boldsymbol{b}_i$ or $\boldsymbol{f}_t$ are treated as random, and the other components are treated fixed; e.g., Sarafidis and Wansbeek (2012), Sarafidis, Yamagata, and Robertson (2009), etc., consider treating $\boldsymbol{b}_i$ random and $\boldsymbol{f}_t$ fixed.

10.4.1 Fixed Time Effects and Random Individual Effects

Suppose $\boldsymbol{f}_t$ are fixed and $\boldsymbol{b}_i$ are independently distributed over i with $E(\boldsymbol{b}_i) = 0$ and covariance matrix Σ_b, v_{it} is independently distributed over i with

$$E v_{it} = \boldsymbol{f}_t' E(\boldsymbol{b}_i) + E u_{it} = 0, \tag{10.4.2}$$

$$E(v_{it}v_{jt}) = \boldsymbol{f}_t' E(\boldsymbol{b}_i \boldsymbol{b}_j') \boldsymbol{f}_t + E(u_{it}u_{jt})$$
$$\begin{cases} = 0, & \text{if } i \neq j, \\ = \boldsymbol{f}_t' \Sigma_b \boldsymbol{f}_t + \sigma_u^2, & \text{if } i = j. \end{cases} \tag{10.4.3}$$

However, for given i,

$$E(v_{it}v_{is}) = \boldsymbol{f}_t' \Sigma_b \boldsymbol{f}_s \tag{10.4.4}$$

can be different from zero. In other words, allowing $\boldsymbol{f}_t$ to be fixed constants not only can allow the common factor $\boldsymbol{f}_t$ to be correlated with $\boldsymbol{x}_{it}$ but also allow serial correlation in v_{it}.

When $\boldsymbol{x}_{it}$ are strictly exogenous with respect to u_{it}, $E(\boldsymbol{x}_{it}u_{is}) = 0$ and are independent of $\boldsymbol{b}_i$, then

$$\begin{aligned} \prod_{i=1}^N f(\boldsymbol{y}_i \mid x_i, F) &= \prod_{i=1}^N \int f(\boldsymbol{y}_i \mid \boldsymbol{x}_i, F, \boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i \\ &= \prod_{i=1}^N (2\pi)^{-\frac{T}{2}} \mid V \mid^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{y}_i - X_i \boldsymbol{\beta})' V^{-1} (\boldsymbol{y}_i - X_i \boldsymbol{\beta}) \right\}, \end{aligned} \tag{10.4.5}$$

where

$$V = \sigma_u^2 I_T + F \Sigma_b F'. \tag{10.4.6}$$

One can obtain the QMLE of $\boldsymbol{\beta}, F, \Sigma_b$, and σ_u^2 by maximizing the logarithmic transformation of the quasi-likelihood function (10.4.5). However, just like the fixed-effects case, F and Σ_b cannot be uniquely determined. To obtain the unique decomposition of V, we follow Anderson and Rubin (1956), Bai (2009), etc., to impose the normalization conditions that $\frac{1}{T} F'F = I_r$ and Σ_b is diagonal.

The QMLE of $\boldsymbol{\beta}$ conditional on F, Σ_b, and σ_u^2 is equal to

$$\hat{\boldsymbol{\beta}}_{\text{QMLE}} = \left\{ \sum_{i=1}^N X_i' V^{-1} X_i \right\}^{-1} \left\{ \sum_{i=1}^N X_i' V^{-1} \boldsymbol{y}_i \right\}. \tag{10.4.7}$$

Substituting (10.4.7) into the log-likelihood function of (10.4.5) yields the concentrated log-likelihood function

$$-\frac{N}{2} \log \mid V \mid - tr V^{-1} S, \tag{10.4.8}$$

where

$$S = \frac{1}{N} \sum_{i=1}^N (\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_{\text{QMLE}})(\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_{\text{QMLE}})'. \tag{10.4.9}$$

Thus, conditional on $\boldsymbol{\beta}$ and σ_u^2, the QMLE of F is

$$\hat{F}_{\text{QMLE}} = Q(M - \sigma_u^2 I_r)^{\frac{1}{2}}, \tag{10.4.10}$$

where M is a $r \times r$ diagonal matrix containing the r largest eigenvalues of S, and Q is the $T \times r$ matrix of corresponding orthonormal eigenvectors.

Conditional on $\boldsymbol{\beta}$ and F,

$$\hat{\sigma}_u^2 = \frac{1}{T}\, tr\ (S - FF'). \tag{10.4.11}$$

Thus, the QMLE can be obtained by iterating between (10.4.7), (10.4.10), and (10.4.11) until the solution converges.

The QMLE (10.4.7) is consistent and asymptotically normally distributed as $N \longrightarrow \infty$ whether T is fixed or also goes to infinity with $\frac{T}{N} \longrightarrow a \neq 0 < 1$ as $N \longrightarrow \infty$,

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{\text{QMLE}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \tilde{D} + \tilde{D}C\tilde{D}), \tag{10.4.12}$$

where

$$\tilde{D} = \sigma_u^2 \left\{ \frac{1}{NT} \sum_{i=1}^{N} X_i' V^{-1} X_i \right\}^{-1} \tag{10.4.13}$$

and

$$C = \frac{1}{NT} \sum_{i=1}^{N} X_i' V^{-1} \left(\boldsymbol{e}'\ \text{Cov}\ (\hat{\boldsymbol{\theta}}) \boldsymbol{e} \otimes \left(\sum_{\ell=1}^{q} \frac{\partial}{\partial \theta_\ell} V \right)^2 \right) V^{-1} X_i = O\left(\sqrt{\frac{T}{N}} \right), \tag{10.4.14}$$

where $\boldsymbol{\theta}$ is a vector consisting of σ_u^2 and $\text{Vec}(F)$, and q denotes the dimension of $\boldsymbol{\theta}$. The covariance matrix of $\hat{\boldsymbol{\theta}}$ and matrix differentiation are given by Hayashi and Sen (1998), Jennrich and Thayer (1973), and Lawley (1961).

The second term of the asymptotic covariance matrix is of $O\left(\sqrt{\frac{T}{N}}\right)$; thus, if T is fixed and N is large, the asymptotic covariance matrix of QMLE converges to $\tilde{D}$, which is fairly straightforward to compute. However, if $\frac{T}{N} \longrightarrow a \neq 0$ as $N \longrightarrow \infty$, the computation of the asymptotic covariance of the QMLE is very tedious. One way to derive the approximate asymptotic covariance matrix of QMLE is to follow the suggestion of Clarkson (1979) using the jackknife procedure. For detail, see Hsiao (2018).

Remark 10.4.1 When N is large and T is fixed and $\boldsymbol{b}_i$ are treated random with $\boldsymbol{f}_t$ fixed, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{\text{QMLE}}$ is given by (10.4.13). Noting that

$$\sigma_u^2 \left\{ \sum_{i=1}^{N} X_i'(I_T - F(F'F)^{-1}F')X_i \right\} \leq \text{Var}\ (\hat{\boldsymbol{\beta}}_{\text{QMLE}})^{-1}. \tag{10.4.15}$$

The inequality follows from

$$V^{-1} = \sigma_u^2 (F'F)^{-1} - \sigma_u^2 (F'F)^{-1} [\sigma^2 \Sigma_b + (F'F)^{-1}]^{-1} (F'F)^{-1}.$$

Therefore, the inverse of the covariance matrix of (10.2.10) subtract $\text{Var}\ (\hat{\boldsymbol{\beta}}_{\text{QMLE}})^{-1}$ is a nonnegative matrix. In other words, treating $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ both as fixed yields an estimator of $\boldsymbol{\beta}$ that is not as efficient as treating $\boldsymbol{b}_i$ as random and $\boldsymbol{f}_t$ as fixed if $\boldsymbol{b}_i$ are indeed independent of $\boldsymbol{x}_{it}$. The difference between treating both $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ fixed and treating $\boldsymbol{f}_t$ fixed and $\boldsymbol{b}_i$ random could be viewed as the difference between conditional and unconditional inference.

Remark 10.4.2 If F and σ_u^2 are known, (10.4.7) is just the generalized least squares estimator. However, if $\frac{T}{N} \longrightarrow 0$ as $N \longrightarrow \infty$, one can ignore the restrictions on V in (10.4.6) and obtain a $\frac{1}{\sqrt{N}}$ consistent estimate of V using the simple formula $\hat{V} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i'$. Thus, if $\frac{T}{N} \longrightarrow 0$ as $(N, T) \longrightarrow \infty$, the QMLE of $\boldsymbol{\beta}$ can be approximated by the usual two-step procedure. In the first step we obtain estimates of $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iT})'$ by

$$\hat{\boldsymbol{v}}_i = \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}, i = 1, \ldots, N, \tag{10.4.16}$$

from any $\sqrt{N}$ consistent estimator of $\boldsymbol{\beta}$, and estimate V by

$$\hat{V} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i'. \tag{10.4.17}$$

The second step is to use $\hat{V} = (\hat{\sigma}_{st})$ in lieu of $V = \sigma_u^2 I_T + F F'$ in

$$\tilde{\boldsymbol{\beta}}_{\text{FGLS}} = \left(\sum_{i=1}^{N} X_i' \hat{V}^{-1} X_i \right)^{-1} \left\{ \sum_{i=1}^{N} X_i' \hat{V}^{-1} \boldsymbol{y}_i \right\}. \tag{10.4.18}$$

It can be shown that if $\frac{T}{N} \longrightarrow 0$ as $(N, T) \longrightarrow \infty$ that (Hsiao 2018)

$$\sqrt{NT} \left(\tilde{\boldsymbol{\beta}}_{\text{FGLS}} - \hat{\boldsymbol{\beta}}_{\text{QMLE}} \right) = o(1). \tag{10.4.19}$$

The limited Monte Carlo conducted by Hsiao (2018) shows that the feasible GLS that ignores the restrictions on V performs equally well in terms of the root mean square error and the size of the test if N is large.

Remark 10.4.3 When $\boldsymbol{x}_{it}$ contains a lag dependent variable, (10.4.8) is not the proper likelihood function. Furthermore, $E(v_{it} \mid \boldsymbol{x}_{it}) \neq 0$ because $\boldsymbol{f}_t$ appear in both v_{it} and $\boldsymbol{x}_{it}$. Neither (10.4.7) nor (10.4.18) is consistent. Moreover, (10.4.3) shows that all $(y_{i1}, \ldots, y_{iT})$ are correlated due to nonzero $\boldsymbol{f}_t \boldsymbol{f}_s'$. In other words, no lag dependent variables can be legitimate instruments through the conventional first difference method.

The QMLE is consistent and asymptotically unbiased only if initial value distribution is properly taken into account. If initial value, y_{i0}, is treated fixed, just like the additive effects models (Chapter 3), the QMLE is inconsistent if T is fixed. It is consistent if $T \to \infty$, but it is asymptotically biased of order $O(\sqrt{a})$ where $a = \frac{N}{T}$ as $T \to \infty$ (Hsiao 2018).

The Ahn and Schmidt (1995) and Ahn, Lee, and Schmidt (2001, 2013) quasi-differencing method can be used to eliminate the impact of $\boldsymbol{b}_i' \boldsymbol{f}_t$ from (10.2.1). If T is fixed, the estimator is biased of order $\frac{1}{T}$. If $T \longrightarrow \infty$, it is consistent, but it is asymptotically biased of order $\sqrt{\frac{N}{T}}$ (Hsiao and Zhang 2015). Robertson and Sarafidis (2015) consider a GMM approach of estimating (10.2.1) assuming there do exist instruments z_{it} such that $E(z_{it} \boldsymbol{b}_i') = \sum_{zb} = \mathbf{0}$. Apart from the issue of how to find appropriate z_{it} empirically, there is also an identification issue if the instruments are weak (e.g., Ahn 2015).

10.4.2 Fixed Individual Effects and Random Time Effects

Treating $\boldsymbol{b}_i$ as fixed and $\boldsymbol{f}_t$ as random leads to essentially the same estimation procedure as in the case of $\boldsymbol{b}_i$ as random and $\boldsymbol{f}_t$ as fixed, except that models (10.2.1) and (10.2.2) are now arranged in the form

$$\boldsymbol{y}_t = X_t\boldsymbol{\beta} + B\boldsymbol{f}_t + \boldsymbol{u}_t, \quad t = 1, \ldots, T, \tag{10.4.20}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $X_t = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$, and $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})'$. However, now

$$E(v_{it}v_{jt}) = \boldsymbol{b}_i' E\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{b}_j \neq 0, \tag{10.4.21}$$

implies that the errors are cross-sectionally dependent. In other words, the error terms are cross-correlated. When T is fixed, as discussed in Section 10.1, the least squares estimator is inconsistent. Moreover, when the errors are cross-correlated, the most efficient estimator is the generalized least squares estimator. However,

$$E\boldsymbol{v}_t\boldsymbol{v}_t' = B\Sigma_f B' + \sigma_u^2 I_N = \Sigma_v \tag{10.4.22}$$

cannot be uniquely decomposed where $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)'$. To unique decompose (10.4.22) we impose the normalization condition $\Sigma_f = I_r$ and $\frac{1}{N}B'B = D$, a nonsingular diagonal matrix.

If $\boldsymbol{f}_t$ is independent normal $N(0, \Sigma_f)$, under the assumption that u_{it} is independent normal over i and t, the likelihood function of $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ takes the form

$$\prod_{t=1}^{T} (2\pi)^{-\frac{N}{2}} \mid \Sigma_v \mid^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{y}_t - X_t\boldsymbol{\beta})'\Sigma_v^{-1}(\boldsymbol{y}_t - X_t\boldsymbol{\beta})'\right\}. \tag{10.4.23}$$

Then conditional on Σ_v, the QMLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \left\{\sum_{t=1}^{T} X_t'(\sigma_u^2 I_N + BB')^{-1}X_t\right\}^{-1} \left\{\sum_{t=1}^{T} X_t'(\sigma_u^2 I_N + BB')^{-1}\boldsymbol{y}_t\right\}. \tag{10.4.24}$$

The derivation of the QMLE requires the simultaneous satisfaction of the first-order conditions of (10.4.23) with respect to $\boldsymbol{\beta}$, B, and σ_u^2. Similar to the fixed $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ case we can iterate between (10.4.24) and $\hat{B}$ until the solution converges where $\hat{B}$ are the r eigenvectors corresponding to the r largest eigenvalues of

$$\hat{B}_{\text{QMLE}} = \tilde{Q}(\tilde{M} - \sigma_u^2 I_r)^{\frac{1}{2}} \tag{10.4.25}$$

and

$$\hat{\sigma}_u^2 = \frac{1}{N} tr(\tilde{S} - BB'), \tag{10.4.26}$$

where $\tilde{M}$ is an $r \times r$ diagonal matrix containing the largest eigenvalues of

$$\tilde{S} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{y}_t - X_t\hat{\boldsymbol{\beta}})(\boldsymbol{y}_t - X_t\hat{\boldsymbol{\beta}})', \tag{10.4.27}$$

and $\tilde{Q}$ is the $N \times r$ matrix of corresponding orthonormal eigenvectors.

When $\frac{N}{T} \to 0$ as $(N, T) \to \infty$, Hsiao (2018) suggests using the two-step procedure feasible generalized least square estimator by first estimating Σ_v by $\hat{\Sigma}_v = (\hat{\sigma}_{ij})$,

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_{it}\hat{v}_{jt}. \tag{10.4.28}$$

Remark 10.4.4 Treating $\boldsymbol{f}_t$ as random and independently distributed over t and $\boldsymbol{b}_i$ fixed, contrary to the additive-effects model (Chapter 3), there is no need to separate $\boldsymbol{x}_{it}$ as strictly exogenous with regard to u_{it} or predetermined. The QMLE is consistent and asymptotically unbiased whether N is fixed or large. Moreover, it is more efficient than the estimator treating both $\boldsymbol{b}_i$ and $\boldsymbol{f}_t$ fixed (Hsiao 2018).

Remark 10.4.5 The interactive effects $\boldsymbol{b}_i'\boldsymbol{f}_t$ are unobserved. The statistical inference procedure for treating $\boldsymbol{f}_t$ fixed and $\boldsymbol{b}_i$ random is analogous to treating $\boldsymbol{b}_i$ fixed and $\boldsymbol{f}_t$ random (for detail, see Hsiao 2018). However, there are important differences between the two. First, treating $\boldsymbol{b}_i$ as fixed and $\boldsymbol{f}_t$ as random allows the individual-specific effects to be correlated with $\boldsymbol{x}_{it}$ but does not allow the time effects to be correlated with $\boldsymbol{x}_{it}$; while treating $\boldsymbol{f}_t$ as fixed and $\boldsymbol{b}_i$ as random allows $\boldsymbol{f}_t$ to be correlated with $\boldsymbol{x}_{it}$, but not $\boldsymbol{b}_i$. Second, there is an important difference in modeling the theoretical properties of the random error term. Treating $\boldsymbol{b}_i$ as fixed but $\boldsymbol{f}_t$ as random assumes the error terms are cross-sectionally correlated. On the other hand, under the assumption that $\boldsymbol{b}_i$ are independent of $\boldsymbol{x}_{it}$ and are independently distributed over i with mean zero, the error terms v_{it} are cross-sectionally independent, $Ev_{it}v_{jt} = E\boldsymbol{b}_i'\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{b}_j = 0$. Third, under the assumption the $\boldsymbol{f}_t$ is independently distributed over t, the QMLE or the generalized least squares estimator is consistent and asymptotically unbiased independent of whether $\boldsymbol{x}_{it}$ is strictly exogenous or predetermined. On the other hand, if $\boldsymbol{f}_t$ is fixed and $\boldsymbol{b}_i$ random, the QMLE or the generalized least squares (FGLS) estimator is consistent and asymptotically unbiased if $\boldsymbol{x}_{it}$ is strictly exogenous. If $\boldsymbol{x}_{it}$ contains lag dependant variables, then the QMLE or generalized least squares estimator treating initial value as fixed is inconsistent if T is fixed. If $T \to \infty$, then it is consistent. However, if N also goes to infinity and $\frac{N}{T} \to a \neq 0$, then it is asymptotically biased of order $\sqrt{a}$. Although in principle, one can formulate the initial value distribution y_{i0} to obtain asymptotically unbiased estimator, it is difficult to make a reasonable assumption under interactive effects formulations (Hsiao 2018).

10.5 QUANTILE MODELS

An analogous quantile specification for the linear models with interactive effects is

$$y_{it}(\tau) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \boldsymbol{b}_i'(\tau)\boldsymbol{f}_t. \tag{10.5.1}$$

Thus, a quantile estimator of $\boldsymbol{\beta}(\tau)$ and $\boldsymbol{b}_i(\tau)$ conditional on $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$ can be obtained by minimizing the objective function,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) - \boldsymbol{b}_i'(\tau)\boldsymbol{f}_t\right), \tag{10.5.2}$$

where

$$\rho_\tau(\nu_{it}) = (\tau - 1(\nu_{it} < 0))\nu_{it}, \tag{10.5.3}$$

$1(\cdot)$ is an indicator function, and $\nu_{it} = y_{it} - y_{it}(\tau)$. However, just like the quantile estimator for the additive individual specific effects model discussed in Section 9.1, conditional on $\boldsymbol{\beta}(\tau)$, the estimator of $\boldsymbol{b}_i(\tau)$ depends only on the ith individual's T time series observations, and the estimator of $\boldsymbol{\beta}(\tau)$ depends on $(\boldsymbol{b}_1(\tau), \ldots, \boldsymbol{b}_N(\tau))$. The estimator derived from the objective function (10.5.2) is sensitive to sample variability. To reduce the sample variability, one may follow the suggestions of Koenker (2004) to obtain the quantile estimator by minimizing the objective function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau)-\boldsymbol{b}_i(\tau)\boldsymbol{f}_t\right)+d\sum_{i=1}^{N}\sum_{j=1}^{r}\mid b_{ij}'(\tau)\mid. \tag{10.5.4}$$

However, $\boldsymbol{f}_t$ in the objective function (10.5.2) or (10.5.4) are unknown. To implement (10.5.2) or (10.5.4) we can consider the following two step estimation method:

Step 1: Estimate $\boldsymbol{\beta}$ and $F=(\boldsymbol{f}_1,\ldots,\boldsymbol{f}_T)'$ by the Bai (2009) least square method. It is shown by Bai (2009) that the r eigenvectors corresponding to the r largest eigenvalues of the matrix

$$\frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{y}_i-X_i\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i-X_i\hat{\boldsymbol{\beta}})' \tag{10.5.5}$$

is consistent when $\hat{\boldsymbol{\beta}}$ is consistent.

Step 2: Substitute $\hat{F}$ in lieu of F in (10.5.2) or (10.5.4), then implement the quantile estimation method.

Ando and Bai (2020) consider $\boldsymbol{\beta}_i(\tau)$, $\boldsymbol{b}_i(\tau)$, and $\boldsymbol{f}_t(\tau)$ as random functions of τ and are interested in estimating the function

$$y_{it}^*(\tau)=\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(\tau)+\boldsymbol{b}_i'(\tau)\boldsymbol{f}_t(\tau), \tag{10.5.6}$$

where the right-hand side of (10.5.6) is increasing in τ. Then the quantile function is to minimize the objective function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(\tau)-\boldsymbol{b}_i'(\tau)\boldsymbol{f}_t(\tau)\right). \tag{10.5.7}$$

Let $\tilde{B}(\tau)=(\boldsymbol{\beta}_1(\tau),\ldots,\boldsymbol{\beta}_N(\tau))$, $B(\tau)=(\boldsymbol{b}_1(\tau),\ldots,\boldsymbol{b}_N(\tau))$ and $F_\tau=(\boldsymbol{f}_1(\tau),\ldots,\boldsymbol{f}_T(\tau))$. Ando and Bai (2020) suggest the following algorithm:

Step 1: Initialize parameters $\hat{\tilde{B}}(\tau)$, $\hat{F}(\tau)$, and $\hat{B}(\tau)$.

Step 2: Conditional on $\hat{F}(\tau)$, update $\hat{\boldsymbol{\beta}}_i(\tau)$, and $\hat{\boldsymbol{f}}_t(\tau)$ by minimizing

$$\sum_{t=1}^{T}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{\beta}_i(\tau)-\boldsymbol{b}_i'(\tau)\hat{\boldsymbol{f}}_t(\tau)\right), \tag{10.5.8}$$

for each i, $i=1,\ldots,N$.

Step 3: Given $\hat{\boldsymbol{\beta}}_i(\tau)$ and $\hat{\boldsymbol{b}}_i(\tau)$, $i=1,\ldots,N$, update $\hat{\boldsymbol{f}}_t(\tau)$ by minimizing

$$\sum_{i=1}^{N}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}_i(\tau)-\hat{\boldsymbol{b}}_i'(\tau)\boldsymbol{f}_t(\tau)\right), \tag{10.5.9}$$

for each t, $t=1,\ldots,T$.

Step 4: Update $F(\tau)$ and $B(\tau)$ in terms of the newly obtained $\boldsymbol{f}_t(\tau)$ and $\boldsymbol{b}_i(\tau)$ in step 2 and 3.

Step 5: Repeat steps 2–4 until convergence.

The convergence of the above algorithm is quick. However, the convergence to a global optimum is not guaranteed.

10.6 FACTOR DIMENSION DETERMINATION

10.6.1 Factor Model

The desirable properties of the least squares or quasi-difference estimator depend on the knowledge of the dimension of common factors r. When r is unknown but the slope coefficients $\boldsymbol{\beta}$ are known for models (10.2.1) and (10.2.2), Bai and Ng (2002) suggest an information criterion to select the unknown dimension from $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iT})' = (\boldsymbol{y}_i - X_i \boldsymbol{\beta}_i)$, by choosing the dimension p to

$$\min_p \quad PC(p) = V(p) + pg(N, T), \tag{10.6.1}$$

where $g(N, T)$ is some suitable chosen penalty function,

$$V(p) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (v_{it} - \hat{\boldsymbol{b}}_i^{(p)'} \hat{\boldsymbol{f}}_t^{(p)})^2, \tag{10.6.2}$$

and

$$\hat{F}^{(p)} = (\hat{\boldsymbol{f}}_1^{(p)}, \ldots, \hat{\boldsymbol{f}}_p^{(p)})' \tag{10.6.3}$$

are the $\sqrt{T}$ times the eigenvectors corresponding to the p largest eigenvalues of the $T \times T$ matrix

$$\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{v}_i \boldsymbol{v}_i' \tag{10.6.4}$$

and

$$\hat{\boldsymbol{b}}_i^{(p)} = \hat{F}^{(p)'} \boldsymbol{v}_i, \tag{10.6.5}$$

under the assumption that the dimension of common factor is p. The intuition behind the Bai and Ng (2002) criterion is that if p is less than r, plim $V(p) >$ plim $V(r)$. If p is greater than r, plim $V(p)$ = plim $V(r)$. Therefore, they suggest that the penalty function $pg(N, T)$ should go to zero as $(N, T) \to \infty$, but increase with p in a finite sample such as:

$$IC_1(p) = \log(V(p)) + p\left(\frac{N+T}{NT}\right) \log\left(\frac{NT}{N+T}\right), \tag{10.6.6}$$

$$IC_2(p) = \log(V(p)) + p\left(\frac{N+T}{NT}\right) \log C_{NT}^2, \tag{10.6.7}$$

$$IC_3(p) = \log(V(p)) + p\left(\frac{\log C_{NT}^2}{C_{NT}^2}\right). \tag{10.6.8}$$

where $C_{NT} = \min(\sqrt{N}, \sqrt{T})$. The criterion is implemented with a priori specified maximum numbers of factors, $p_{\max}$. When both N and T go to infinity, Bai and Ng (2002) show that either criterion of (10.6.6)–(10.6.8) can lead to the selection of $p = r$.[5] The finite sample simulation conducted by Hsiao, Xie and Zhou (2021) appears to show that

[5] When u_{it} exhibit considerable serial correlation and the sample size is not sufficiently large, the Bai and Ng (2002) criterion may overfit (e.g., Greenaway-McGrevy, Han, and Sul 2012).

$IC_2(p)$ has the highest frequency of selecting p that is equal to r, while (10.6.8) tends to select p much greater than r.

To further improve the performance of the Bai and Ng (2002) information criterion, Alessi et al. (2010) suggest introducing a trimming parameter, $c > 0$, to the penalty function:

$$\min_{p,c} \quad IC^*_{1,c}(p) = \log V(p) + c \cdot p \left(\frac{N+T}{NT}\right) \log \left(\frac{NT}{N+T}\right), \tag{10.6.9}$$

$$\min_{p,c} \quad IC^*_{2,c}(p) = \log V(p) + c \cdot p \left(\frac{N+T}{NT}\right) \log C^2_{NT}. \tag{10.6.10}$$

The estimated p is now also a function of c as well. Alessi et al. (2010) suggest the following procedure to select c:

Divide the NT samples into subsample of sizes (N_j, T_j) with N_o between $0 < N_1 < N_2 < \cdots < N_j = N$ and T_o between $0 < T_1 < \cdots < T_j = T$. For any j, compute $\hat{p}_{c,(N_j,T_j)}$. Due to the monotonicity of $\hat{p}_{c,(T_j,N_j)}$ as a function of c, between the "small" under-penalizing values of c and the "large" over-penalizing values, there must exist a range of "moderate" values of c such that $\hat{p}_{c,(T_j,N_j)}$ is a stable function of the subsample size (T_j, N_j). The stability with respect to sample size can be measured by the empirical variance of $\hat{p}_{c,(T_j,N_j)}$ as a function of j, i.e.,

$$S_c = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{p}_{c,(T_j,N_j)} - \frac{1}{J}\sum_{l=1}^{J}\hat{p}_{c,(T_l,N_l)}\right)^2 \tag{10.6.11}$$

Consequently, the search of c can be made automatic by considering the mapping $c \to S_c$ and by choosing $\hat{p}_c = \hat{p}_{c,(NT)}$ where $\hat{c}$ belongs to an interval of c implying $S_c = 0$ and therefore a constant value of $\hat{p}_{c,(N,T)}$ is a function of c.

Alternatively, Ahn and Horenstein (2013) note that

$$\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{v}_i \boldsymbol{v}_i' \longrightarrow BB' + \frac{1}{N}\Omega. \tag{10.6.12}$$

Since rank $(BB') = r$ and $\Omega = O(1)$, the first r eigenvalues are $O(1)$ while the rest of $(T - r)$ eigenvalues converge to zero as $N \to \infty$. Ahn and Horenstein (2013) suggest an eigenvalue ratio test for selecting the dimension of $\boldsymbol{f}_t$. Let the T eigenvalues be arranged in decreasing order, $\delta_1 > \delta_2 > \cdots > \delta_T$. Form the eigenvalue ratio

$$d_j = \frac{\delta_j}{\delta_{j+1}}, \quad j = 1, \ldots, T-1. \tag{10.6.13}$$

and select the dimension of $\boldsymbol{f}_t$ as j if $d_j > c^*$ some critical value c^*.

10.6.2 Interactive Effects Model

For a panel interactive model (10.2.1) and (10.2.2), neither $\boldsymbol{\beta}$ nor r is known a priori. To implement the Bai and Ng (2002) information criterion, the Alessi et al. (2010) improved information criterion, or the Ahn and Horenstein (2013) eigenvalue ratio test, we need the knowledge of $\boldsymbol{\beta}$.

10.6.2.1 Bai (2009b) Criterion

Bai (2009b, p. 28) noted that if $k \geq r$, $\hat{\boldsymbol{\beta}}^k \to_p \boldsymbol{\beta}$ as $(N, T) \to \infty$, then $\hat{v}_{it}^k = y_{it} - \boldsymbol{x}_{it}' \hat{\boldsymbol{\beta}}^{\boldsymbol{k}} = \boldsymbol{b}_i' \boldsymbol{f}_t + u_{it} + O_p\left((NT)^{-1/2}\right)$, where $\hat{\boldsymbol{\beta}}^k$ is the least squares estimator of $\boldsymbol{\beta}$ based on the dimension of $\boldsymbol{f}_t$ is k. Then $\hat{v}_{it}^k$ is approximately a pure factor model, the remainder term $O_p\left((NT)^{-1/2}\right)$ does not affect the analysis of Bai and Ng (2002), and the aforementioned IC remains valid. On the other hand, if $k < p, \boldsymbol{\beta}$ cannot be consistently estimated. Any penalty function that converges to zero but is of greater magnitude than $O_p(1/\min(N, T))$ will lead to consistent estimation of the number of factors. Hence Bai (2009b) suggests using either

$$\overline{CP}(k) = \hat{\sigma}^2(k) + \left[k(N+T) - k^2\right] \frac{\log(NT)}{NT}, \tag{10.6.14}$$

or

$$\overline{IC}(k) = \log\left(\hat{\sigma}^2(p_{\max})\right) + \left[k(N+T) - k^2\right] \frac{\log(NT)}{NT}, \tag{10.6.15}$$

will work where $\hat{\sigma}^2(k) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^{(k)2}$ and $\hat{\sigma}^2(p_{\max}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^{(p_{\max})2}$. Or equivalently, choosing

$$\hat{p} = \underset{0 \leq k \leq p_{\max}}{\arg\min}\, \overline{CP}(k), \quad \text{or} \quad \hat{p} = \underset{0 \leq k \leq p_{\max}}{\arg\min}\, \overline{IC}(k)$$

10.6.2.2 Recursively Iterating Procedure

Hsiao, Xie, and Zhou (2021) suggest a recursively iterating procedure to implement the Bai and Ng (2002) IC and its variants:

Step 1: From any initial $\hat{\boldsymbol{\beta}}^0$, say the least squares or least squares dummy-variable estimate (Hsiao 2014, 2018) of y_{it} on $\boldsymbol{x}_{it}$, compute

$$\frac{1}{NT} \sum_{i=1}^N (\boldsymbol{y}_{it} - X_i \hat{\boldsymbol{\beta}}^{(0)})(\boldsymbol{y}_{it} - X_i \hat{\boldsymbol{\beta}}^0)', \tag{10.6.16}$$

and the corresponding eigenvalues and eigenvectors.

Step 2: Find the value of $\hat{\boldsymbol{\beta}}_p^{(j)}$ and $\hat{p}^{(j)}$ that minimizes the Bai and Ng (2002) IC or Alessi et al. (2010) improved IC where

$$V(\hat{p}^{(j)}) = \frac{1}{NT} \sum_{i=1}^N (\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_p^{(j)} - \hat{F}_{(p)}^{(j)} \boldsymbol{b}_{i(p)}^{(j)})'(\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_p^{(j)} - \hat{F}_{(p)}^{(j)} \boldsymbol{b}_{i(p)}^{(j)}), \tag{10.6.17}$$

while

$$\hat{\boldsymbol{\beta}}_p^{(j)} = \left(\sum_{i=1}^N X_i' M_{\hat{p}} X_i\right)^{-1} \left(\sum_{i=1}^N X_i' M_{\hat{p}} \boldsymbol{y}_i\right), \tag{10.6.18}$$

$$M_{\hat{p}} = \left[I_T - \left(\frac{1}{T}\right) \hat{F}_{(p)}^{(j)} \hat{F}_{(p)}^{(j)\prime}\right], \tag{10.6.19}$$

$\hat{F}^{(p)}_{(j)}$ is a $T \times \hat{p}^{(j)}$ matrix consisting of the $\sqrt{T} \times \hat{p}^{(j)}$ eigenvectors of (10.6.16) that correspond to the $\hat{p}^{(j)}$ largest eigenvalues, and

$$\hat{\boldsymbol{b}}^{(j)}_{i(p)} = \left(\frac{1}{T}\right) \hat{F}^{(j)'}_{(p)} (\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}^{(j)}_p). \tag{10.6.20}$$

Step 3: Repeat steps 1 and 2 until the solution converges or stops at a prespecified maximum iteration, say j^*.

Step 4: Compute

$$\hat{\boldsymbol{v}}^{(j)*}_i = \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}^{j*}. \tag{10.6.21}$$

Treat $\hat{\boldsymbol{v}}^{(j)*}_i$ as $\boldsymbol{v}_i$, implement either Bai and Ng (2002) information criterion or ABC or ER as in (10.6.1) or (10.6.9)–(10.6.10) or (10.6.13).

10.6.2.3 Orthogonal Projection Method

The iterative procedure is computationally intensive. Moreover, the implementation of the recursively iterating procedures requires prior assumption that $p_{\max}$ is known. Hsiao, Xie, and Zhou (2021) suggest an orthogonal projection method to implement the aforementioned criterion that needs no iteration. The basic idea is to multiply the panel interactive effects model (10.2.1) by the orthogonal projection matrix,

$$\mathbf{M}_T = \mathbf{I}_T - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-}\tilde{\mathbf{X}}, \tag{10.6.22}$$

where $\tilde{X}$ is a $T \times G$ matrix consisting of distinct columns of $X = (X_1, \ldots, X_N)$ and $(\tilde{X}'\tilde{X})^{-}$ denotes the Moore–Penrose generalized inverse of matrix $\tilde{X}'\tilde{X}$. Then

$$\mathbf{M}_T \mathbf{X}_i = 0 \text{ for } i = 1, \ldots, N. \tag{10.6.23}$$

Hence, multiplying $\mathbf{M}_T$ to both sides of (10.2.3) yields

$$\boldsymbol{y}^*_i = \mathbf{M}_T \mathbf{y}_i = \mathbf{M}_T \mathbf{F} \boldsymbol{b}_i + \mathbf{M}_T \mathbf{u}_i, \quad i = 1, \ldots, N, \tag{10.6.24}$$

where $\mathbf{M}_T \mathbf{F} = F^*$ is identical for all $i = 1, \ldots, N$. Then (10.6.24) is the factor model considered by Bai and Ng (2002), Alessi et al. (2010), or Ahn and Horenstein (2013).

When $\frac{1}{T}\boldsymbol{F}^{*'}\boldsymbol{F}^*$ converges to a nonsingular matrix of rank r, one can apply Bai and Ng's (2002) information criterion to (10.6.24) to identify the rank of the factor structure. However, the requirement for *rank* $(\frac{1}{T}\boldsymbol{F}^{*'}\boldsymbol{F}^*) = r$ requires *rank* $(\mathbf{M}_T) \geq r$. If *rank* $(\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-}\tilde{\boldsymbol{X}}') = T$, then $(\mathbf{M}_T)$ is a null matrix; thus, (10.6.24) does not exist. In the case $\mathbf{M}_T$ is a null matrix, we may consider using the orthogonal projection matrix.

$$\mathbf{M}_N = I_N - \breve{\boldsymbol{X}}^*(\breve{\boldsymbol{X}}^{*'}\breve{\boldsymbol{X}}^*)^{-}\breve{\boldsymbol{X}}^{*'}, \tag{10.6.27}$$

to eliminate $\boldsymbol{X}_t\boldsymbol{\beta}$ from (10.2.42), where $\breve{\boldsymbol{X}}^* = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)$. Multiplying (10.6.27) to (10.2.42) for each t yields

$$\breve{\boldsymbol{y}}_t = \tilde{B}\boldsymbol{f}_t + \breve{\boldsymbol{u}}_t, \quad t = 1, \ldots, T. \tag{10.6.28}$$

where $\breve{\boldsymbol{y}}_t = \mathbf{M}_N \boldsymbol{y}_t$, $\tilde{B} = \mathbf{M}_N B$ and $\breve{\boldsymbol{u}}_t = \mathbf{M}_N \boldsymbol{u}_t$. Provided $\frac{1}{N}\tilde{B}'\tilde{B}$ converges to an $r \times r$ nonsingular matrix, the Bai and Ng (2002) information criterion can again be applied to model (10.6.28) to select the dimension of factor structure.

The orthogonal projection approach avoids the computationally intensive steps 1–3 in the recursively iterating procedure, nor does it need iteration. However, to implement the orthogonal projection approach, it is important to distinguish the case between $T > N$ or $N > T$ because the stability of the eigenvalues and eigenvectors depend on the stability of the $N \times N$ or $T \times T$ estimated covariance matrix of $\boldsymbol{y}_i^*$ or $\check{\boldsymbol{y}}_t$.

Let p be the assumed maximum possible rank for the factor structure. Although we know for sure if $\frac{T}{N} > p, \mathbf{M}_T$ will have rank greater than p and if $\frac{T}{N} < \frac{1}{p}, \mathbf{M}_N$ will have rank greater than p when $(N, T) \to \infty$. If $\frac{1}{p} \leq \frac{T}{N} \leq p$, we are not sure if rank$(\mathbf{M}_T) > p$ or rank$(\mathbf{M}_N) > p$. One quick way to check this is to compute the

$$trace(\mathbf{M}_T) > p \text{ or } rank\left(\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-}\tilde{\boldsymbol{X}}'\right) \leq T - p \tag{10.6.29}$$

If (10.6.29) is not satisfied, one can compute

$$trace(\mathbf{M}_N) > p \text{ or } rank\left(\check{\boldsymbol{X}}^*(\check{\boldsymbol{X}}^{*'}\check{\boldsymbol{X}}^*)^{-}\check{\boldsymbol{X}}^{*'}\right) \leq N - p \tag{10.6.30}$$

before applying the orthogonal projection approach to (10.2.3) or (10.2.42). When either (10.6.29) or (10.6.30) holds, the standard assumptions for the panel interactive effects models (e.g., the assumptions made by Bai 2009a) ensures $rank\ (\frac{1}{T}\boldsymbol{F}^{*'}\boldsymbol{F}^*) = r$ or $rank\ (\frac{1}{N}\tilde{B}'\tilde{B}) = r$. If neither (10.6.29) nor (10.6.30) hold, one will have to resort to the computationally cumbersome joint estimation of $(\boldsymbol{\beta}, B, \boldsymbol{F})$, bearing in mind that the convergence of the recursively iterative estimation procedure between $\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{F}}^{j+1}$, and $\hat{B}^{j+1}$ depends critically on the initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ (Jiang et al. 2020).

Hsiao, Xie, and Zhou (2021) conducted some Monte Carlo simulation to consider the feasibility of the orthogonal projection method to implement the Bai and Ng (2002) information criterion to select factor dimension in the panel interactive effects model. For comparison, they also consider the recursively iterating method discussed above and the method suggested by Bai (2009b, p. 28):

$$\overline{CP}(k) = \hat{\sigma}^2(k) + \hat{\sigma}^2(p_{\max})\left[k(N+T) - k^2\right]\frac{\log(NT)}{NT}, \tag{10.6.31}$$

$$\overline{IC}(k) = \log\left(\hat{\sigma}^2(k)\right) + \left[k(N+T) - k^2\right]\frac{\log(NT)}{NT}, \tag{10.6.32}$$

where $\hat{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{v}_{it}^{(k)2}$ and $\hat{\sigma}^2(p_{\max}) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{v}_{it}^{(p_{\max})2}$, and to select the factor dimension that is

$$\hat{p} = \underset{0\leq k\leq p_{\max}}{\arg\min}\ \overline{CP}(k), \text{ or } \hat{p} = \underset{0\leq k\leq p_{\max}}{\arg\min}\ \overline{IC}(k). \tag{10.6.33}$$

They consider the following data generating process (DGP) as the base models for investigation.

DGP: Model with *three* factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \lambda_{1,i}f_{1,t} + \lambda_{2,i}f_{2,t} + \lambda_{3,i}f_{3,t} + u_{it}; \tag{10.6.34}$$

and

$$x_{k,it} = \alpha_{x,ki} + \gamma_{k,1}f_{1t} + \gamma_{k,2}f_{2t} + \eta_{k,it},\ k = 1, 2.$$

Table 10.4a. *Average number of factors selected during replications for case 1 with three factors*

		Projection					Recursive			
		BN			ABC		Bai		ABC	
N	T	IC1	IC2	IC3	ABC1	ABC2	CP	IC	ABC1	ABC2
50	40	–	–	–	–	–	1	1	3	3
	60	–	–	–	–	–	1	1	3	3
	40	10	3	10	3	3	1	1	3	3
100	80	–	–	–	–	–	1	1	3	3
	100	–	–	–	–	–	1	1	3	3
	150	–	–	–	–	–	1	1	3	3
	40	3	3	10	3	3	1	1	3	3
200	60	3	3	10	3	3	1	1	3	3
	80	10	3	10	3	3	1	1	3	3
	100	–	–	–	–	–	1	1	3	3
	150	–	–	–	–	–	1	1	3	3
	200	–	–	–	–	–	1	1	3	3
	40	3	3	10	3	3	1	1	3	3
500	60	3	3	10	3	3	1	1	3	3
	80	3	3	10	3	3	1	1	3	3
	100	3	3	10	3	3	1	1	3	3
	150	3	3	10	3	3	1	1	3	3
	200	3.2	3	10	3	3	1	1	3	3
	40	3	3	10	3	3	1	1	3	3
1000	60	3	3	10	3	3	1	1	3	3
	80	3	3	10	3	3	1	1	3	3
	100	3	3	10	3	3	1	1	3	3
	150	3	3	10	3	3	1	1	3	3
	200	3	3	10	3	3	1	1	3	3

Notes: 1. "Projection" and "Recursive" refer to orthogonal projection and recursive method, respectively.
2. "BN", "Bai," and "ABC" refers to the IC of Bai and Ngs (2002), Bai (2009b), and Alessi et al. (2010), respectively.
3. "-" indicates that the orthogonal projection does not work for this particular (N, T).
Source: Hsiao, Xie, and Zhou (2021).

For this DGP, they assume that $\alpha_{x,1i}, \alpha_{x,2i} \sim_{iid} N(0,1)$ and $\eta_{k,it} = \rho_{ki}\eta_{k,it-1} + v_{k,it}$ with $\rho_{ki} \sim_{iid} U(0.1, 0.9)$, and that $v_{k,it} \sim_{iid} \left(0, \sigma^2_{vk,i}\right)$ for $k = 1, 2$ and $\sigma^2_{v1,i}$, $\sigma^2_{v2,i}$ are independent draws from $0.5\left(1 + \chi^2(2)\right)$. The common factors $f_{j,t}$ are i.i.d. draw from $N(0,1)$ for $j = 1, 2, 3,$ and the factor loadings are set as

- $\lambda_{1,i} \sim_{iid} N(1,1)$, $\lambda_{2,i} \sim_{iid} N(1, 0.5)$, $\lambda_{3,i} \sim_{iid} N(1,2)$.
- $\gamma_{1,1} \sim_{iid} U(0,1)$, $\gamma_{1,2} \sim_{iid} U(0,2)$, and $\gamma_{1,3} \sim_{iid} U(-1,2)$.
- $\gamma_{2,1} \sim_{iid} U(-1,1)$, $\gamma_{2,2} \sim_{iid} U(1,2)$, and $\gamma_{2,3} \sim_{iid} U(0,1)$.

For the errors in (10.6.34), they consider the following three specifications.

Case 1: Heteroscedastic errors
The errors u_{it} are assumed independent $N\left(0, \sigma^2_{u,i}\right)$ where $\sigma^2_{u,i}$ are independent draws from $0.5\left(1 + \chi^2(2)\right)$.

Case 2: Serially correlated errors
The errors are generated as

$$u_{it} = \rho_{ui} u_{it-1} + \epsilon_{it}, \tag{10.6.35}$$

Table 10.4b. *Percentage of correctly estimating the number of factors for Case 1 with three factors*

		Projection					Recursive			
		BN			ABC		Bai		ABC	
N	T	IC1	IC2	IC3	ABC1	ABC2	CP	IC	ABC1	ABC2
50	40	–	–	–	–	–	0%	0%	98%	99%
	60	–	–	–	–	–	0%	0%	99%	100%
100	40	0%	98%	0%	98%	99%	0%	0%	100%	100%
	80	–	–	–	–	–	0%	0%	100%	100%
	100	–	–	–	–	–	0%	0%	100%	100%
	150	–	–	–	–	–	0%	0%	100%	100%
200	40	100%	100%	0%	100%	100%	0%	0%	100%	100%
	60	100%	100%	0%	100%	100%	0%	0%	100%	100%
	80	0%	100%	0%	100%	100%	0%	0%	100%	100%
	100	–	–	–	–	–	0%	0%	100%	100%
	150	–	–	–	–	–	0%	0%	100%	100%
	200	–	–	–	–	–	0%	0%	100%	100%
500	40	100%	100%	0%	100%	100%	0%	0%	100%	100%
	60	100%	100%	0%	100%	100%	0%	0%	100%	100%
	80	100%	100%	0%	100%	100%	0%	0%	100%	100%
	100	100%	100%	0%	100%	100%	0%	0%	100%	100%
	150	100%	100%	0%	100%	100%	0%	0%	100%	100%
	200	83%	100%	0%	100%	100%	0%	0%	100%	100%
1000	40	100%	100%	0%	100%	100%	0%	0%	100%	100%
	60	100%	100%	0%	100%	100%	0%	0%	100%	100%
	80	100%	100%	0%	100%	100%	0%	0%	100%	100%
	100	100%	100%	0%	100%	100%	0%	0%	100%	100%
	150	100%	100%	0%	100%	100%	0%	0%	100%	100%
	200	100%	100%	0%	100%	100%	0%	0%	100%	100%

See notes of Table 10.4a.

where $\rho_{ui} \sim_{iid} U(0.1, 0.9)$ and $\epsilon_{it} \sim_{iid} N\left(0, \sigma^2_{\epsilon,i}\right)$ with $\sigma^2_{\epsilon,i}$ is independent draws from $0.5\left(1 + \chi^2(2)\right)$.

Case 3: Weakly dependent errors
Hsiao, Xie, and Zhou (2021) also consider errors generated as

$$u_{it} = \left(1 + b^2\right)v_{it} + bv_{i+1,t} + bv_{i-1,t}, \tag{10.6.36}$$

where $b = 0.2$ and $v_{it} \sim_{iid} N\left(0, \sigma^2_{v,i}\right)$ where $\sigma^2_{v,i}$ are independent draws from $0.5\left(1 + \chi^2(2)\right)$.

For the DGP described above, the true value of β_1 and β_2 are set at $\beta_1 = 1$ and $\beta_2 = 2$. The combinations of $N = 50, 100, 200, 500, 1{,}000$ and $T = 40, 60, 80, 100, 150, 200$ are used for simulation. The number of replication is set at 1,000. For the implementation of the above approaches, they the projection model (10.6.23) when (10.6.29) is satisfied, and consider the projection model (10.6.27) when (10.6.30) is satisfied. For the choice of the number of blocks in the ABC approach, following Alessi et al. (2010), they let $c_{\max} = 2$ and $J = \left[s_y N/20\right]$ where s^2_y denotes the sample variance of the dependent variable y and $[\cdot]$ denotes the integer part of $\cdot$. Moreover, following Bai and Ng (2002), they standardize

Table 10.5a. *Average number of factors selected during replications for case 3 with three factors*

		Projection					Recursive			
		BN			ABC		Bai		ABC	
N	*T*	IC1	IC2	IC3	ABC1	ABC2	CP	IC	ABC1	ABC2
50	40	–	–	–	–	–	1	1	3	3
	60	–	–	–	–	–	1	1	3.2	3.1
	40	10	2.9	10	3	3	1	1	3	3
100	80	–	–	–	–	–	1	1	3	3
	100	–	–	–	–	–	1	1	3	3
	150	–	–	–	–	–	1	1	3	3
	40	3	2.9	10	3	3	1	1	3	3
200	60	3	3	10	3	3	1	1	3	3
	80	10	3	10	3	3	1	1	3	3
	100	–	–	–	–	–	1	1	3	3
	150	–	–	–	–	–	1	1	3	3
	200	–	–	–	–	–	1	1	3	3
	40	3	3	10	3	3	1	1	3	3
500	60	3	3	10	3	3	1	1	3	3
	80	3	3	10	3	3	1	1	3	3
	100	3	3	10	3	3	1	1	3	3
	150	3	3	10	3	3	1	1	3	3
	200	3.7	3	10	3	3	1	1	3	3
	40	3	3	10	3	3	1	1	3	3
1000	60	3	3	10	3	3	1	1	3	3
	80	3	3	10	3	3	1	1	3	3
	100	3	3	10	3	3	1	1	3	3
	150	3	3	10	3	3	1	1	3	3
	200	3	3	10	3	3	1	1	3	3

See notes of Table 10.4a.

the projection data for both the BN and ABC approaches. To implement Bai's (2009) iterative approach, for a given number of factors, they first use the pooled estimator as the initial estimator, then iterate between the least squares and principal component analysis to estimate the coefficient and the dimension of factors, respectively. They set the maximum number of iterations at 10. Based on these estimates, they use (10.6.31) and (10.6.32) to determine the dimension of unobserved factors. Finally, for the implementation of the recursively iteratings approach to implement IC, they set the maximum number of iteration to 5 to reduce the computation time.

Hsiao, Xie and Zhou (2021) report the average number of dimensions in each replication and the percentage of correctly estimating the number of factors for these cases. Their results for case 1 and 3 are provided in Tables 10.4a–10.5a. Their limited simulation results appear to show the following: (i) Bai's (2009) CP and IC tend to select the factor dimension less than the true dimension. (ii) Among the three information criterion suggested by Bai and Ng (2002), IC2 appears to dominate. (iii) The Alessi et al. (2010)-modified BN information criterion does improve the accuracy of selecting the correct number of factors in a finite sample. In general, the recursively iterating procedure to implement the Alessi et al. (2010) modified IC performs very well. (iv) The performance of the orthogonal projection method depends on the stability of the estimated $T \times T$ covariance matrix

Table 10.5b. *Percentage of correctly estimating the number of factors for Case 3 with three factors*

		Projection					Recursive			
		BN			ABC		Bai		ABC	
N	T	IC1	IC2	IC3	ABC1	ABC2	CP	IC	ABC1	ABC2
50	40	–	–	–	–	–	0%	0%	96%	98%
	60	–	–	–	–	–	0%	0%	96%	98%
	40	0%	94%	0%	97%	97%	0%	0%	100%	100%
100	80	–	–	–	–	–	0%	0%	100%	100%
	100	–	–	–	–	–	0%	0%	100%	100%
	150	–	–	–	–	–	0%	0%	100%	100%
	40	100%	97%	0%	100%	100%	0%	0%	100%	100%
200	60	99%	100%	0%	100%	100%	0%	0%	100%	100%
	80	0%	100%	0%	100%	100%	0%	0%	100%	100%
	100	–	–	–	–	–	0%	0%	100%	100%
	150	–	–	–	–	–	0%	0%	100%	100%
	200	–	–	–	–	–	0%	0%	100%	100%
	40	100%	100%	0%	100%	100%	0%	0%	100%	100%
500	60	100%	100%	0%	100%	100%	0%	0%	100%	100%
	80	100%	100%	0%	100%	100%	0%	0%	100%	100%
	100	100%	100%	0%	100%	100%	0%	0%	100%	100%
	150	100%	100%	0%	100%	100%	0%	0%	100%	100%
	200	50%	100%	0%	100%	100%	0%	0%	100%	100%
	40	100%	100%	0%	100%	100%	0%	0%	100%	100%
1000	60	100%	100%	0%	100%	100%	0%	0%	100%	100%
	80	100%	100%	0%	100%	100%	0%	0%	100%	100%
	100	100%	100%	0%	100%	100%	0%	0%	100%	100%
	150	100%	100%	0%	100%	100%	0%	0%	100%	100%
	200	100%	100%	0%	100%	100%	0%	0%	100%	100%

See notes of Table 10.4a.

of $\mathbf{y}_i^*$ or or the $N \times N$ covariance matrix $\check{\mathbf{y}}_t$. It is important to distinguish the relative size between N and T for the conditions (10.6.29) or (10.6.30) in implementing the orthogonal projection approach. (v) When N and T satisfy (10.6.29) or (10.6.30), the performance of the orthogonal projection method is comparable with the recursively iterating procedure of Alessi et al. (2010). However, the computation time for the orthogonal projection approach is at least seven times faster than the recursively iterating approach based on Alessi et al. (2010) for case 1 when $N = 1000$ and $T = 40$ if we set the maximum number of iteration to 10.