

# PANEL COINTEGRATION WITH GLOBAL STOCHASTIC TRENDS

Jushan Bai \*      Chihwa Kao<sup>†</sup>      Serena Ng <sup>‡</sup>

This Draft: November 28, 2006

## Abstract

This paper studies estimation of panel cointegrated models with cross-sectional dependence generated by unobserved global stochastic trends. The standard least squares estimator is, in general, inconsistent owing to the spuriousness induced by the unobservable I(1) trends. We propose two iterative procedures that jointly estimate the slope parameters and the stochastic trends. The resulting estimators are referred to respectively as CupBC (continuously-updated and bias-corrected) and the CupFM (continuously-updated and fully-modified) estimators. We establish their consistency and derive their limiting distributions. Both are asymptotically unbiased and asymptotically normal and permit inference to be conducted using standard test statistics. The estimators are also valid when there are mixed stationary and non-stationary factors, as well as when the factors are all stationary.

*JEL Classification:* C13; C33

*Keywords:* Panel data; Cross-sectional dependence; Factor analysis; Fully modified estimator.

---

\*Department of Economics, New York University, New York, NY 10003, USA, and Email Jushan.Bai@nyu.edu

<sup>†</sup>Center for Policy Research and Department of Economics, Syracuse University, Syracuse, NY 13244-1020, USA.  
Email cdkao@maxwell.syr.edu

<sup>‡</sup>Department of Economics, University of Michigan, Ann Arbor, MI 48109, USA, Email:serena.Ng@umich.edu  
We are grateful for discussions with Yixiao Sun, Joon Park, Yoosoon Chang, and Joakim Westerlund. We would also like to thank participants in econometrics seminars at NYU, Texas A&M, CESG 2006, Academia Sinica and National Taiwan University for helpful comments. Bai and Ng gratefully acknowledge financial support from the NSF (grants and SES-0549978)

## 1 Introduction

This paper is concerned with estimating panel cointegrated models using a large panel of data. Our focus is on the slope parameters on the non-stationary regressors when the units share common sources of non-stationary variation in the form of global stochastic trends. We provide a framework for estimation and inference. We propose two iterative procedures that estimate the latent common trends (hereafter factors) and the slope parameters jointly. The estimators are  $\sqrt{nT}$  consistent and asymptotically normal. As such, inference can be made using standard  $t$  and Wald tests. The estimators are also valid when some or all of the common factors are stationary, and when some of the regressors are stationary.

Panel data have long been used to study and test economic hypothesis. The two dimensions of panel data bring in additional information to permit analysis that would otherwise be inefficient, if not impossible, with time series or cross-sectional data alone. A new development in recent years is the use of ‘large dimensional panels’, meaning that the sample size in the time series ( $T$ ) and the cross-section ( $n$ ) dimensions are both large. This is in contrast to traditional panels in which we have data of many units over a short time span, or of a few variables over a long horizon. Many researchers have come up with new ideas to exploit the rich information in large panels.<sup>1</sup> However, large panels also raise econometric issues of their own. In this analysis, we tackle two of these issues:- the data  $(y_{it}, x_{it})$  are non stationary, and the structural errors  $e_{it} = y_{it} - x'_{it}\beta$  are neither iid across  $i$  nor over  $t$ . Instead, they are cross-sectionally dependent and strongly persistent and possibly non-stationary. In addition,  $e_{it}$  are also correlated with the explanatory variables  $x_{it}$ . These problems are dealt with by putting a factor structure on  $e_{it}$  and modelling the factor process explicitly.

The presence of common sources of non-stationarity leads naturally to the concept of cointegration. In a small panel made up of individually I(1) (or unit root) processes  $y_t$  and  $x_t$ , where small means that the dimension of  $y_t$  plus the dimension of  $x_t$  is treated as fixed in asymptotic analysis, cointegration as defined in Engle and Granger (1986) means that there exists a cointegrating vector,  $(1 \quad -\beta')$ , such that the linear combinations  $y_t - x'_t\beta$  are stationary, or are an I(0) processes. In a panel data model specified by  $y_{it} = x'_{it}\beta + e_{it}$  where  $y_{it}$  and  $x_{it}$  are I(1) processes, and that  $e_{it}$  are iid across  $i$ , cointegration is said to hold if  $e_{it}$  are ‘jointly’ I(0), or in other words,  $(1, -\beta)$  is the common cointegrating vector between  $y_{it}$  and  $x_{it}$  for *all*  $n$  units. A large literature

---

<sup>1</sup>See, for example, Baltagi (2005), Hsiao (2003), Pesaran and Smith (1995), Kao (1999), Bai and Ng (2002), and Moon and Phillips (2000, 2004) in the context of testing the unit root hypothesis using panel data. Stock and Watson (2002) suggests a new approach to forecasting, while Bernanke Boivin (2003) suggest new formulations of vectorautoregressions to exploit the information in large panels.

on panel cointegration already exists<sup>2</sup> for modeling panel cointegration when  $e_{it}$  is cross-sectionally independent.

In practice, we have large panels of data of which  $y_{it}$  are variables like output of firms, or consumption of households, or the national product of countries, or the value-added of industries, while the corresponding  $x_{it}$  are factor inputs, household earnings, national employment, and sectoral factor prices. While macroeconomic theory often starts with the premise that firms, households, and industries are affected by common shocks such as arising aggregate productivity, from monetary and fiscal policies, the panel cointegration model under cross-section independence has no role for such common sources of variation. Failure to account for cross-section dependence will invalidate estimation and inference of  $\beta$ .<sup>3</sup> In view of this, more recent work has allowed for cross-sectional dependence of  $e_{it}$  when testing for the null hypothesis panel cointegration.<sup>4</sup> However, not much attention has been paid to the issue of estimation and inference of parameters in a panel cointegrated model with cross-sectional dependence. The present paper fills this gap.

The framework we adopt is that  $e_{it}$  has a common component and a stationary idiosyncratic component. That is,  $e_{it} = \lambda_i' F_t + u_{it}$ , so that panel cointegration holds when  $u_{it} = y_{it} - \beta x_{it} - \lambda_i' F_t$  is jointly stationary. We focus on estimation and inference about  $\beta$  when  $F_t$  is non-stationary. A regression of  $y_{it}$  on  $x_{it}$  will give a consistent estimator for  $\beta$  when  $F_t$  is  $I(0)$ . However, if  $F_t$  is  $I(1)$ , a regression of  $y_{it}$  on  $x_{it}$  is spurious since  $e_{it}$  is not only cross-sectionally correlated, but also non-stationary. We deal with the problem by treating the common  $I(1)$  variables as parameters. These are estimated jointly with  $\beta$  using an iterated procedure. The procedure is shown to yield a consistent estimator of  $\beta$ , but the estimator is asymptotically biased. We then construct two estimators to account for the bias arising from endogeneity and serial correlation so as to re-center the limiting distribution around zero. The first, denoted CupBC, estimates the asymptotic bias directly. The second, denoted CupFM, modifies the data so that the limiting distribution does not depend on nuisance parameters. Both are ‘continuously updated’ (Cup) procedures and require iteration till convergence. The estimators are  $\sqrt{nT}$  consistent for the common slope coefficient vector,  $\beta$ . The estimators enables use of standard test statistics such as  $t$ ,  $F$ , and  $\chi^2$  for inference. The estimators are robust to mixed  $I(1)/I(0)$  factors, as well as mixed  $I(1)/I(0)$  regressors. Thus,

---

<sup>2</sup>See, for example, literature, e.g., Phillips and Moon (1999) and Kao (1999). Recent surveys can be found in Baltagi and Kao (2000) and Breitung and Pesaran (2005).

<sup>3</sup>Andrews (2005) showed that cross-section dependence induced by common shocks can yield inconsistent estimates. Andrews’ argument is made in the context of a single cross section and for stationary regressors and errors. For a single cross section, not much can be done about common shocks. But for panel data, we can explore the common shocks to yield consistent procedures.

<sup>4</sup>See, for example, Phillips and Sul (2003), Moon and Perron (2004), Bai and Ng (2004), Harris et al. (2004), Breitung and Das (2005a, 2005b), Gengenbach et al. (2004, 2005), Pesaran (2005), Westerlund (2005, 2006), Westerlund and Edgerton (2006), Choi (2006), Chang (2002, 2004), and Chang and Song (2005).

our approach is an alternative to the solution proposed in Bai and Kao (2005) for stationary factors. As we argue below, the Cup estimators have some advantages that make an analysis of their properties interesting in its own right.

The rest of the paper is organized as follows. Section 2 describes the basic model of panel cointegration with unobservable common stochastic trends. Section 3 develops the asymptotic theory for the continuously-updated and fully-modified estimators. Section 4 discusses the problem of testing cross-sectional dependence. Section 5 examines issues related to incidental trends, mixed I(0)/I(1) regressors and mixed I(0)/I(1) common shocks. Section 6 presents Monte Carlo results to illustrate the finite sample properties of the proposed estimators. Section 7 provides a brief conclusion. The appendix contains the technical materials.

## 2 The Model

Consider the model

$$y_{it} = x'_{it}\beta + e_{it}$$

where for  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ ,  $y_{it}$  is a scalar,

$$x_{it} = x_{it-1} + \varepsilon_{it}. \tag{1}$$

is a set of  $k$  non-stationary regressors,  $\beta$  is a  $k \times 1$  vector of the common slope parameters, and  $e_{it}$  is the regression error. Suppose  $e_{it}$  is stationary and *iid* across  $i$ . Then it is easy to show that the pooled least squares estimator of  $\beta$  defined by

$$\hat{\beta}_{LS} = \frac{\sum_{i=1}^n \sum_{t=1}^T x_{it} y_{it}}{\sum_i \sum_t x_{it}^2} \tag{2}$$

is, in general,  $T$  consistent.<sup>5</sup> Similar to the case of time series regression considered by Phillips and Hansen (1990), the limiting distribution is shifted away from zero due to an asymptotic bias induced by the long run correlation between  $e_{it}$  and  $\varepsilon_{it}$ . The exception is when  $x_{it}$  is strictly exogenous, in which case the estimator is  $\sqrt{nT}$  consistent. The asymptotic bias can be estimated, and a panel fully-modified estimator can be developed along the lines of Phillips and Hansen (1990) to achieve  $\sqrt{nT}$  consistency and asymptotic normality.

The cross-section independence assumption is restrictive and difficult to justify when the data under investigation are economic time series. We now relax this assumption and model the cross-section dependence by imposing a factor structure on  $e_{it}$ . That is,

$$e_{it} = \lambda'_i F_t + u_{it}$$

---

<sup>5</sup>The estimator can be regarded as  $\sqrt{nT}$  consistent but with a bias of order  $O(\sqrt{n})$ . Up to the bias, the estimator is also asymptotically normal.

where  $F_t$  is a  $r \times 1$  vector of latent common factors,  $\lambda_i$  is a  $r \times 1$  vector of factor loadings and  $u_{it}$ . If we continue to assume that  $e_{it}$  is stationary, then it must be the case that  $F_t$  and  $u_{it}$  are also stationary. Note that the moment matrix of the I(1) regressors are  $O_p(T^2)$  while the sample cross-covariance of the stationary errors  $e_{it}$  is  $O_p(T)$ . Therefore, a consistent estimator of the regression coefficients can still be obtained even when the cross-section dependence is ignored. Using this property, Bai and Kao (2005) considered a two-step fully modified estimator (2sFM). In the first step, pooled OLS is used to obtain a consistent estimate of  $\beta$ . The residuals are then used to construct a fully-modified (FM) estimator along the line of Phillips and Hansen (1990). Essentially, nuisance parameters induced by cross-section correlation are dealt with just like serial correlation by suitable estimation of the long-run covariance matrices.

The 2sFM treats the I(0) common shocks as part of the error processes. However, an alternative estimator can be developed by rewriting the regression model as

$$y_{it} = x'_{it}\beta + \lambda'_i F_t + u_{it} \quad (3)$$

If  $F_t$  was observed but stationary and  $u_{it}$  is also stationary, then  $\beta$  can be estimated by pooled OLS. This regression involves mixed I(1) and I(0) regressors and the limit theory is provided by Park and Phillips (1988). Except for the case of strict exogeneity, the limiting distribution once again contains nuisance parameters but a modified estimator can be constructed to remove the nuisance parameters. When  $F_t$  is unobservable but I(0), moving  $F_t$  from the error term to the regression function (treated as parameters) is desirable for the following reason. If some components of  $x_{it}$  are actually I(0), treating  $F_t$  as part of error process will yield an inconsistent estimate for  $\beta$  when  $F_t$  and  $x_{it}$  are correlated. The simultaneity bias is now of the same order as the convergence rate of the coefficient estimates on the stationary regressors. Additionally, the regression error after explicitly controlling for  $F_t$  is necessarily no larger than the one that does not control for  $F_t$ . The approach of estimating  $\beta$  from (3) was suggested but not explored in Bai and Kao (2005).

When  $F_t$  is nonstationary, which is the primary focus of this paper, estimating  $\beta$  from (3) versus pooled OLS in (2) is no longer valid. More precisely, if

$$F_t = F_{t-1} + \eta_t,$$

then  $e_{it}$  is I(1) and pooled OLS in (2) is, in general, not consistent. To see this, consider the following data generating process for  $x_{it}$

$$x_{it} = \tau'_i F_t + \xi_{it} \quad (4)$$

with  $\xi_{it}$  being I(1) such that  $\xi_{it} = \xi_{it-1} + \zeta_{it}$ . For simplicity, assume there is a single factor. It follows that  $x_{it}$  is I(1) and can be written as (1) with  $\varepsilon_{it} = \tau'_i \eta_t + \zeta_{it}$ . The pooled OLS can be

written as

$$\hat{\beta}_{LS} - \beta = \frac{(\frac{1}{n} \sum_{i=1}^n \tau_i \lambda_i) (\frac{1}{T^2} \sum_{t=1}^T F_t^2)}{\frac{1}{nT^2} \sum_i \sum_t x_{it}^2} + O_p(n^{-1/2}) + O_p(T^{-1})$$

If  $\tau_i$  and  $\lambda_i$  are correlated, or when they have non-zero means, the first term on the right hand side is  $O_p(1)$ , implying inconsistency of the pooled OLS. The best convergence rate is  $\sqrt{n}$  when  $x_{it}$  and  $F_t$  are independent random walks. The problem arises because as seen from (3), we now have a panel model with non-stationary regressors  $x_{it}$  and  $F_t$ , and in which  $u_{it}$  is stationary by assumption. This means that  $y_{it}$  cointegrates with  $x_{it}$  and  $F_t$  with cointegrating vector  $(1, -\beta', \lambda_i)$ . Omitting  $F_t$  creates a first order bias that is not correctable. It is worth noting that the cointegrating vector varies with  $i$  because the factor loading is unit specific. Thus, in a panel with  $n$  units and common stochastic trends, there can be as many as  $n$  cointegrating vectors. Estimation of the parameter of interest  $\beta$  involves a new methodology.

In the rest of the paper, we will show how to obtain  $\sqrt{nT}$  consistent and asymptotically normal estimates of  $\beta$  when the data generating process is characterized by (3) assuming that  $x_{it}$  and  $F_t$  are both  $I(1)$ , and that  $x_{it}$ ,  $F_t$  and  $u_{it}$  are potentially correlated. We will refer to  $F_t$  as the global stochastic trends since they are shared by each cross-sectional unit. Hereafter, we write the integral  $\int_0^1 W(s)ds$  as  $\int W$  when there is no ambiguity. We define  $\Omega^{1/2}$  to be any matrix such that  $\Omega = (\Omega^{1/2}) (\Omega^{1/2})'$ , and  $BM(\Omega)$  to denote Brownian motion with the covariance matrix  $\Omega$ . We use  $\|A\|$  to denote  $(tr(A'A))^{1/2}$ ,  $\xrightarrow{d}$  to denote convergence in distribution,  $\xrightarrow{p}$  to denote convergence in probability,  $[x]$  to denote the largest integer less than or equal to  $x$ . We let  $M < \infty$  be a generic positive number, not depending on  $T$  or  $n$ . Unless indicated explicitly, all limits are taken as  $(n, T) \rightarrow \infty$ . We also define the matrix that projects onto the orthogonal space of  $z$  as  $M_z = I_T - z(z'z)^{-1}z'$ . We will use  $\beta^0$ ,  $F_t^0$ , and  $\lambda_i^0$  to denote the true common slope parameters, true common trends, and the true factor loading coefficients.

Our analysis is based on the following assumptions.

**Assumption 1** *Factor and Loading:*

- (a)  $E \|\lambda_i^0\|^4 \leq M$ . As  $n \rightarrow \infty$ ,  $\frac{1}{n} \sum_{i=1}^n \lambda_i^0 \lambda_i^{0'} \xrightarrow{p} \Sigma_\lambda$ , a  $r \times r$  positive definite matrix.
- (b)  $E \|F_t^0\|^4 \leq M$ . As  $T \rightarrow \infty$ ,  $\frac{1}{T^2} \sum_{i=1}^n F_t^0 F_t^{0'} \xrightarrow{d} \int B_\eta B_\eta'$ , a  $r \times r$  random matrix, where  $B_\eta$  is a vector of Brownian motions with covariance matrix  $\Omega_\eta$  a positive definite matrix.

**Assumption 2** Let  $w_{it} = (u_{it}, \varepsilon'_{it}, \eta_t)'$ . For each  $i$ ,  $w_{it} = \Pi_i(L)v_{it} = \sum_{j=0}^{\infty} \Pi_{ij} v_{it-j}$ ,  $\sum_{j=0}^{\infty} j^a \|\Pi_{ij}\| < \infty$ ,  $|\Pi_i(1)| \neq 0$  for some  $a > 1$ , where  $v_{it}$  is i.i.d. over  $t$ . In addition,  $E v_{it} = 0$ ,  $E(v_{it} v_{it}') = \Sigma_v > 0$ , and  $E \|v_{it}\|^8 \leq M < \infty$ .

**Assumption 3** *Serial and cross-sectional weak dependence and heterokedasticity*

- (a)  $E(u_{it}) = 0$ ,  $E|u_{it}|^8 < M$ .
- (b)  $E(u_{it}u_{js}) = \sigma_{ij,ts}$ ,  $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$  for all  $(t, s)$  and  $|\sigma_{ij,ts}| \leq \tau_{ts}$  for all  $(i, j)$  such that (i)  $\frac{1}{n} \sum_{i,j=1}^n \bar{\sigma}_{ij} \leq M$ , (ii)  $\frac{1}{T} \sum_{t,s=1}^T \tau_{ts} \leq M$ , and (iii)  $\frac{1}{nT} \sum_{i,j,t,s=1} |\sigma_{ij,ts}| \leq M$ .
- (c) For every  $(t, s)$ ,  $E \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n [u_{is}u_{it} - E(u_{is}u_{it})] \right|^4 \leq M$ .
- (d)  $\frac{1}{nT^2} \sum_{t,s,u,v} \sum_{i,j} |\text{cov}(u_{it}u_{is}, u_{ju}u_{jv})| \leq M$  and  $\frac{1}{nT^2} \sum_{t,s} \sum_{i,j,k,l} |\text{cov}(u_{it}u_{js}, u_{ku}u_{ls})| \leq M$ .

**Assumption 4**  $\{x_{it}, F_t^0\}$  are not cointegrated.

Assumption 1 is standard in the panel factor literature. Assumption 3 allows for limited time series and cross-sectional dependence in the error term,  $u_{it}$ . Heteroskedasticity in both time series and cross-sectional dimensions for  $u_{it}$  is allowed as well. The assumption that  $\Omega_\eta$  is positive definite rules out cointegration among the components of  $F_t^0$ . Assumption 4 also rules out the cointegration between  $x_{it}$  and  $F_t^0$ .

Assumption 2 implies that a multivariate invariance principle for  $w_{it}$  holds, i.e., the partial sum process  $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T \cdot]} w_{it}$  satisfies:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T \cdot]} w_{it} \xrightarrow{d} B_i(\cdot) = B(\Omega_i) \text{ as } T \rightarrow \infty \text{ for all } i,$$

where

$$B_i = [ B_{ui} \quad B_{\varepsilon i} \quad B_\eta ]'.$$

The long-run covariance matrix of  $\{w_{it}\}$  is given by

$$\Omega_i = \sum_{j=-\infty}^{\infty} E(w_{i0}w'_{ij}) = \begin{bmatrix} \Omega_{ui} & \Omega_{u\varepsilon i} & \Omega_{u\eta i} \\ \Omega_{\varepsilon ui} & \Omega_{\varepsilon i} & \Omega_{\varepsilon \eta i} \\ \Omega_{\eta ui} & \Omega_{\eta \varepsilon i} & \Omega_\eta \end{bmatrix} \quad (5)$$

are partitioned conformably with  $w_{it}$ . Define the one-sided long-run covariance

$$\Delta_i = \sum_{j=0}^{\infty} E(w_{i0}w'_{ij}) = \begin{bmatrix} \Delta_{ui} & \Delta_{u\varepsilon i} & \Delta_{u\eta i} \\ \Delta_{\varepsilon ui} & \Delta_{\varepsilon i} & \Delta_{\varepsilon \eta i} \\ \Delta_{\eta ui} & \Delta_{\eta \varepsilon i} & \Delta_\eta \end{bmatrix}. \quad (6)$$

For future reference, it will be convenient to group elements corresponding to  $\varepsilon_{it}$  and  $\eta_t$  taken together. Let

$$B_{bi} = [ B_{\varepsilon i} \quad B_\eta ]' \quad \Omega_{bi} = \begin{bmatrix} \Omega_{\varepsilon i} & \Omega_{\varepsilon \eta i} \\ \Omega_{\eta \varepsilon i} & \Omega_\eta \end{bmatrix}.$$

Then  $B_i$  can be rewritten as

$$B_i = \begin{bmatrix} B_{ui} \\ B_{bi} \end{bmatrix} = \begin{bmatrix} \Omega_{u.bi}^{1/2} & \Omega_{ubi}\Omega_{bi}^{-1/2} \\ 0 & \Omega_{bi}^{1/2} \end{bmatrix} \begin{bmatrix} V_i \\ W_i \end{bmatrix}$$

where  $[V_i \ W_i]'$  is a standardized Brownian motion and

$$\Omega_{u.bi} = \Omega_{ui} - \Omega_{ubi}\Omega_{bi}^{-1}\Omega_{bui}$$

is the long-run conditional variance of  $u_{it}$  given  $(\Delta x'_{it}, \Delta F_t^{0'})'$ . Note that  $\Omega_{bi} > 0$  since we assume that there is no cointegration relationship in  $(x'_{it}, F_t^{0'})'$  in Assumption 4.

The following assumption is needed only when we derive the limiting distribution of various estimators.

**Assumption 5** *The idiosyncratic errors  $u_{it}$  are cross-sectionally independent.*

Consistency of the Cup estimator (to be introduced later) does not require this assumption. It is also possible to derive the limiting distribution of the Cup estimator under cross-sectional correlation, but estimating nuisance parameters would be more difficult.

### 3 Estimation

In this section, we first consider the problem of estimating  $\beta$  when  $F$  is observed. We then consider two iterative procedures that jointly estimate  $\beta$  and  $F$ . The procedures yield two estimators that are  $\sqrt{nT}$  consistent and asymptotically normal. These estimators, denoted CupBC and CupFM, are presented in subsections 3.2 and 3.3.

#### 3.1 Estimation when $F$ is observed

The true model (3) in vector form, is

$$y_i = x_i\beta^0 + F^0\lambda_i^0 + u_i$$

where

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, x_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ \vdots \\ x'_{iT} \end{bmatrix}, F = \begin{bmatrix} F'_1 \\ F'_2 \\ \vdots \\ F'_T \end{bmatrix}, u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

Define  $\Lambda = (\lambda_1, \dots, \lambda_n)'$  to be an  $n \times r$  matrix. In matrix notation

$$y = X\beta^0 + F^0\Lambda^0 + u.$$

Given data  $y$ ,  $x$ , and  $F^0$ , the least squares objective function is

$$S_{nT}^0(\beta, \Lambda) = \sum_{i=1}^n (y - x_i\beta - F^0\lambda_i)' (y - x_i\beta - F^0\lambda_i).$$

After concentrating out  $\lambda$ , the least squares estimator for  $\beta$  is then

$$\tilde{\beta}_{LS} = \left( \sum_{i=1}^n x_i' M_{F^0} x_i \right)^{-1} \sum_{i=1}^n x_i' M_{F^0} y_i.$$

Let The least squares estimator has the following properties.

**Proposition 1** *Under Assumptions 1-5,*

$$\sqrt{nT} (\tilde{\beta}_{LS} - \beta^0) - \sqrt{n}\phi_{nT}^0 \xrightarrow{d} N(0, \Sigma^0)$$

where

$$\phi_{nT}^0 = \left[ \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{F^0} x_i \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \theta_i^0 \right] \quad (7)$$

$$\Sigma^0 = D^{-1} \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{u,bi} E \left( \int Q_i Q_i' \right) \right] D^{-1}, \quad D = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int Q_i Q_i' \right) \quad (8)$$

with

$$\begin{aligned} Q_i &= B_{\varepsilon i} - \left( \int B_{\varepsilon i} B_{\eta}' \right) \left( \int B_{\eta} B_{\eta}' \right)^{-1} B_{\eta}, \\ \theta_i^0 &= \frac{1}{T} x_i' M_{F^0} \Delta b_i \Omega_{bi}^{-1} \Omega_{bui} + \begin{pmatrix} I_k & -\delta_i' \end{pmatrix} \begin{pmatrix} \Delta_{\varepsilon ui}^+ \\ \Delta_{\eta u}^+ \end{pmatrix} \\ &= \frac{1}{T} x_i' M_{F^0} \Delta b_i \Omega_{bi}^{-1} \Omega_{bui} + \left( \Delta_{\varepsilon ui}^+ - \delta_i^{0'} \Delta_{\eta u}^+ \right), \\ \delta_i^0 &= (F^{0'} F^0)^{-1} F^{0'} x_i, \quad \Delta b_i = \begin{pmatrix} \Delta x_i & \Delta F^0 \end{pmatrix} \end{aligned}$$

$$\Delta_{bui}^+ = \begin{pmatrix} \Delta_{\varepsilon ui}^+ \\ \Delta_{\eta u}^+ \end{pmatrix} = \begin{pmatrix} \Delta_{bui} & \Delta_{bi} \end{pmatrix} \begin{pmatrix} I_k \\ -\Omega_{bi}^{-1} \Omega_{bui} \end{pmatrix} = \Delta_{bui} - \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui}.$$

The estimator is  $\sqrt{nT}$  consistent if  $\phi_{nT}^0 = 0$ , which occurs when  $x_{it}$  is strictly exogenous. Otherwise, the estimator is  $T$  consistent as there is an asymptotic bias given by the term  $\sqrt{n}\phi_{nT}^0$ . This is an average of individual biases that are data specific as seen from the definition of  $\theta_i$ . The individual biases arise from the contemporaneous and low frequency correlations between the regression error and the innovations of the I(1) regressors as given by terms such as  $\Omega_{bui}$  and  $\Delta_{bui}$ .

We use the fully modified procedure of Phillips and Hansen (1990) to remove the bias. For this purpose, we need the kernel method to consistently estimate the nuisance parameters. Consider

$$\begin{aligned}\widehat{\Omega}_i &= \sum_{j=T+1}^{T-1} w\left(\frac{j}{K}\right) \widehat{\Gamma}_i(j), \\ \widehat{\Delta}_i &= \sum_{j=0}^{T-1} w\left(\frac{j}{K}\right) \widehat{\Gamma}_i(j) \\ \widehat{\Gamma}_i(j) &= \frac{1}{T} \sum_{t=1}^{T-j} \widehat{w}_{it+j} \widehat{w}'_{it}.\end{aligned}$$

where  $\widehat{w}_{it} = (\widehat{u}_{it}, \Delta x'_{it}, \Delta F_t^{0'})'$ . To state the asymptotic theory for the fully modified estimator, we need the following assumption, as used in Moon and Perron (2004):

**Assumption 6** (a)  $\liminf_{n, T \rightarrow \infty} (\log T / \log n) > 1$ .

(b) the kernel function  $w(\cdot) : \mathbb{R} \rightarrow [-1, 1]$  satisfies (i)  $w(0) = 1$ ,  $w(x) = w(-x)$ , (ii)  $\int_{-1}^1 w(x)^2 dx < \infty$  and with Parzen's exponent  $q \in (0, \infty)$  such that  $\lim_{|x| \rightarrow 0} \frac{1-w(x)}{|x|^q} < \infty$ .

(c) The bandwidth parameter  $K$  satisfies  $K \sim n^b \frac{1}{2q} < b < \liminf \frac{\log T}{\log n} - 1$ .

Let

$$\widehat{\phi}_{nT}^0 = \left[ \frac{1}{nT^2} \sum_{i=1}^n x'_i M_{F^0} x_i \right]^{-1} \widehat{\theta}^n$$

where  $\widehat{\theta}^n = \frac{1}{n} \sum_{i=1}^n \widehat{\theta}_i$ ,  $\widehat{\theta}_i$  is a consistent estimate of  $\theta_i^0$ . The resulting bias-corrected estimator is

$$\widetilde{\beta}_{LSBC} = \widetilde{\beta}_{LS} - \frac{1}{T} \widehat{\phi}_{nT}^0. \quad (9)$$

This is the fully-modified estimator of Phillips and Hansen (1990); also see Phillips and Moon (1999) and Bai and Kao (2005) for the panel version of the fully modified estimator. The estimator can be rewritten in the following format. Define

$$y_{it}^+ = y_{it} - \Omega_{ubi} \Omega_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix} \quad u_{it}^+ = u_{it} - \Omega_{ubi} \Omega_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix}$$

By construction  $u_{it}^+$  has a zero long-run covariance with  $(\Delta x'_{it} \quad \Delta F_t^{0'})'$  and hence the endogeneity can be removed. Furthermore, nuisance parameters arising from the low frequency correlation of the errors are summarized in  $\Delta_{bui}^+$ . Let

$$\widetilde{\beta}_{LSFM} = \left( \sum_{i=1}^n x'_i M_{F^0} x_i \right)^{-1} \sum_{i=1}^n \left( x'_i M_{F^0} \widetilde{y}_i^+ - T \left( \widetilde{\Delta}_{\varepsilon ui}^+ - \delta_i^{0'} \widetilde{\Delta}_{\eta u}^+ \right) \right) \quad (10)$$

where  $\tilde{y}^+$  and  $\tilde{\Delta}^+$  are consistent estimates of  $y^+$  and  $\Delta^+$  etc. It is not difficult to verify that  $\hat{\beta}_{LSBC}$  and  $\hat{\beta}_{LSFM}$  are identical. The estimator extends the fully-modified estimator of Phillips and Moon (1995) to the situation in which common stochastic trends are present.

**Proposition 2** *Let  $\tilde{\beta}_{LSFM}$  be defined by (10). Under Assumptions 1-6,*

$$\sqrt{nT}(\tilde{\beta}_{LSFM} - \beta) \xrightarrow{d} N(0, \Sigma).$$

In small scale cointegrated systems, cointegrated vectors are  $T$  consistent, and this fast rate of convergence is already accelerated relative to the case of stationary regressions, which is  $\sqrt{T}$ . Here in a panel data context with observed global stochastic trends, the estimates converge to the true values at an even faster rate of  $\sqrt{nT}$  and the limiting distributions are normal. To take advantage of this fast convergence rate made possible by large panels, we need to deal with the fact that  $F$  is not observed. This problem is considered in the next two subsections.

### 3.2 The Cup Estimator

The LSFM considered above is a linear estimator and can be obtained if  $F$  is observed. When  $F$  is not observed, the previous estimator is infeasible. Recall that least squares estimator that ignore  $F$  is, in general, inconsistent. In this section, we consider estimating  $F$  along with  $\beta$  and  $\Lambda$  by minimizing the objective function

$$S_{nT}(\beta, F, \Lambda) = \sum_{i=1}^n (y - x_i\beta - F\lambda_i)' (y - x_i\beta - F\lambda_i) \quad (11)$$

subject to the constraint  $T^{-2}F'F = I_r$  and  $\Lambda'\Lambda$  being diagonal. The least squares estimator for  $\beta$  for a given  $F$  is

$$\hat{\beta} = \left( \sum_{i=1}^n x_i' M_F x_i \right)^{-1} \sum_{i=1}^n x_i' M_F y_i.$$

Define

$$\begin{aligned} w_i &= y_i - x_i\beta \\ &= F\lambda_i + u_i. \end{aligned}$$

Notice that given  $\beta$ ,  $w_i$  has a pure factor structure. Let  $W = (w_1, \dots, w_n)$  be a  $T \times n$  matrix. We can rewrite the objective function (11) as  $tr[(W - F\Lambda')(W - F\Lambda)']$ . If we concentrate out  $\Lambda = W'F(F'F)^{-1} = T^{-2}W'F$ , we have the concentrated objective function:

$$tr(W' M_F W) = tr(W'W) - tr(F'WW'F/T^2). \quad (12)$$

Since the first term does not depend on  $F$ , minimizing (12) with respect to  $F$  is equivalent to maximizing  $\text{tr} \left( T^{-2} F' W W' F \right)$  subject to the constraint  $T^{-2} F' F = I_r$ . The solution, denoted  $\widehat{F}$ , is a matrix of the first  $r$  eigenvectors (multiplied by  $T$ ) of the matrix  $\frac{1}{nT^2} \sum_{i=1}^n (y_i - x_i \beta) (y_i - x_i \beta)'$ .

Although  $F$  is not observed when estimating  $\beta$ , and similarly,  $\beta$  is not observed when estimating  $F$ , we can replace the unobserved quantities by initial estimates and iterate until convergence. Such a solution is more easily seen if we rewrite the left hand side of (12) with  $y - x\beta$  substituting in for  $W$ . Define

$$S_{nT}(\beta, F) = \frac{1}{nT^2} \sum_{i=1}^n (y_i - x_i \beta)' M_F (y_i - x_i \beta)$$

The continuous updated estimator (Cup) for  $(\beta, F)$  is defined as

$$\left( \widehat{\beta}_{Cup}, \widehat{F}_{Cup} \right) = \underset{\beta, F}{\text{argmin}} S_{nT}(\beta, F).$$

More precisely,  $(\widehat{\beta}_{Cup}, \widehat{F}_{Cup})$  is the solution to the following two nonlinear equations

$$\widehat{\beta} = \left( \sum_{i=1}^n x_i' M_{\widehat{F}} x_i \right)^{-1} \sum_{i=1}^n x_i' M_{\widehat{F}} y_i \quad (13)$$

$$\widehat{F} V_{nT} = \left[ \frac{1}{nT^2} \sum_{i=1}^n (y_i - x_i \widehat{\beta}) (y_i - x_i \widehat{\beta})' \right] \widehat{F} \quad (14)$$

where  $M_{\widehat{F}} = I_T - T^{-2} \widehat{F} \widehat{F}'$  since  $\widehat{F}' \widehat{F} / T^2 = I_r$ , and  $V_{nT}$  is a diagonal matrix consisting of the  $r$  largest eigenvalues of the matrix inside the brackets, arranged in decreasing order. Note that the estimator is obtained by iteratively solving for  $\widehat{\beta}$  and  $\widehat{F}$  using (13) and (14). It is a non-linear estimator even though linear least squares estimation is involved at each iteration. An estimate of  $\Lambda$  can be obtained as:

$$\widehat{\Lambda} = T^{-2} \widehat{F}' (Y - X \widehat{\beta}).$$

The triplet  $(\widehat{\beta}, \widehat{F}, \widehat{\Lambda})$  jointly minimizes the objective function (11).

The estimator  $\widehat{\beta}_{Cup}$  is consistent for  $\beta$ . We state this result in the following proposition.

**Proposition 3** *Under Assumptions 1-4, as  $(n, T) \rightarrow \infty$ ,*

$$\widehat{\beta}_{Cup} \xrightarrow{p} \beta^0.$$

Given consistency, we can further establish an asymptotic representation for  $\widehat{\beta}_{Cup}$ .

**Proposition 4** *Suppose Assumptions 1-4 hold and  $(n, T) \rightarrow \infty$ , then*

$$\sqrt{nT} \left( \widehat{\beta}_{Cup} - \beta^0 \right) = D (F^0)^{-1} \left[ \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{F^0} - \frac{1}{n} \sum_{k=1}^n a_{ik} x_i' M_{F^0} \right) u_i \right] + o_p(1),$$

where  $a_{ik} = \lambda'_i \left( \frac{\Lambda' \Lambda}{n} \right) \lambda_k$ ,  $D(F^0) = \frac{1}{nT^2} \sum_{i=1}^n Z'_i Z_i$  and  $Z_i = M_{F^0} x_i - \frac{1}{n} \sum_{k=1}^n M_{F^0} x_k a_{ik}$ .

In comparison with the pooled least squares estimator for the case of known  $F^0$ , estimating the stochastic trends clearly affects the limiting behavior of the estimator. The term involving  $a_{ik}$  is due to the estimation of  $F$ . The extra term is carried over to the limiting distribution and to the asymptotic bias, as shown in Theorem 1 below. Let  $\bar{w}_{it} = \left( u_{it}, \Delta \bar{x}'_i, \eta_t \right)'$  where  $\bar{x}_i = x_i - \frac{1}{n} \sum_{k=1}^n x_k a_{ik}$  and  $a_{ik} = \lambda'_i \left( \frac{\Lambda' \Lambda}{n} \right) \lambda_k$ . Define  $\bar{\Omega}_i$  to be the long run covariance matrix of  $\bar{w}_{it}$  as in (5), and define  $\bar{\Delta}_i$  similarly as in (6). These quantities depend on  $n$ , but this dependence is suppressed for notional simplicity. For the rest of the paper, we use bar to denote those long run covariance matrices (including one sided and conditional covariances and so on) are generated from  $\bar{w}_{it}$  instead of  $w_{it}$ .

Because the right hand side of the representation does not depend on estimated quantities, it is not difficult to derive the limiting distribution of  $\hat{\beta}_{Cup}$ .

**Theorem 1** *Suppose that assumptions 1-5 hold. Let  $\hat{\beta}_{Cup}$  be obtained by iteratively updating (13) and (14). As  $(n, T) \rightarrow \infty$  with  $n/T \rightarrow 0$ , we have*

$$\sqrt{nT} \left( \hat{\beta}_{Cup} - \beta \right) - \sqrt{n} \phi_{nT} \xrightarrow{d} N(0, \Sigma)$$

where

$$\begin{aligned} \phi_{nT} &= \left[ \frac{1}{nT^2} \sum_{i=1}^n Z'_i Z_i \right]^{-1} \left( \frac{1}{n} \sum_{i=1}^n \theta_i \right) \\ \theta_i &= \frac{1}{T} Z'_i \Delta \bar{b}_i \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \left( \bar{\Delta}_{\varepsilon ui}^+ - \bar{\delta}'_i \bar{\Delta}_{\eta u}^+ \right), \\ \Sigma &= D_Z^{-1} \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R'_{ni} \right) \right] D_Z^{-1}, \quad D_Z = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int R_{ni} R'_{ni} \right), \quad (15) \\ R_{ni} &= Q_i - \frac{1}{n} \sum_{k=1}^n Q_k a_{ik}, \\ \Delta \bar{b}_i &= \left( \Delta \bar{x}_i \quad \Delta F^0 \right), \\ \bar{x}_i &= x_i - \frac{1}{n} \sum_{k=1}^n x_k a_{ik}, \\ \bar{\delta}_i &= \delta_i - \frac{1}{n} \sum_{k=1}^n \delta_k a_{ik}. \end{aligned}$$

Theorem 1 establishes the large sample properties of the Cup estimator. As mentioned earlier, the  $a_{ik}$  term arises from having to estimate  $F_t$ . In consequence, the bias is now a function of terms not present in Proposition 1, which is valid when  $F_t$  is observed. Since  $\phi_{nT} = O_p(1)$ , the Cup estimator is also  $T$  consistent. This is in contrast with pooled OLS in Section 2, where it was shown to be inconsistent in general. Nevertheless, as in the case when  $F$  is observed, the Cup estimator has an asymptotic bias and thus the limiting distribution is not centered around zero. This motivates removing the bias by constructing a consistent estimate of  $\phi_{nT}$ . This can be obtained upon replacing  $F^0$ ,  $\Delta\bar{b}_i$ ,  $\bar{\Omega}_{bi}$ ,  $\bar{\Omega}_{bui}$ ,  $\bar{\Delta}_{\varepsilon ui}^+$ ,  $\bar{\Delta}_{\eta u}^+$  by their consistent estimates.

We consider two fully-modified estimators. The first one directly corrects the bias of  $\hat{\beta}_{Cup}$ , and is denoted by  $\hat{\beta}_{CupBC}$ . The second one will be considered in the next subsection, where correction is made during each iteration, and will be denoted by  $\hat{\beta}_{CupFM}$ .

Consider

$$\begin{aligned}\hat{\Omega}_i &= \sum_{j=T+1}^{T-1} w\left(\frac{j}{K}\right) \hat{\Gamma}_i(j), \\ \hat{\Delta}_i &= \sum_{j=0}^{T-1} w\left(\frac{j}{K}\right) \hat{\Gamma}_i(j) \\ \hat{\Gamma}_i(j) &= \frac{1}{T} \sum_{t=1}^{T-j} \hat{w}_{it+j} \hat{w}'_{it}.\end{aligned}$$

where

$$\hat{w}_{it} = (\hat{u}_{it}, \Delta\hat{x}'_{it}, \Delta\hat{F}'_t)' \quad \text{with} \quad \Delta\hat{x}_{it} = \Delta x_{it} - \frac{1}{n} \sum_{k=1}^n \Delta x_{kt} \hat{a}_{ik}$$

The bias-corrected Cup estimator is defined as

$$\hat{\beta}_{CupBC} = \hat{\beta}_{Cup} - \frac{1}{T} \hat{\phi}_{nT}$$

where

$$\begin{aligned}\hat{\phi}_{nT} &= \left[ \frac{1}{nT^2} \sum_{i=1}^n \hat{Z}'_i \hat{Z}_i \right]^{-1} \left( \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i \right) \\ \hat{\theta}_i &= \hat{Z}'_i \Delta\hat{b}_i \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui} + \left( \hat{\Delta}_{\varepsilon ui}^+ \quad -\hat{\delta}'_i \quad \hat{\Delta}_{\eta u}^+ \right), \\ \hat{\delta}_i &= \left( \hat{F}' \hat{F} \right)^{-1} \hat{F}' \hat{x}_i \quad \Delta\hat{b}_i = \left( \Delta\hat{x}_i \quad \Delta\hat{F} \right) \\ \hat{x}_i &= x_i - \frac{1}{n} \sum_{k=1}^n x_{kt} \hat{a}_{ik}, \quad \hat{a}_{ik} = \hat{\lambda}'_i \left( \hat{\Lambda}' \hat{\Lambda} / n \right)^{-1} \hat{\lambda}_k.\end{aligned}$$

The bias corrected estimator is also a fully modified estimator, and in fact, it can be rewritten as (16) below. This will be discussed further in the next subsection.

**Theorem 2** *Assume Assumptions 1-6 hold. Then as  $(n, T) \rightarrow \infty$  with  $n/T \rightarrow 0$ ,*

$$\sqrt{nT} \left( \widehat{\beta}_{CupBC} - \beta^0 \right) \xrightarrow{d} N(0, \Sigma).$$

The CupBC is  $\sqrt{nT}$  consistent with a limiting distribution that is centered at zero as long as  $(n, T) \rightarrow \infty$  and  $\frac{n}{T} \rightarrow 0$ . This type of bias correction approach is also used in Hahn and Kuersteiner (2002), for example, and is not uncommon in panel data analysis. Because the bias-corrected estimator is  $\sqrt{nT}$  and has a normal limit distribution, the usual  $t$  and Wald tests can be used for inference. Note that the limiting distribution is different from that of the infeasible LFSM estimator, implying that the estimation of  $F$  does affect the limiting distribution.

### 3.3 Fully Modified Cup Estimator

The CupBC just considered is constructed by estimating the asymptotic bias of  $\widehat{\beta}_{Cup}$ , and then subtracting it from  $\widehat{\beta}_{Cup}$ . In this subsection, we consider a different fully-modified estimator, denoted by  $\widehat{\beta}_{CupFM}$ . Let

$$\begin{aligned} y_{it}^+ &= y_{it} - \widehat{\Omega}_{ubi} \widehat{\Omega}_{bi}^{-1} \begin{pmatrix} \Delta \widehat{x}_{it} \\ \Delta \widehat{F}_t \end{pmatrix} \\ \widehat{\delta}_i &= \left( \widehat{F}' \widehat{F} \right)^{-1} \widehat{F}' \widehat{x}_i. \end{aligned}$$

where  $\widehat{\Omega}_{ubi}$ ,  $\widehat{\Omega}_{bi}$ , and  $\widehat{\Delta}_{bui}$  are estimates of  $\bar{\Omega}_{ubi}$ ,  $\bar{\Omega}_{bi}$  and  $\bar{\Delta}_{bui}$ , respectively. Recall that  $\widehat{\beta}_{Cup}$  is obtained by jointly solving (13) and (14). Consider replacing these equations by the following:

$$\widehat{\beta}_{CupFM} = \left( \sum_{i=1}^n x_i' M_{\widehat{F}} x_i \right)^{-1} \sum_{i=1}^n \left( x_i' M_{\widehat{F}} y_i^+ - T \left( \widehat{\Delta}_{\varepsilon ui}^+ - \widehat{\delta}_i' \widehat{\Delta}_{\eta u}^+ \right) \right) \quad (16)$$

$$\widehat{F} V_{nT} = \left[ \frac{1}{nT^2} \sum_{i=1}^n \left( y_i - x_i \widehat{\beta}_{CupFM} \right) \left( y_i - x_i \widehat{\beta}_{CupFM} \right)' \right] \widehat{F} \quad (17)$$

Like the FM estimator of Phillips and Hansen (1990), the corrections are made to the data to remove serial correlation and endogeneity. The CupFM estimator for  $(\beta, F)$  is obtained by iteratively solving (16) and (17). Thus correction to endogeneity and serial correlation is made during each iteration.

**Theorem 3** *Assume Assumptions 1-6 hold. Then as  $(n, T) \rightarrow \infty$  with  $n/T \rightarrow 0$ ,*

$$\sqrt{nT} \left( \widehat{\beta}_{CupFM} - \beta^0 \right) \xrightarrow{d} N(0, \Sigma),$$

where  $\Sigma$  is given in (15)

The CupFM and CupBC have the same asymptotic distribution, but they are constructed differently. The estimator  $\hat{\beta}_{CupBC}$  does the bias correction only once, i.e., at the final stage of the iteration, and  $\hat{\beta}_{CupFM}$  does the correction at every iteration. Thus  $\hat{\beta}_{CupFM}$  is computationally more costly but may have better finite sample properties. The situation is different from the case of known  $F$ , in which the bias-corrected estimator and the fully-modified estimator are identical due to the absence of iteration.

The preceding results assume that the number of stochastic trends,  $r$ , is known. If this is not the case,  $r$  can be consistently estimated using the information criterion function developed in Bai and Ng (2002). In particular, let

$$\hat{r} = \arg \min_{1 \leq r \leq r_{\max}} IC(r)$$

where  $r \leq r_{\max}$ ,  $r_{\max}$  is a bounded integer and

$$IC_1(r) = \hat{\sigma}^2(r) + r\hat{\sigma}^2(r_{\max})g_{nT}$$

where  $g_{nT} \rightarrow 0$  as  $n, T \rightarrow \infty$  and  $\min[n, T]g_{nT} \rightarrow \infty$ . For example,  $g_{nT}$  can be  $\log(a_{nT})/a_{nT}$ , with  $a_{nT} = \frac{nT}{n+T}$ .

#### 4 Test of Cross-Sectional Dependence

The results in this paper can be used to test the null hypothesis of no cross-sectional dependence

$$H_0 : \lambda_i = 0 \quad \text{for all } i \tag{18}$$

for all  $i$  in (3) against the alternative that <sup>6</sup>

$$H_0 : \lambda_i \neq 0 \quad \text{for some } i.$$

For each  $i$ , let  $RSS_{1i} = \sum_{t=1}^T \hat{w}_{it}^2$  be the sum of squared residuals from the restricted model:

$$\hat{y}_{it} = x'_{it}\hat{\beta}_{FM} + \hat{w}_{it}$$

where  $\hat{\beta}_{FM}$  is the FM estimator of Phillips and Moon (1999). Also let  $RSS_{2i} = \sum_{t=1}^T \hat{u}_{it}^2$  be the sum of squared residuals from the unrestricted model:

$$\hat{y}_{it} = x'_{it}\hat{\beta}_{CupFM} + \hat{\lambda}'_i\hat{F}_t + \hat{u}_{it}$$

Let  $\hat{\Omega}_{u,\varepsilon i}$  is a consistent estimate of the long run variance,  $\Omega_{u,\varepsilon i}$ . Define

$$J_i = \frac{RSS_{1i} - RSS_{2i}}{\hat{\Omega}_{u,\varepsilon i}}. \tag{19}$$

---

<sup>6</sup>We thank Joon Park for this suggestion.

The  $J_i$  is similar to the variable addition test for cointegration in Park (1990). By Theorem 4.1 in Park (1990), the  $J_i$  statistic has a limiting  $\chi^2$  distribution as  $T \rightarrow \infty$  with degree of freedom equal to  $r$ , under the null hypothesis of  $r$  common factors. Our proposed test is based on averaging the individual  $J_i$  as follows:

$$J = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{J_i - r}{\sqrt{2r}}$$

It can be shown<sup>7</sup> that as  $(n, T) \rightarrow \infty$

$$J \xrightarrow{d} N(0, 1)$$

The result follows because  $E(J_i) = r + O(\frac{1}{T})$  and  $Var(J_i) = 2r + O(\frac{1}{T})$  for each  $i$ .

## 5 Further issues

The preceding analysis assumes that there are no deterministic components and that the regressors and the common factors are all I(1) without drifts. This section considers construction of the estimator when these restrictions are relaxed. It will be shown that when there are deterministic components, we can apply the same estimation procedure to the demeaned or detrended series provided the Brownian motion processes in the limiting distribution are replaced by the demeaned and/or detrended versions. Furthermore, the procedure is robust to the presence of mixed I(1)/I(0) regressors and/or factors. Of course, the convergence rates for I(0) and I(1) regressors will be different, but asymptotic normality and the construction of test statistics (and their limiting distribution) do not depend on the convergence rate.

### 5.1 Incidental trends

The Cup estimator can be easily extended to models with incidental trends,

$$y_{it} = \alpha_i + \rho_i t + x'_{it}\beta + \lambda'_i F_t + u_{it}. \quad (20)$$

In the intercept only case ( $\rho_i = 0$ , for all  $i$ ), we define the projection matrix

$$M_T = I_T - \nu_T \nu_T' / T$$

where  $\nu_T$  is a vector of 1's. When a linear trend is also included in the estimation, we define  $M_T$  to be the projection matrix orthogonal to  $\nu_T$  and to the linear trend. Then

$$M_T y_i = M_T x_i \beta + M_T F \lambda_i + M_T u_i,$$

---

<sup>7</sup>See e.g., similar to Theorem 3 in Pesaran and Yamagata (2006)

or

$$\dot{y}_i = \dot{x}_i\beta + \dot{F}_t\lambda_i + \dot{u}_i$$

where the dotted variables are demeaned and/or detrended versions. The estimation procedure for the cup estimator is identical to that of Section 3, except that we use dotted variables.

With the intercept only case, the construction of FM estimator is also the same as before. Theorems 1-3 hold with the following modification for the limiting distribution. The random processes  $B_{\varepsilon,i}$  and  $B_\eta$  in  $Q_i$  are replaced by the demeaned Brownian motions.

When linear trends allowed,  $\Delta x_{it}$  is now replaced by demeaned version of  $\Delta x_{it}$ , i.e.,  $\hat{\varepsilon}_{it} = \Delta x_{it} - \overline{\Delta x_{it}}$ , which is detrended residual of  $x_{it}$ . We then use  $\hat{\varepsilon}_{it} - \frac{1}{n} \sum_{k=1}^n \hat{\varepsilon}_{kt} \hat{a}_{ik}$  in place of  $\Delta \hat{x}_{it}$  in Section 3. Similarly,  $\hat{F}$  should also be detrended. More specifically, we use  $\hat{\eta}_t = \Delta \hat{F}_t - \overline{\Delta \hat{F}_t}$  in place of  $\Delta \hat{F}_t$ . Alternatively, since  $\dot{x}_i$  is an already detrended series, and  $\dot{F}$  is also asymptotically detrended (since it is estimating  $\dot{F}$ ),  $\Delta \dot{x}_{it}$  and  $\Delta \dot{F}_t$  are also estimating the detrended residuals. Thus we can simply apply the same procedure prescribed in Section 3 with the dotted variables. The limiting distribution in Theorems 2 and consequently in Theorem 3 is modified as follows. The random processes  $B_{\varepsilon,i}$  and  $B_\eta$  are replaced by the demeaned and detrended Brownian motions.

In either case, the test statistics (t and  $\chi^2$ ) have standard asymptotic distribution, not depending on whether the underlying Brownian motion is demeaned or detrended.

When linear trends are included in the estimation, the limiting distribution is invariant to whether or not  $y_{it}$ ,  $x_{it}$  and  $F_t$  contain a linear trend. Now suppose that these variables do contain a linear trend (drifted random walks). With deterministic cointegration holding (i.e., cointegrating vector eliminates the trends), the estimated  $\beta$  will have a faster convergence rate when a separate linear trend is not included in the estimation. But we do not consider this case. Interested readers are referred to Hansen (1992).

## 5.2 Mixed I(0)/I(1) Regressors and Common Shocks

So far, we have considered estimation of panel cointegration models when all the regressors and common shocks are I(1). There are no stationary regressors or stationary common shocks. In this section we suggest that the results are robust to mixed I(1)/I(0) regressors and mixed I(1)/I(0) common shocks. Below, we sketch the arguments for the LS estimator assuming the factors are observed. If they are not observed, the limiting distribution is different, but the idea of argument is the same.

Recall that the LS estimator is  $\hat{\beta}_{LS} = (\sum_{i=1}^n x_i' M_{F^0} x_i)^{-1} \sum_{i=1}^n x_i' M_{F^0} y_i$ . The term

$$M_{F^0} x_i = (I_T - F^0 (F^{0'} F^0)^{-1} F^{0'}) x_i = x_i - F^0 \delta_i$$

with  $\delta_i = (F^{0'} F^0)^{-1} F^{0'} x_i$  plays an important role in the properties of the LS. When  $x_{it}$  and  $F_t$  are I(1),  $\delta_i = O_p(1)$  and thus

$$\frac{(M_{F^0} x_i)_t}{\sqrt{T}} = \frac{x_{it}}{\sqrt{T}} - \frac{\delta_i' F_t^0}{\sqrt{T}} = O_p(1).$$

We now consider this term under alternative assumptions.

**I(1) Regressors, I(0) Factors** Suppose all regressors are I(1) and all common shocks are I(0). With I(0) factors, we have  $T^{-1} F^{0'} F^0 \xrightarrow{p} \Sigma_F = O_p(1)$ . Thus

$$\delta_i = \left( T^{-1} F^{0'} F^0 \right)^{-1} \frac{1}{T} \sum_{t=1}^T F_t^0 x'_{it} \xrightarrow{d} \Sigma_F^{-1} \int dB_\eta B'_{\varepsilon i} = O_p(1).$$

It follows that

$$\frac{(M_{F^0} x_i)_t}{\sqrt{T}} = \frac{x_{it} - \delta_i' F_t^0}{\sqrt{T}} = \frac{x_{it}}{\sqrt{T}} + o_p(1)$$

and  $\frac{x_{it}}{\sqrt{T}} \xrightarrow{d} B_{\varepsilon i}$  as  $T \rightarrow \infty$ . The limiting distribution of the LS when the factors are I(0) is the same as when all factors are I(1), except that  $Q_i$  is now asymptotically the same as  $B_{\varepsilon i}$ . For the FM, observe that the submatrix  $\Omega_\eta$  in

$$\Omega_{bi} = \begin{bmatrix} \Omega_{\varepsilon i} & \Omega_{\varepsilon \eta i} \\ \Omega_{\eta \varepsilon i} & \Omega_\eta \end{bmatrix}$$

is a zero matrix since  $\eta = \Delta F_t^0$  is an  $I(-1)$  process and has zero long-run variance. Similarly,  $\Omega_{\varepsilon \eta i}$  is also zero. The submatrix  $\Omega_{u \eta i}$  in  $\Omega_{u.bi} = \Omega_{ui} - \Omega_{ubi} \Omega_{bi}^{-1} \Omega_{bui}$  as well as the submatrices  $(\Delta_{\eta ui} \quad \Delta_{\eta i})$  in  $(\Delta_{bui} \quad \Delta_{bi})$  are also degenerate because of the factors are I(0). Note that  $\Omega_{bi}$  is not invertible. Under appropriate choice of bandwidth, see Phillips (1995),  $\Omega_{bi}^{-1} \Omega_{bui}$  can be consistently estimated, so that FM estimators can be constructed. This argument treats  $F_t$  as if it were I(1). If it is known that  $F_t$  is I(0), we will simply use  $F_t$  instead of  $\Delta F_t$  in the FM construction. This case is analyzed by Bai and Kao (2005).

**I(1) Regressors, Mixed I(0)/I(1) Factors** Consider the model

$$y_{it} = x'_{it} \beta + \lambda'_{1i} F_{1t} + \lambda'_{2i} F_{2t} + u_{it} \quad (21)$$

where  $F_{1t} = \eta_{1t}$  is  $r_1 \times 1$  and  $\Delta F_{2t} = \eta_{2t}$  is  $r_2 \times 1$ . We again have  $M_{F^0} x_i = x_i - F^0 \delta_i$  but  $\delta_i = [\delta_{1i} \quad \delta_{2i}]'$ . Then

$$\begin{aligned} \frac{(M_{F^0} x_i)_t}{\sqrt{T}} &= \frac{x_{it}}{\sqrt{T}} - \frac{1}{\sqrt{T}} \begin{bmatrix} \delta'_{1i} & \delta'_{2i} \end{bmatrix} \begin{bmatrix} F_{1t}^0 \\ F_{2t}^0 \end{bmatrix} = \frac{x_{it}}{\sqrt{T}} - \frac{1}{\sqrt{T}} \left( \delta'_{1i} F_{1t}^0 + \delta'_{2i} F_{2t}^0 \right) \\ &= \frac{x_{it}}{\sqrt{T}} - \frac{\delta'_{2i} F_{2t}^0}{\sqrt{T}} + o_p(1) \end{aligned}$$

since  $\delta_{1i} = O_p(1)$ ,  $\delta_{2i} = O_p(1)$  but  $\frac{F_{1t}^0}{\sqrt{T}} = o_p(1)$ . The random matrix  $Q_i$  involves  $B_{\varepsilon i}$  and  $B_{2\eta}$ . In the FM correction, the long run variance  $(u_{it}, \Delta x'_{it}, \Delta F'_{1t}, \Delta F'_{2t})'$  is degenerate. With an appropriate choice of bandwidth as in Phillips (1995), the limiting normality still holds.

**Mixed I(1)/I(0) Regressors and I(1) Factors** Suppose  $k_2$  regressors denoted by  $x_{2it}$  are I(1), and  $k_1$  regressors denoted by  $x_{1it}$  are I(0). Assume  $F_t$  is I(1) and  $u_{it}$  is I(0) as in (3). Consider

$$\begin{aligned} y_{it} &= \alpha_i + x'_{1it}\beta_1 + x'_{2it}\beta_2 + \lambda'_i F_t + u_{it} \\ \Delta x_{2it} &= \varepsilon_{2it}. \end{aligned}$$

With the inclusion of an intercept, there is no loss of generality to assume  $x_{1it}$  having a zero mean. For this model, we add the assumption that

$$E(x_{1it}u_{it}) = 0 \tag{22}$$

to rule out simultaneity bias with I(0) regressors. Otherwise  $\beta_1$  cannot be consistently estimated. Alternatively, if  $u_{it}$  is correlated with  $x_{1it}$ , we can project  $u_{it}$  onto  $x_{1it}$  to obtain the projection residual and still denote it by  $u_{it}$  (with abuse of notation), and by definition,  $u_{it}$  is uncorrelated with  $x_{1it}$ . But then  $\beta_1$  is no longer the structural parameter. The dynamic least squares approach by adding  $\Delta x_{2it}$  is exactly based on this argument, with the purpose of more efficient estimation of  $\beta_2$ .

If one knows which variable is I(0) and which is I(1), the situation is very simple. The I(1) and I(0) variables are asymptotically orthogonal, we can separately analyze the distribution of estimated  $\beta_1$  and  $\beta_2$ . The estimated  $\beta_1$  needs no correction and is asymptotically normal, and the estimated  $\beta_2$  has a distribution as if there is no I(0) regressors except the intercept. Note that FM construction for  $\hat{\beta}_2$  is based on the residuals with all regressors included. The rest of analysis is identical to the situation of all I(1) regressors with an intercept.

In practice, the separation of I(0) or I(1) regressors may not be known in advance. One can proceed with a pretesting to identify the integration order for each variable, and then apply the above argument. One major purpose of separating I(0) and I(1) variables is to derive relevant rate of convergence for the estimated parameters. But if the ultimate purpose is to do hypothesis testing, there is no need to know the rate of convergence for the estimator since the scaling factor  $n$  or  $T$  are cancelled out in the end. One can proceed as if all regressors are I(1). Then care should be taken since the long-run covariance matrix is of deficient rank. Phillips (1995) shows that FM estimators can be constructed with appropriate choice of bandwidth. Interested readers are referred to Phillips (1995) for details.

Finally, there is the case of mixed I(1)/I(0) regressors and mixed I(1)/I(0) factors. As explained earlier, I(0) factors do not change the result. Also, in actual computation, there is no need to know whether  $F^0$  is I(1) and I(0), since the Cup estimator only depends on  $M_{\hat{F}}$ , scaling in  $\hat{F}$  does not alter the numerical value of  $\hat{\beta}_{Cup}$ . The important implication is that researchers can use the CupBC and CupFM without pre-testing the regressors for I(0) or I(1).

## 6 Monte Carlo Simulations

In this section, we conduct Monte Carlo experiments to assess the finite sample properties of the proposed CupFM estimator. Random numbers for error terms,  $(u_{it}, \varepsilon_{it}, \eta_t)$  are generated by the GAUSS procedure RNDNS. At each replication, we generate an  $nT$  length of random numbers and then split it into  $n$  series so that each series has the same mean and variance. The variables  $\{u_{it}\}$ ,  $\{\varepsilon_{it}\}$  and  $\{\eta_t\}$  are constructed with  $u_{i0} = 0$ ,  $\varepsilon_{i0} = 0$  and  $\eta_t = 0$ .

We compare the performance of the LSDV (least squares dummy variables, i.e., the within group estimator), 2sFM (2-stage fully modified), and CupFM estimators by conducting Monte Carlo experiments based on the following design. For  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ ,

$$\begin{aligned} y_{it} &= 2x_{it} + c \left( \lambda_i' F_t \right) + u_{it} \\ F_t &= F_{t-1} + \eta_t, \\ x_{it} &= x_{it-1} + \varepsilon_{it} \end{aligned}$$

where

$$\begin{pmatrix} u_{it} \\ \varepsilon_{it} \\ \eta_t \end{pmatrix} \stackrel{iid}{\sim} N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & 1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & 1 \end{bmatrix} \right). \quad (23)$$

For this experiment, we assume a single factor, i.e.,  $r = 1$ ,  $\lambda_i$  and  $\eta_t$  are generated from i.i.d.  $N(\mu_\lambda, 1)$  and  $N(\mu_\eta, 1)$  respectively. We set  $\mu_\lambda = 2$  and  $\mu_\eta = 0$ . Endogeneity of the system under the design of (23) is controlled by only two parameters,  $\sigma_{21}$  and  $\sigma_{31}$ . We use  $c$  to control the importance of the global stochastic trends and let  $c = (5, 10)$ ,  $\sigma_{32} = 0.4$ ,  $\sigma_{21} = (0, 0.2, -0.2)$  and  $\sigma_{31} = (0, 0.8, -0.8)$ .

The estimate of the long-run covariance matrix is based on the procedure KERNEL in COINT 2.0 with a Bartlett window with the truncation set at five. Results with other kernels, such as Parzen and quadratic spectral kernels, are not reported, because no essential differences were found for most cases.

We report the finite-sample properties of (a) the LSDV estimator, (b) the 2sFM estimator which is similar to the CupFM estimator except the iteration only goes two steps, and (c) the CupFM

estimator  $\widehat{\beta}_{FM}$  in (16). The LSDV is obtained by assuming the cross-sectional independence and the usual fixed effect specification. The maximum number of the iteration for CupFM estimators is set to 20. The results we report are based on 1,000 replications.

Table 1 reports the means and standard deviations (in parentheses) of the LSDV, 2sFM and the CupFM for sample sizes  $T = n = (20, 40, 60, 120)$ . The bias of the LSDV estimator does not decrease as  $(n, T)$  increases in general. In terms of mean bias, the CupFM is distinctly superior to the LSDV and 2sFM estimators for all cases considered. The 2sFM estimator is less efficient than the CupFM estimator as shown from larger standard deviations.

To see how the properties of the estimator vary with  $n$  and  $T$ , Table 2 considers 16 different combinations for  $n$  and  $T$ , each ranging from 20 to 120. From Table 2, we see that the LSDV and 2sFM estimators become heavily biased when the common shock,  $\lambda_i' F_t$ , are magnified from  $c = 5$  and  $c = 10$ . On the other hand, the CupFM estimator is unaffected by the values of  $c$ . The results in Table 2 again indicate that the CupFM performs very well.

The properties of the  $t$ -statistic for testing  $\beta = \beta_0$ , are given in Table 3. Here, the LSDV  $t$ -statistic is the conventional  $t$ -statistic as reported by standard statistical packages. It is clear that LSDV  $t$ -statistics and 2sFM  $t$ -statistics diverge as  $(n, T)$  increases and they are not well approximated by a standard  $N(0,1)$  distribution. The CupFM  $t$ -statistic is well approximated by a standard  $N(0,1)$  except when  $\sigma_{21} = -0.2$ . The CupFM  $t$ -statistic is much closer to the standard normal density than the LSDV  $t$ -statistic and the 2sFM  $t$ -statistic for all cases.

Table 4 shows that both the LSDV  $t$ -statistic, the 2sFM  $t$ -statistics, and CupFM  $t$ -statistic become more negatively biased as the dimension of cross-section  $n$  increases. The large standard deviation of the 2sFM  $t$ -statistic in Tables 3-4 again indicate the poor performance of the 2sFM estimator. For the CupFM, the biases decrease rapidly and the standard errors converge to 1.0 as  $T$  increases.

## 7 Conclusion

This paper develops an asymptotic theory for a panel cointegration model with unobservable global stochastic trends. We derive the rate of convergence and the limiting distributions for CupBC and CupFM estimators. We show that these estimators are  $\sqrt{nT}$  consistent and this holds in spite of spuriousness induced by unobservable  $I(1)$  common shocks. A simulation study shows that the proposed CupFM estimator has good finite sample properties.

## Appendix

The proofs for Propositions 1 and 2 use standard arguments and are hence omitted. Propositions 3 and 4 are proved in Bai et al. (2006). Note that no restriction is placed between  $n$  and  $T$  here. In contrast, Bai (2005) considered stationary regressions with factor errors and required  $n/T$  to converge to zero in the presence of serial correlation in  $\varepsilon_{it}$ .

### Proof.

■

To derive the limiting distribution for  $\widehat{\beta}_{Cup}$ , we need the following lemma. Hereafter, we define  $\delta_{nT} = \min\{\sqrt{n}, T\}$ .

**Lemma 1** *Assume Assumptions 1-4 hold. Let  $Z_i = M_{F^0}x_i - \frac{1}{n} \sum_{k=1}^n M_{F^0}x_k a_{ik}$ . Then we have:*

(a) *As  $(n, T) \rightarrow \infty$*

$$\frac{1}{nT^2} \sum_{i=1}^n Z_i' Z_i \xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int R_{ni} R_{ni}' \right),$$

(b) *If  $\frac{n}{T} \rightarrow 0$ , and if  $M_{F^0}x_i$  and  $u_i$  are uncorrelated for each  $i$ , then*

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n Z_i' u_i \xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{ui} E \left( \int R_{ni} R_{ni}' \right) \right)$$

(c) *If  $\frac{n}{T} \rightarrow 0$ , and if  $M_{F^0}x_i$  and  $u_i$  are possibly correlated, then*

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n Z_i' u_i - \theta^n \xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R_{ni}' \right) \right)$$

where

$$\begin{aligned} R_{ni} &= Q_i - \frac{1}{n} \sum_{k=1}^n Q_k a_{ik}, \\ a_{ik} &= \lambda_i' \left( \Lambda' \Lambda / n \right)^{-1} \lambda_k, \\ Q_i &= B_{\varepsilon i} - \left( \int B_{\varepsilon i} B_{\eta}' \right) \left( \int B_{\eta} B_{\eta}' \right)^{-1} B_{\eta} \\ \theta^n &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{T} Z_i' \left( \Delta \bar{x}_i \quad \Delta F \right) \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \left( I_k \quad -\bar{\delta}_i' \right) \begin{pmatrix} \bar{\Delta}_{\varepsilon ui}^+ \\ \bar{\Delta}_{\eta u}^+ \end{pmatrix} \right] \end{aligned}$$

with  $\bar{\delta}_i = \left( F^{0'} F^0 \right)^{-1} F^{0'} \bar{x}_i$ , and  $\bar{x}_i = x_i - \frac{1}{n} \sum_{k=1}^n x_k a_{ik}$ .

**Proof of (a).** Recall

$$\begin{aligned} M_{F^0}x_i &= \left( I_T - F^0 \left( F^{0'} F^0 \right)^{-1} F^{0'} \right) x_i \\ &= x_i - F^0 \left( F^{0'} F^0 \right)^{-1} F^{0'} x_i = x_i - F^0 \delta_i \end{aligned}$$

where

$$\delta_i = \left( F^{0'} F^0 \right)^{-1} F^{0'} x_i = \left( \frac{F^{0'} F^0}{T^2} \right)^{-1} \frac{1}{T^2} \sum_{t=1}^T F_t^0 x_{it} \xrightarrow{d} \left( \int B_\eta B_\eta' \right)^{-1} \int B_\eta B_{\varepsilon i}' = \pi_i$$

is a  $r \times k$  matrix as  $T \rightarrow \infty$ . First we note that

$$M_{F^0}x_i = x_i - F^0 \delta_i$$

can be seen as the residual from a spurious regression of  $x_i$  on  $F$ . Let

$$\tilde{x}_i = x_i - F^0 \delta_i$$

be a  $T \times k$  matrix. Hence

$$\begin{aligned} Z_i &= M_{F^0}x_i - \frac{1}{n} \sum_{k=1}^n M_{F^0}x_k a_{ik} \\ &= (x_i - F^0 \delta_i) - \frac{1}{n} \sum_{k=1}^n (x_k - F^0 \delta_k) a_{ik} = \tilde{x}_i - \frac{1}{n} \sum_{k=1}^n \tilde{x}_k a_{ik} \end{aligned}$$

where  $a_{ik} = \lambda_i' \left( \Lambda' \Lambda / n \right)^{-1} \lambda_k$  a scalar and

$$\frac{\tilde{x}_{it}}{\sqrt{T}} = \frac{x_{it}}{\sqrt{T}} - \delta_i' \frac{F_t^0}{\sqrt{T}} \xrightarrow{d} B_{\varepsilon i} - \left[ \left( \int B_\eta B_\eta' \right)^{-1} \int B_\eta B_{\varepsilon i}' \right]' B_\eta = Q_i$$

a  $k \times 1$  vector, e.g., Phillips and Ouliaris (1990), p. 169 as  $T \rightarrow \infty$ . It follows that

$$\frac{Z_{it}}{\sqrt{T}} \xrightarrow{d} Q_i - \frac{1}{n} \sum_{k=1}^n Q_k a_{ik} = R_{ni}$$

and as  $n \rightarrow \infty$ ,

$$\frac{1}{nT^2} \sum_{i=1}^n \int R_{ni} R_{ni}' \xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int R_{ni} R_{ni}' \right).$$

Let

$$\xi_{iT} = \frac{1}{T^2} \sum_{t=1}^T Z_{it} Z_{it}'$$

Then as  $T \rightarrow \infty$ ,

$$\xi_{iT} \xrightarrow{d} \xi_i = \int R_{ni} R'_{ni}.$$

It can be shown that  $\|\xi_{iT}\|$  is uniformly integrable in  $T$  for all  $i$ .<sup>8</sup> Apply Theorem 1 in Phillips and Moon (1999) we have

$$\frac{1}{nT^2} \sum_{i=1}^n Z'_i Z_i \xrightarrow{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int R_{ni} R'_{ni} \right)$$

as  $(n, T) \rightarrow \infty$  showing (a).

**Proof of part (b).** Notice that

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{i=1}^n Z'_i u_i &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( M_{F^0} x_i - \frac{1}{n} \sum_{k=1}^n M_{F^0} x_k a_{ik} \right)' u_i \\ &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n (M_{F^0} x_i)' u_i - \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( \frac{1}{n} \sum_{k=1}^n M_{F^0} x_k a_{ik} \right)' u_i \\ &= I_b + II_b. \end{aligned}$$

Consider  $I_b$ .

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{i=1}^n (M_{F^0} x_i)' u_i &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n (x_i - F^0 \delta_i)' u_i \\ &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \tilde{x}'_i u_i = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it} u_{it}. \end{aligned}$$

Note

$$\frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{it} \xrightarrow{d} \int Q_i dB_{ui} \sim \left[ \Omega_{ui} \int Q_i Q'_i \right]^{1/2} \times N(0, I_k)$$

as  $T \rightarrow \infty$ . Let

$$\zeta_{iT} = \frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{it}.$$

It is clear that  $E[\zeta_{iT}] = 0$  and

$$\begin{aligned} E[\zeta_{iT} \zeta'_{iT}] &= E \left[ \left( \frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{it} \right) \left( \frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{it} \right)' \right] \\ &\rightarrow E \left[ \left( \int Q_i dB_{ui} \right) \left( \int Q_i dB_{ui} \right)' \right] \\ &= \Omega_{ui} \int Q_i Q'_i \end{aligned}$$

---

<sup>8</sup>If  $\xi_{iT} \xrightarrow{d} \xi_i$  as  $T \rightarrow \infty$ , the uniformly integrability of  $\|\xi_{iT}\|$  is equivalent to

$$E \|\xi_{iT}\| \rightarrow E \|\xi_i\|$$

as  $T \rightarrow \infty$ .

as  $T \rightarrow \infty$ . The  $\zeta_{iT}$  is an i.i.d. sequence with mean zero and covariance  $\Omega_{ui} \int Q_i Q_i'$ . It can be shown that  $\|\zeta_{iT}\|^2$  is uniformly integrable. Using Theorems 3 and 8 in Phillips and Moon (1999),

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n (M_{F^0} x_i)' u_i \xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{ui} E \left( \int Q_i Q_i' \right) \right)$$

as  $(n, T) \rightarrow \infty$  when  $\frac{n}{T} \rightarrow 0$  if  $\tilde{x}_{it}$  and  $u_{it}$  are uncorrelated.

Similarly, for  $II_b$ , we have

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( \frac{1}{n} \sum_{k=1}^n a_{ik} M_{F^0} x_k \right)' u_i \xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{ui} E C_{ni} \right)$$

where  $C_{ni} = \frac{1}{n} \sum_{k=1}^n a_{ik} \int Q_k Q_k'$  we have used the fact that  $\frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n a_{ik} a_{ij} = \frac{1}{n} \sum_{k=1}^n a_{ik}$ . Thus both  $I_b$  and  $II_b$  have a proper limiting distribution. These distributions are dependent since they depend on the same  $u_i$ . We can also derive their joint limiting distribution. Given the form of  $Z_i$ , it is easy to show that the above convergence implies part (b).

**Proof of part (c).** Now suppose  $\tilde{x}_{it}$  and  $u_{it}$  are correlated. It is known that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{it} &= \frac{1}{T} \sum_{t=1}^T (x_{it} - \delta_i' F_t^0) u_{it} = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} I_k & -\delta_i' \end{pmatrix} \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{it} \\ &= \begin{pmatrix} I_k & -\delta_i' \end{pmatrix} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{it} \\ &\xrightarrow{d} \begin{pmatrix} I_k & -\pi_i' \end{pmatrix} \int \begin{pmatrix} B_{\varepsilon i} dB_{ui} \\ B_{\eta} dB_{ui} \end{pmatrix} + \begin{pmatrix} \Delta_{\varepsilon ui} \\ \Delta_{\eta u} \end{pmatrix} \\ &= \int Q_i dB_{ui} + \begin{pmatrix} I_k & -\pi_i' \end{pmatrix} \begin{pmatrix} \Delta_{\varepsilon ui} \\ \Delta_{\eta u} \end{pmatrix} \end{aligned} \tag{24}$$

as  $T \rightarrow \infty$  (e.g., Phillips and Durlauf, 1986). First we note

$$\begin{aligned} \int Q_i dB_{ui} &= \int Q_i d \left( \Omega_{u,bi}^{1/2} V_i + \Omega_{ubi} \Omega_{bi}^{-1/2} W_i \right) \\ &= \int Q_i dB_{u,bi} + \int Q_i dB_{bi}' \Omega_{bi}^{-1} \Omega_{bui} \end{aligned}$$

such that

$$\begin{aligned} E \left[ \int Q_i dV_i \right] &= E \left[ E \left[ \int Q_i dV_i \right] | \pi_i \right] \\ &= E \left[ E \left[ \int (B_{\varepsilon i} - \delta_i' B_{\eta}) dV_i | \pi_i \right] \right] = 0. \end{aligned}$$

Note

$$\begin{aligned}
& \frac{1}{T} \tilde{x}'_i M_{F^0} \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} \\
&= \frac{1}{T} \tilde{x}'_i \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} \\
&= \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} I_k & -\pi'_i \end{pmatrix} \left[ \int \begin{pmatrix} B_{\varepsilon i} \\ B_{\eta} \end{pmatrix} dB'_{bi} \Omega_{bi}^{-1} \Omega_{bui} + \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui} \right].
\end{aligned}$$

Therefore

$$\begin{aligned}
& \frac{1}{T} \tilde{x}'_i u_i - \left[ \frac{1}{T} \tilde{x}'_i M_{F^0} \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} + \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} [\Delta_{bui} - \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui}] \right] \\
&= \frac{1}{T} \tilde{x}'_i u_i - \left[ \frac{1}{T} \tilde{x}'_i M_{F^0} \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} + \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} \Delta_{bui}^+ \right] \\
&\xrightarrow{d} \Omega_{u.bi}^{1/2} \int Q_i dV_i \sim \left[ \Omega_{u.bi}^{1/2} \int Q_i Q'_i \right]^{1/2} \times N(0, I_k)
\end{aligned} \tag{25}$$

where

$$\Delta_{bui}^+ = \Delta_{bui} - \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui}.$$

Let

$$\theta_1^n = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{T} \tilde{x}'_i M_{F^0} \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} \Omega_{bi}^{-1} \Omega_{bui} + \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} \Delta_{bui}^+ \right].$$

Then we apply Theorem 3 in Phillips and Moon (1999) to get

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \sum_{i=1}^n \tilde{x}'_i u_i - \theta_1^n &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{it} u_{it} - \theta_1^n \\
&\xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{u.bi} E \left( \int Q_i Q'_i \right) \right)
\end{aligned}$$

as  $(n, T) \rightarrow \infty$ .

Note  $Z_i = \tilde{x}_i - \frac{1}{n} \sum_{k=1}^n \tilde{x}_k a_{ik}$  is a demeaned  $\tilde{x}_i$  where  $\frac{1}{n} \sum_{k=1}^n \tilde{x}_k a_{ik}$  is the weighted average of  $\tilde{x}_i$  with the weight  $a_{ik}$ . It follows that

$$\begin{aligned}
Z_i &= \tilde{x}_i - \frac{1}{n} \sum_{k=1}^n \tilde{x}_k a_{ik} \\
&= (x_i - F^0 \delta_i) - \frac{1}{n} \sum_{k=1}^n (x_i - F^0 \delta_i) a_{ik} \\
&= \left( x_i - \frac{1}{n} \sum_{k=1}^n x_k a_{ik} \right) - F^0 \left( \delta_i - \frac{1}{n} \sum_{k=1}^n \delta_k a_{ik} \right)' \\
&= \bar{x}_i - F^0 \bar{\delta}'_i
\end{aligned}$$

where  $\bar{x}_i = x_i - \frac{1}{n} \sum_{k=1}^n x_k a_{ik}$  and  $\bar{\delta}_i = \delta_i - \frac{1}{n} \sum_{k=1}^n \delta_k a_{ik}$ .

We then can modify (24) as

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T Z_{it} u_{it} &= \frac{1}{T} \sum_{t=1}^T (\bar{x}_{it} - \bar{\delta}'_i F_t^0) u_{it} \\
&= \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} I_k & -\bar{\delta}'_i \end{pmatrix} \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix} u_{it} = \begin{pmatrix} I_k & -\bar{\delta}'_i \end{pmatrix} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix} u_{it} \\
&\xrightarrow{d} \begin{pmatrix} I_k & -\bar{\pi}'_i \end{pmatrix} \int \begin{pmatrix} \bar{B}_{\varepsilon i} dB_{ui} \\ B_{\eta} dB_{ui} \end{pmatrix} + \begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} \\
&= \int R_{ni} dB_{ui} + \begin{pmatrix} I_k & -\bar{\pi}'_i \end{pmatrix} \begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} \tag{26}
\end{aligned}$$

where  $\bar{B}_{\varepsilon i} = B_{\varepsilon i} - \frac{1}{n} \sum_{k=1}^n B_{\varepsilon i} a_{ik}$  and

$$\bar{\delta}_i = \delta_i - \frac{1}{n} \sum_{k=1}^n \delta_k a_{ik} \xrightarrow{d} \left( \int B_{\eta} B'_{\eta} \right)^{-1} \int B_{\eta} \bar{B}'_{\varepsilon i} = \bar{\pi}_i.$$

The  $R_{ni}$  terms appears in the last line in (26) this is because

$$\begin{aligned}
\bar{B}_{\varepsilon i} - \bar{\pi}'_i B_{\eta} &= \left( B_{\varepsilon i} - \frac{1}{n} \sum_{k=1}^n B_{\varepsilon k} a_{ik} \right) - \left( \int B_{\eta} B'_{\eta} \right)^{-1} \int B_{\eta} \left( B_{\varepsilon i} - \frac{1}{n} \sum_{k=1}^n B_{\varepsilon k} a_{ik} \right)' B_{\eta} \\
&= B_{\varepsilon i} - \left[ \left( \int B_{\eta} B'_{\eta} \right)^{-1} \int B_{\eta} B'_{\varepsilon i} \right] B_{\eta} - \frac{1}{n} \sum_{k=1}^n \left\{ B_{\varepsilon k} - \left[ \left( \int B_{\eta} B'_{\eta} \right)^{-1} \int B_{\eta} B'_{\varepsilon k} \right] B_{\eta} \right\} a_{ik} \\
&= Q_i - \frac{1}{n} \sum_{k=1}^n Q_k a_{ik} = R_{ni}.
\end{aligned}$$

Let

$$\theta^n = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{T} Z'_i (\Delta \bar{x}_i \quad \Delta F) \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \begin{pmatrix} I_k & -\bar{\delta}'_i \end{pmatrix} \bar{\Delta}_{bui}^+ \right].$$

Clearly

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \sum_{i=1}^n Z'_i u_i - \theta^n &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( \tilde{x}_i - \frac{1}{n} \sum_{k=1}^n \tilde{x}_k a_{ik} \right)' u_i \\
&\xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R'_{ni} \right) \right)
\end{aligned}$$

as  $(n, T \rightarrow \infty)$  with  $R_{ni} = Q_i - \frac{1}{n} \sum_{k=1}^n Q_k a_{ik}$ . This proves (c). ■

■

**Proof of Theorem 1.**

This follows directly from Lemma 1 as  $(n, T) \rightarrow \infty$  when  $\frac{n}{T} \rightarrow 0$

$$\sqrt{n}T \left( \widehat{\beta}_{Cup} - \beta \right) - \sqrt{n}\phi_{nT} \xrightarrow{d} N \left( 0, D_Z^{-1} \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R'_{ni} \right) \right] D_Z^{-1} \right)$$

where  $D_Z = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left( \int R_{ni} R'_{ni} \right)$  and  $\phi_{nT} = \left[ \frac{1}{nT^2} \sum_{i=1}^n Z'_i Z_i \right]^{-1} \theta^n$ . ■

**Proof of Theorem 2.** The proof is similar to that of Theorem 3 below, thus omitted. ■

To prove Theorem 3, we need some preliminary results. First we examine the limiting distribution of the infeasible FM estimator,  $\widetilde{\beta}_{CupFM}$ . The endogeneity correction is achieved by modifying the variable  $y_{it}$  in (3) with the transformation

$$y_{it}^+ = y_{it} - \bar{\Omega}_{ubi} \bar{\Omega}_{bi}^{-1} \begin{pmatrix} \Delta \bar{x}_{it} \\ \Delta F_t^0 \end{pmatrix}$$

and

$$u_{it}^+ = u_{it} - \bar{\Omega}_{ubi} \bar{\Omega}_{bi}^{-1} \begin{pmatrix} \Delta \bar{x}_{it} \\ \Delta F_t^0 \end{pmatrix}.$$

By construction  $u_{it}^+$  has zero long-run covariance with  $\begin{pmatrix} \Delta \bar{x}'_{it} & \Delta F_t^{0'} \end{pmatrix}$  and hence the endogeneity can be removed. The serial correlation correction term has the form

$$\begin{aligned} \bar{\Delta}_{bui}^+ &= \begin{pmatrix} \bar{\Delta}_{\varepsilon ui}^+ \\ \bar{\Delta}_{\eta u}^+ \end{pmatrix} = \begin{pmatrix} \bar{\Delta}_{bui} & \bar{\Delta}_{bi} \end{pmatrix} \begin{pmatrix} I_k \\ -\bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} \end{pmatrix} \\ &= \bar{\Delta}_{bui} - \bar{\Delta}_{bi} \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui}, \end{aligned}$$

where  $\bar{\Delta}_{bui}$  denotes the one-sided long-run covariance between  $u_{it}$  and  $(\varepsilon_{it}, \eta_t)$ . Therefore, the infeasible FM estimator is

$$\widetilde{\beta}_{CupFM} = \left( \sum_{i=1}^n x'_i M_{F^0} x_i \right)^{-1} \sum_{i=1}^n \left( x'_i M_{F^0} y_i^+ - T \left( \bar{\Delta}_{\varepsilon ui}^+ - \bar{\delta}'_i \bar{\Delta}_{\eta u}^+ \right) \right)$$

with  $\bar{\delta}_i = \left( F^{0'} F^0 \right)^{-1} F^{0'} \bar{x}_i$ .

The following Lemma gives the limiting distribution of  $\widetilde{\beta}_{CupFM}$ .

**Lemma 2** *Assume Assumptions in Theorem 1 hold. Then as  $(n, T) \rightarrow \infty$  with  $\frac{n}{T} \rightarrow 0$*

$$\sqrt{n}T \left( \widetilde{\beta}_{CupFM} - \beta^0 \right) \xrightarrow{d} N \left( 0, D_Z^{-1} \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R'_{ni} \right) \right] D_Z^{-1} \right).$$

**Proof.** Let  $w_{it}^+ = \begin{pmatrix} u_{it}^+ & \varepsilon'_{it} & \eta' \end{pmatrix}$  and we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_{it}^+ \xrightarrow{d} \begin{bmatrix} B_{ui}^+ \\ B_{\varepsilon i} \\ B_{\eta} \end{bmatrix} = \begin{bmatrix} B_{ui}^+ \\ B_{bi} \end{bmatrix} = BM(\Omega_i^+) \text{ as } T \rightarrow \infty, \quad (27)$$

where

$$\begin{aligned}
B_{bi} &= \begin{bmatrix} B_{\varepsilon i} \\ B_{\eta} \end{bmatrix}, & \Omega_{u.bi} &= \Omega_{ui} - \Omega_{ubi}\Omega_{bi}^{-1}\Omega_{bui}, \\
\Omega_i^+ &= \begin{bmatrix} \Omega_{u.bi} & 0 \\ 0 & \Omega_{bi} \end{bmatrix} = \begin{bmatrix} \Omega_{u.bi} & 0 & 0 \\ 0 & \Omega_{\varepsilon i} & \Omega_{\varepsilon\eta i} \\ 0 & \Omega_{\eta\varepsilon i} & \Omega_{\eta} \end{bmatrix} \\
&= \Sigma^+ + \Gamma^+ + \Gamma^{+'}, \\
\begin{bmatrix} B_{ui}^+ \\ B_{bi} \end{bmatrix} &= \begin{bmatrix} I & -\Omega_{ubi}\Omega_{bi}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} B_{ui} \\ B_{bi} \end{bmatrix}.
\end{aligned}$$

Define  $\Delta_i^+ = \Sigma_i^+ + \Gamma_i^+$ . and let  $u_{1it}^+ = u_{it} - \Omega_{ubi}\Omega_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t \end{pmatrix}$ . First we notice from (25) in Lemma 1 that

$$\begin{aligned}
\zeta_{1iT}^+ &= \frac{1}{T} \sum_{t=1}^T \tilde{x}_{it} u_{1it}^+ = \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{1it}^+ \\
&= \begin{pmatrix} I_k & -\delta'_i \end{pmatrix} \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{it} - \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} \Omega_{ubi}\Omega_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix} \right] \\
&= \Omega_{u.bi}^{1/2} \int Q_i dV_i + \left( \Delta_{\varepsilon ui}^+ - \pi'_i \Delta_{\eta u}^+ \right) \tag{28}
\end{aligned}$$

as  $T \rightarrow \infty$ . Now let

$$\zeta_{1iT}^* = \zeta_{1iT}^+ - \left( \Delta_{\varepsilon ui}^+ - \delta'_i \Delta_{\eta u}^+ \right).$$

Clearly,

$$\zeta_{1iT}^* \xrightarrow{d} \Omega_{u.bi}^{1/2} \int Q_i dV_i.$$

Thus,

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x'_i M_{F^0} u_{1i}^+ - T \left( \Delta_{\varepsilon ui}^+ - \delta'_i \Delta_{\eta u}^+ \right) \right) &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( \sum_{t=1}^T \tilde{x}_{it} u_{1it}^+ - T \left( \Delta_{\varepsilon ui}^+ - \delta'_i \Delta_{\eta u}^+ \right) \right) \\
&\xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{u.bi} E \left( \int Q_i Q'_i \right) \right).
\end{aligned}$$

The result in joint limit follows in the same manner as in Theorem 1. Next, we modify (28).

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T Z_{it} u_{it}^+ &= \frac{1}{T} \sum_{t=1}^T \left( \bar{x}_{it} - \bar{\delta}'_i F_t^0 \right) u_{it}^+ \\
&= \left( I_k \quad -\bar{\delta}'_i \right) \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix} u_{it}^+ - \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix} \Omega_{ubi} \Omega_{bi}^{-1} \begin{pmatrix} \Delta \bar{x}_{it} \\ \Delta F_t^0 \end{pmatrix} \right] \\
&\xrightarrow{d} \left( I_k \quad -\bar{\pi}'_i \right) \left\{ \int \begin{pmatrix} \bar{B}_{\varepsilon i} \\ B_{\eta} \end{pmatrix} dB_{ui} + \begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} - \left[ \int \begin{pmatrix} \bar{B}_{\varepsilon i} \\ B_{\eta} \end{pmatrix} dB'_{bi} \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \bar{\Delta}_{bi} \right] \right\} \\
&= \int R_{ni} dB_{ui} + \left( I_k \quad -\bar{\pi}'_i \right) \begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} - \int \left[ R_{ni} dB'_{bi} \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \left( I_k \quad -\bar{\pi}'_i \right) \begin{pmatrix} \bar{\Delta}_{\varepsilon i} \\ \bar{\Delta}_{\eta} \end{pmatrix} \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} \right] \\
&= \bar{\Omega}_{u.bi}^{1/2} \int R_{ni} dV_i + \left( \bar{\Delta}_{\varepsilon ui}^+ - \bar{\pi}'_i \bar{\Delta}_{\eta u}^+ \right)
\end{aligned}$$

Therefore,

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( Z'_i u_i^+ - T \left( \bar{\Delta}_{\varepsilon ui}^+ - \bar{\delta}'_i \bar{\Delta}_{\eta u}^+ \right) \right) \xrightarrow{d} N \left( 0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \bar{\Omega}_{u.bi} E \left( \int R_{ni} R'_{ni} \right) \right)$$

as  $(n, T) \rightarrow \infty$ . Then

$$\begin{aligned}
&\sqrt{nT} \left( \tilde{\beta}_{CupFM} - \beta^0 \right) \\
&\xrightarrow{d} N \left( 0, D_Z^{-1} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{u.bi} E \left( \int R_{ni} R'_{ni} \right) D_Z^{-1} \right)
\end{aligned}$$

as  $(n, T) \rightarrow \infty$  when  $\frac{n}{T} \rightarrow 0$ . This proves the theorem. ■

Next we show  $\sqrt{nT} \left( \hat{\beta}_{CupFM} - \tilde{\beta}_{CupFM} \right) = o_p(1)$ . The following lemma is needed.

**Lemma 3** *Under Assumptions of Theorem 2 and Assumption 6. If  $(n, T) \rightarrow \infty$ , then*

- (a)  $\sqrt{n} \left( \hat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) = o_p(1)$ ,
- (b)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \delta'_i \hat{\Delta}_{\eta u}^+ - \delta'_i \Delta_{\eta u}^+ \right) = o_p(1)$ ,
- (c)  $\frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x'_i M_{\hat{F}} \hat{u}_i^+ - x'_i M_{F^0} u_i^+ \right) = o_p(1)$

where  $\hat{u}_{it}^+ = u_{it} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix}$ ,  $\hat{\Delta}_{\varepsilon un}^+ = \frac{1}{n} \sum_{i=1}^n \hat{\Delta}_{\varepsilon ui}^+$  and  $\Delta_{\varepsilon un}^+ = \frac{1}{n} \sum_{i=1}^n \Delta_{\varepsilon ui}^+$ .

**Proof.** To simplify the notations, in this lemma we only show the results for the long run covariances without the bar since the steps are the same. First, note that

$$\Delta_{bui}^+ = \begin{pmatrix} \Delta_{\varepsilon ui}^+ \\ \Delta_{\eta u}^+ \end{pmatrix} = \begin{pmatrix} \Delta_{bui} & \Delta_{bi} \end{pmatrix} \begin{pmatrix} 1 \\ -\Omega_{bi}^{-1} \Omega_{bui} \end{pmatrix} = \Delta_{bui} - \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui}.$$

Then

$$\Delta_{\varepsilon ui}^+ = \Delta_{\varepsilon ui} - \Delta_{\varepsilon i} \Omega_{\varepsilon i}^{*-1} \Omega_{\varepsilon ui}$$

where  $\Omega_{\varepsilon i}^{*-1}$  is the first  $k \times k$  block of  $\Omega_{bi}^{-1}$ . Following the same lines the proofs of Theorems 9 and 10 of Hannan (1970), we have

$$\begin{aligned} & E \left\| \sqrt{n} \left( \widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) \right\|^2 \\ & \leq \sup_i E \left\| \widehat{\Delta}_{\varepsilon ui}^+ - E \widehat{\Delta}_{\varepsilon ui}^+ \right\|^2 + n \sup_i \left\| E \widehat{\Delta}_{\varepsilon ui}^+ - \Delta_{\varepsilon ui}^+ \right\|^2 \\ & = O\left(\frac{K}{T}\right) + O\left(\frac{n}{K^{2q}}\right). \end{aligned}$$

It follows that

$$\sqrt{n} \left( \widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) = O_p \left( \max \sqrt{\frac{K}{T}}, \sqrt{\frac{n}{K^{2q}}} \right).$$

From Assumption 6.  $K \sim n^b$ . Then

$$\frac{n}{K^{2q}} \sim \frac{n}{n^{2qb}} = n^{(1-2qb)} \rightarrow 0$$

if  $1 < 2qb$  or  $\frac{1}{2q} < b$ . Next

$$\frac{K}{T} \sim \frac{n^b}{T} = \exp \left( \log \left( \frac{n^b}{T} \right) \right) = \exp \left( b - \frac{\log T}{\log n} \right) \log n = n^{b - \frac{\log T}{\log n}} \leq n^{b - \liminf \frac{\log T}{\log n}} \rightarrow 0$$

if  $b < \liminf \frac{\log T}{\log n}$ . by Assumption 6. Then

$$\begin{aligned} \sqrt{n} \left( \widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) & = O_p \left( \max \sqrt{\frac{K}{T}}, \sqrt{\frac{n}{K^{2q}}} \right) \\ & = o_p(1) \end{aligned}$$

as required. This proves (a).

To establish (b), we note

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \delta'_i \widehat{\Delta}_{\eta u}^+ - \delta'_i \Delta_{\eta u}^+ \right) & = \left( \frac{1}{n} \sum_{i=1}^n \delta'_i \right) \sqrt{n} \left( \widehat{\Delta}_{\eta u}^+ - \Delta_{\eta u}^+ \right) \\ & = O_p(1) O_p \left( \max \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{n}{K^{2q}}} \right\} \right) = o_p(1) \end{aligned}$$

as required for part (b).

Let  $\tilde{u}_{it}^+ = u_{it} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \begin{pmatrix} \Delta x_{it} \\ \Delta F_t \end{pmatrix}$ . Next,

$$\begin{aligned}
& \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{\hat{F}} \tilde{u}_i^+ - x_i' M_{F^0} u_i^+ \right) \\
= & \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{\hat{F}} \tilde{u}_i^+ - x_i' M_{\hat{F}} \tilde{u}_i^+ + x_i' M_{\hat{F}} \tilde{u}_i^+ - x_i' M_{\hat{F}} u_i^+ + x_i' M_{\hat{F}} u_i^+ - x_i' M_{F^0} u_i^+ \right) \\
= & \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{\hat{F}} \tilde{u}_i^+ - x_i' M_{\hat{F}} u_i^+ \right) + \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{\hat{F}} u_i^+ - x_i' M_{F^0} u_i^+ \right) + \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left( x_i' M_{\hat{F}} \tilde{u}_i^+ - x_i' M_{\hat{F}} \tilde{u}_i^+ \right) \\
= & \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\tilde{u}_i^+ - u_i^+) + \frac{1}{\sqrt{nT}} \sum_{i=1}^n (x_i' M_{\hat{F}} - x_i' M_{F^0}) u_i^+ + \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\tilde{u}_i^+ - \tilde{u}_i^+) \\
= & \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\tilde{u}_i^+ - u_i^+) + \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' (M_{\hat{F}} - M_{F^0}) u_i^+ + \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\tilde{u}_i^+ - \tilde{u}_i^+) \\
= & I + II + III.
\end{aligned}$$

From Bai et al. (2006) we know that

$$II = \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' (M_{\hat{F}} - M_{F^0}) u_i^+ = o_p(1)$$

if we replace  $u_i$  by  $u_i^+$ . Let  $\Delta b_i = \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix}$  be a  $T \times (k+r)$  matrix. Consider  $I$ .

$$\begin{aligned}
\frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\tilde{u}_i^+ - u_i^+) &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} \left( u_i - \Delta b_i \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui} - u_i + \Delta b_i \Omega_{bi}^{-1} \Omega_{bui} \right) \\
&= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} \left( \Delta b_i \left( \Omega_{ubi} \Omega_{bi}^{-1} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \right) \right) \\
&= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' \left( I_T - \frac{\hat{F} \hat{F}'}{T^2} \right) \left( \Delta b_i \left( \Omega_{ubi} \Omega_{bi}^{-1} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \right) \right) \\
&= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' \Delta b_i \left( \Omega_{ubi} \Omega_{bi}^{-1} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \right) \\
&\quad - \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' \frac{\hat{F} \hat{F}'}{T^2} \left( \Delta b_i \left( \Omega_{ubi} \Omega_{bi}^{-1} - \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} \right) \right) \\
&= I_c + II_c.
\end{aligned}$$

Along the same lines as the proofs of Theorems 9 and 10 of Hannan (1970), we can show that

$$\sup_i E \left\| \hat{\Omega}_{ubi} \hat{\Omega}_{bi}^{-1} - \Omega_{ubi} \Omega_{bi}^{-1} \right\|^2 = O\left(\frac{K}{T}\right) + O\left(\frac{1}{K^2 q}\right).$$

Then we have

$$\Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} = O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right).$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 &= \sqrt{n} \frac{1}{n} \sum_{i=1}^n \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 \\ &\leq \sqrt{n} \sup_i \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 \\ &= \sqrt{n} \left[ O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right) \right]^2. \end{aligned}$$

For  $I_c$ , by the Cauchy Schwarz inequality,

$$\begin{aligned} \|I_c\| &= \left\| \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' \Delta b_i \left( \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right) \right\| \\ &\leq \left( \sqrt{n} \frac{1}{n} \sum_{i=1}^n \left\| \frac{x_i' \Delta b_i}{T} \right\|^2 \right)^{1/2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 \right)^{1/2} \\ &\leq [O_p(\sqrt{n})]^{1/2} (\sqrt{n})^{1/2} O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right) \\ &= O_p(\sqrt{n}) O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right) \end{aligned}$$

Similarly,

$$\begin{aligned} \|II_c\| &= \left\| \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' \frac{\widehat{F}\widehat{F}'}{T^2} \left( \Delta b_i \left( \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right) \right) \right\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{x_i' \widehat{F}}{T^2} \frac{\widehat{F}' \Delta b_i}{T} \left( \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right) \right\| \\ &\leq \left\| \left( \sqrt{n} \frac{1}{n} \sum_{i=1}^n \left\| \frac{x_i' \widehat{F}}{T^2} \frac{\widehat{F}' \Delta b_i}{T} \right\|^2 \right)^{1/2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 \right)^{1/2} \right\| \\ &= O_p(\sqrt{n}) O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right) \end{aligned}$$

Combining  $I_c$  and  $II_c$ , we have

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\widehat{F}} (\widehat{u}_i^+ - \widetilde{u}_i^+) &= O_p(\sqrt{n}) O_p \left( \text{Max} \left\{ \sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}} \right\} \right) \\ &= O_p \left( \text{Max} \left\{ \sqrt{\frac{nK}{T}}, \sqrt{\frac{n}{K^{2q}}} \right\} \right) \end{aligned}$$

Recall  $K \sim n^b$  and  $\liminf \frac{\log T}{\log n} > 1$  from Assumption 6. It follows that

$$\begin{aligned} \frac{nK}{T} &\sim \frac{n^{b+1}}{T} = \exp\left(\log\left(\frac{n^{b+1}}{T}\right)\right) = \exp\left(b+1 - \frac{\log T}{\log n}\right) \log n \\ &= n^{b+1 - \frac{\log T}{\log n}} \leq n^{b+1 - \liminf \frac{\log T}{\log n}} \rightarrow 0 \end{aligned}$$

by Assumption 6 and  $b < \liminf \frac{\log T}{\log n} - 1$ . Also note

$$\frac{n}{K^{2q}} \sim \frac{n}{n^{2qb}} = n^{(1-2qb)} \rightarrow 0$$

by Assumption 6 and  $\frac{1}{2q} < b$ . Therefore

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\hat{u}_i^+ - \tilde{u}_i^+) = O_p\left(\text{Max}\left\{\sqrt{\frac{nK}{T}}, \sqrt{\frac{n}{K^{2q}}}\right\}\right) = o_p(1).$$

Let

$$\Delta \hat{b}_i = \begin{pmatrix} \Delta x_i & \Delta \hat{F} \end{pmatrix}.$$

Note that

$$\Delta b_i - \Delta \hat{b}_i = \begin{pmatrix} \Delta x_i & \Delta F \end{pmatrix} - \begin{pmatrix} \Delta x_i & \Delta \hat{F} \end{pmatrix} = \begin{pmatrix} 0 & \Delta F - \Delta \hat{F} \end{pmatrix}.$$

Consider III.

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\hat{u}_i^+ - \tilde{u}_i^+) &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (u_i - \Delta \hat{b}_i \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui} - u_i + \Delta b_i \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui}) \\ &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\Delta b_i - \Delta \hat{b}_i) \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui} \\ &= \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\Delta F - \Delta \hat{F}) \hat{\Omega}_{bi}^{-1} \hat{\Omega}_{bui}. \end{aligned}$$

We use Lemma 12.3 in Bai (2005) to get

$$\frac{1}{nT} \sum_{i=1}^n x_i' M_{\hat{F}} (\Delta F - \Delta \hat{F}) = O_p(\hat{\beta} - \beta^0) + O_p\left(\frac{1}{\min(n, T)}\right).$$

It follows that

$$\begin{aligned} \frac{1}{\sqrt{nT}} \sum_{i=1}^n x_i' M_{\hat{F}} (\Delta F - \Delta \hat{F}) &= \sqrt{n} \left[ O_p(\hat{\beta} - \beta^0) + O_p\left(\frac{1}{\min(n, T)}\right) \right] \\ &= \sqrt{n} O_p\left(\frac{1}{T}\right) + O_p\left(\frac{\sqrt{n}}{\min(n, T)}\right) = o_p(1) \end{aligned}$$

since  $\frac{n}{T} \rightarrow 0$  as  $(n, T) \rightarrow \infty$ . Collecting I – III we prove (c). ■

**Proposition 5** Assume Assumptions 1-6 hold. Then

$$\sqrt{n}T \left( \widehat{\beta}_{CupFM} - \widetilde{\beta}_{CupFM} \right) = o_p(1).$$

**Proof.** To save the notations, we only show that results when  $\bar{x}_i = x_i$  and  $\bar{\delta}_i = \delta_i$  since the steps are basically the same. It is easy to show, see Bai et al (2006), that

$$\left( \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{\widehat{F}} x_i \right) = \left( \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{F^0} x_i \right) + o_p(1).$$

Then

$$\begin{aligned} & \sqrt{n}T \left( \widehat{\beta}_{CupFM} - \widetilde{\beta}_{CupFM} \right) \\ &= \left( \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{F^0} x_i \right)^{-1} \frac{1}{\sqrt{n}T} \left\{ \begin{array}{l} \sum_{i=1}^n \left( x_i' M_{\widehat{F}} \widehat{u}_i^+ - T \left( \widehat{\Delta}_{\varepsilon ui}^+ - \delta_i' \widehat{\Delta}_{\eta u}^+ \right) \right) \\ - \sum_{i=1}^n \left( x_i' M_{F^0} u_i^+ - T \left( \Delta_{\varepsilon ui}^+ - \delta_i' \Delta_{\eta u}^+ \right) \right) \end{array} \right\} + o_p(1) \\ &= \left( \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{F^0} x_i \right)^{-1} \frac{1}{\sqrt{n}T} \left\{ \begin{array}{l} \sum_{i=1}^n \left( x_i' M_{\widehat{F}} \widehat{u}_i^+ - x_i' M_{F^0} u_i^+ \right) \\ - nT \left( \widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) - T \sum_{i=1}^n \left( \delta_i' \widehat{\Delta}_{\eta u}^+ - \delta_i' \Delta_{\eta u}^+ \right) \end{array} \right\} + o_p(1) \\ &= \left( \frac{1}{nT^2} \sum_{i=1}^n x_i' M_{F^0} x_i \right)^{-1} \left\{ \begin{array}{l} \frac{1}{\sqrt{n}T} \sum_{i=1}^n \left( x_i' M_{\widehat{F}} \widehat{u}_i^+ - x_i' M_{F^0} u_i^+ \right) \\ - \sqrt{n} \left( \widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+ \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \delta_i' \widehat{\Delta}_{\eta u}^+ - \delta_i' \Delta_{\eta u}^+ \right) \end{array} \right\} + o_p(1) \end{aligned}$$

where  $\widehat{\Delta}_{\varepsilon un}^+ = \frac{1}{n} \sum_{i=1}^n \widehat{\Delta}_{\varepsilon ui}^+$  and  $\Delta_{\varepsilon un}^+ = \frac{1}{n} \sum_{i=1}^n \Delta_{\varepsilon ui}^+$ . Finally using Lemma 3,

$$\sqrt{n}T \left( \widehat{\beta}_{CupFM} - \widetilde{\beta}_{CupFM} \right) = o_p(1). \blacksquare$$

■

**Proof of Theorem 3:** This follows directly from Proposition 5. ■

## References

- [1] Andrews, D. W. K. (2005), "Cross-section Regression with Common Shocks," *Econometrica*, 73, 1551-1585.
- [2] Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica* 71, 135-171.
- [3] Bai, J. (2004), "Estimating Cross-section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics*, 122, 137-183.
- [4] Bai, J. (2005), "Panel Data Models with Interactive Fixed Effects", working paper.
- [5] Bai, J., and Kao, C. (2005), "On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence," forthcoming in *Contributions to Economic Analysis*, edited by Badi Baltagi, Elsevier.
- [6] Bai, J., Kao, C., and Ng, S. (2006). "Panel Cointegration with Global Stochastic Trends: Supplementary Appendix," Working paper.
- [7] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.
- [8] Bai, J., and Ng, S., (2004), "A Panic Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127-1177.
- [9] Baltagi, B. (2005), *Econometric Analysis of Panel Data*, New York, Wiley.
- [10] Baltagi, B., and Kao, C., (2000), Nonstationary panels, cointegration in panels and dynamic panels: A survey, *Advances in Econometrics*, 15, 7-51.
- [11] Breitung, J. (2005), "A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data," *Econometric Review*, 24, 151-173.
- [12] Breitung, J., and Das, S. (2005a), "Panel Unit Root Tests Under Cross Sectional Dependence," *Statistica Neerlandica*, 59, 414-433.
- [13] Breitung, J., and Das, S. (2005b), "Testing for Unit Roots in Panels with a Factor Structure," working paper.
- [14] Breitung, J., and Pesaran, M. H. (2005), "Unit Roots and Cointegration in Panels," working paper.

- [15] Chang, Y. (2002), "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 110, 261-292.
- [16] Chang, Y. (2004), "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 120, 263-293.
- [17] Chang, Y., and Song, W. (2005), "Unit Root Tests for Panels in the Presence of Short-run and Long-run Dependencies: Nonlinear IV Approach with Fixed n and large T." working paper.
- [18] Choi, I. (2006), "Combination Unit Root Tests for Cross-Sectionally Correlated Panels," *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, Cambridge, Cambridge University Press.
- [19] Gengenbach, C., Palm, F. C., and Urbain, J.-P. (2004), "Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications of Modelling," working paper.
- [20] Gengenbach, C., Palm, F. C., and Urbain, J.-P. (2005), "Panel Cointegration Testing in the Presence of Common Factors," working paper.
- [21] Hannan, E. (1970), *Multiple Time Series*, New York, Wiley.
- [22] Hansen, B.E. (1992), Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics*, 53, 87-121.
- [23] Harris, D., Leybourne, S., and McCabe, B. (2004), "Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence," working paper.
- [24] Hsiao, C., (2003), *Analysis of Panel Data*, Cambridge, Cambridge University Press.
- [25] Kao, C. (1999), "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90, 1-44.
- [26] Kao, C., and Chiang, M. H. (2000), "On the Estimation and Inference of a Cointegrated Regression in Panel Data," *Advances in Econometrics*, 15, 179-222.
- [27] Mark, N. C., and Sul, D. (2003), "Cointegration Vector Estimation by Panel DOLS and Long-run Money Demand," *Oxford Bulletin of Economics and Statistics*, 65, 655-680.
- [28] Moon, H. R., and Perron, B. (2004), "Testing for a Unit Root in Panels with Dynamic Factor," *Journal of Econometrics*, 122, 81-126.

- [29] Moon, H. R., and Phillips, P. C. B. (2000), "Estimation of Autoregressive Roots of Near Unit Using Panel Data," *Econometric Theory*, 927-997.
- [30] Moon, H. R., and Phillips, P. C. B. (2004), "GMM Estimation of Autoregressive Roots Near Unit with Panel Data," *Econometric Theory*, 927-997.
- [31] Park, J. (1990), "Testing for Unit Roots and Cointegration By Variable Addition," *Advances in Econometrics*, ed. by Fomby and Rhodes, 107-133, JAI Press.
- [32] Park, J. and Phillips, P. C. B. (1988), "Regressions with Integrated Regressors: Part I," *Econometric Theory*, , 468-498.
- [33] Pesaran, M. H. (2005), "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence," working paper.
- [34] Pesaran, M. H., and Smith, R. (1995), "Estimation of Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79-114.
- [35] Pesaran, M. H., and Yamagata, T. (2006), "Testing Slope Homogeneity in Large Panels," working paper.
- [36] Phillips, P. C. B. (1995), "Fully-modified least squares and autoregression," *Econometrica*, 63, 1023-1079.
- [37] Phillips, P. C. B., and Durlauf, S. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.
- [38] Phillips, P. C. B., and B.E. Hansen (1990), "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-125.
- [39] Phillips, P. C. B., and Moon, H., (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1111.
- [40] Phillips, P. C. B., and Ouliaris, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-193.
- [41] Phillips, P. C. B., and Sul, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *Econometrics Journal*, 6, 217-259.
- [42] Sun, Y. (2004), "Estimation on the Long-run Average Relationship in Nonstationary Panel Time Series," *Econometric Theory*, 20, 1227-1260.

- [43] Westerlund, J. (2005), “New Simple Tests for Panel Cointegration,” *Econometrics Reviews*, 24, 297-316.
- [44] Westerlund, J. (2006), “On the Estimation of Cointegrated Panel Regressions with Common Factors,” working paper.
- [45] Westerlund, J., and Edgerton, D. (2006), “Simple Tests for Cointegration Panels and Structural Breaks,” working paper.

Table 1: Means Biases and Standard Deviation of Estimators

	$\sigma_{31} = 0$			$\sigma_{31} = 0.8$			$\sigma_{31} = -0.8$		
	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM
$\sigma_{21} = 0$									
T=20	0.038 (1.559)	0.006 (0.387)	0.001 (0.029)	0.058 (1.505)	-0.010 (0.372)	-0.000 (0.028)	-0.005 (1.524)	0.013 (0.385)	0.000 (0.023)
T=40	0.047 (1.175)	-0.005 (0.235)	-0.004 (0.009)	0.036 (1.186)	-0.001 (0.416)	-0.000 (0.009)	0.022 (1.181)	-0.005 (0.232)	-0.000 (0.009)
T=60	-0.056 (0.904)	0.005 (0.132)	-0.000 (0.005)	-0.012 (0.924)	-0.002 (0.138)	-0.000 (0.005)	-0.043 (0.966)	0.005 (0.141)	-0.000 (0.005)
T=120	0.020 (0.633)	0.001 (0.057)	-0.000 (0.002)	0.026 (0.632)	-0.003 (0.053)	-0.000 (0.002)	-0.004 (0.616)	0.005 (0.057)	-0.000 (0.002)
$\sigma_{21} = 0.2$									
T=20	0.059 (1.563)	0.005 (0.372)	0.004 (0.028)	0.082 (1.529)	-0.011 (0.382)	0.003 (0.028)	0.018 (1.527)	0.005 (0.404)	0.003 (0.027)
T=40	0.047 (1.175)	-0.001 (0.236)	0.000 (0.009)	0.045 (1.171)	-0.003 (0.232)	0.001 (0.009)	0.019 (1.165)	-0.002 (0.224)	0.001 (0.009)
T=60	-0.047 (0.909)	0.000 (0.133)	0.000 (0.005)	-0.007 (0.935)	-0.001 (0.137)	0.000 (0.005)	-0.019 (0.988)	0.000 (0.145)	0.000 (0.005)
T=120	0.024 (0.641)	0.001 (0.058)	0.000 (0.002)	0.031 (0.629)	-0.001 (0.053)	0.000 (0.002)	-0.007 (0.616)	0.000 (0.057)	0.000 (0.002)
$\sigma_{21} = -0.2$									
T=20	-0.017 (1.544)	0.014 (0.389)	-0.002 (0.029)	0.026 (1.484)	-0.014 (0.345)	-0.003 (0.028)	0.026 (1.066)	-0.007 (0.263)	-0.002 (0.029)
T=40	0.024 (1.169)	-0.004 (0.242)	-0.001 (0.009)	0.020 (0.838)	0.001 (0.169)	-0.000 (0.007)	0.020 (0.838)	-0.001 (0.169)	-0.001 (0.007)
T=60	-0.069 (0.897)	0.008 (0.142)	-0.001 (0.005)	-0.007 (0.909)	-0.002 (0.138)	-0.001 (0.005)	-0.014 (0.649)	-0.001 (0.010)	-0.000 (0.004)
T=120	0.002 (0.636)	-0.000 (0.058)	-0.000 (0.002)	0.020 (0.637)	0.001 (0.055)	-0.000 (0.002)	-0.005 (0.437)	0.000 (0.039)	-0.000 (0.001)

Note:

(a)  $T = n$ ,  $c = 5$ ,  $\sigma_{32} = 0.4$ .

**Table 2: Means Biases and Standard Deviation of Estimators for Different  $n$  and  $T$**

(n, T)	$c = 5$						$c = 10$					
	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM	LSDV	2sFM	2sFM	CupFM	CupFM	
(20, 20)	0.082 (1.529)	-0.011 (0.382)	0.003 (0.028)	-0.135 (3.058)	-0.021 (0.761)	-0.003 (0.028)	0.082 (1.529)	-0.011 (0.382)	0.003 (0.028)	-0.135 (3.058)	-0.021 (0.761)	-0.003 (0.028)
(20, 40)	0.039 (1.554)	-0.003 (0.283)	0.000 (0.014)	0.064 (3.107)	-0.006 (0.566)	-0.001 (0.014)	0.039 (1.554)	-0.003 (0.283)	0.000 (0.014)	0.064 (3.107)	-0.006 (0.566)	-0.001 (0.014)
(20, 60)	0.092 (1.580)	-0.009 (0.217)	-0.000 (0.009)	0.174 (3.160)	-0.019 (0.434)	-0.001 (0.009)	0.092 (1.580)	-0.009 (0.217)	-0.000 (0.009)	0.174 (3.160)	-0.019 (0.434)	-0.001 (0.009)
(20, 120)	0.018 (1.646)	0.018 (0.129)	0.000 (0.004)	0.001 (0.231)	-0.000 (0.003)	-0.000 (0.003)	0.018 (1.646)	0.018 (0.129)	0.000 (0.004)	0.001 (0.231)	-0.000 (0.003)	-0.000 (0.003)
(40, 20)	-0.026 (1.073)	0.000 (0.280)	0.003 (0.019)	-0.083 (2.147)	0.000 (0.558)	0.002 (0.019)	-0.026 (1.073)	0.000 (0.280)	0.003 (0.019)	-0.083 (2.147)	0.000 (0.558)	0.002 (0.019)
(40, 40)	0.037 (1.137)	-0.006 (0.222)	0.001 (0.009)	0.060 (2.274)	-0.011 (0.443)	0.001 (0.009)	0.037 (1.137)	-0.006 (0.222)	0.001 (0.009)	0.060 (2.274)	-0.011 (0.443)	0.001 (0.009)
(40, 60)	0.032 (1.077)	-0.007 (0.156)	0.000 (0.006)	0.054 (2.154)	-0.014 (0.312)	0.000 (0.006)	0.032 (1.077)	-0.007 (0.156)	0.000 (0.006)	0.054 (2.154)	-0.014 (0.312)	0.000 (0.006)
(40, 120)	-0.068 (1.121)	0.002 (0.093)	0.000 (0.003)	-0.142 (2.243)	0.004 (0.187)	0.000 (0.003)	-0.068 (1.121)	0.002 (0.093)	0.000 (0.003)	-0.142 (2.243)	0.004 (0.187)	0.000 (0.003)
(60, 20)	-0.003 (0.897)	0.008 (0.229)	0.003 (0.017)	-0.069 (1.730)	0.014 (0.456)	-0.003 (0.017)	-0.003 (0.897)	0.008 (0.229)	0.003 (0.017)	-0.069 (1.730)	0.014 (0.456)	-0.003 (0.017)
(60, 40)	-0.001 (0.919)	-0.001 (0.194)	0.001 (0.008)	-0.014 (1.296)	-0.001 (0.027)	0.001 (0.006)	-0.001 (0.919)	-0.001 (0.194)	0.001 (0.008)	-0.014 (1.296)	-0.001 (0.027)	0.001 (0.006)
(60, 60)	-0.009 (0.946)	-0.000 (0.141)	0.000 (0.005)	-0.023 (1.332)	-0.000 (0.199)	-0.000 (0.004)	-0.009 (0.946)	-0.000 (0.141)	0.000 (0.005)	-0.023 (1.332)	-0.000 (0.199)	-0.000 (0.004)
(60, 120)	-0.015 (0.945)	-0.002 (0.082)	0.000 (0.003)	-0.009 (1.843)	-0.006 (0.153)	0.000 (0.002)	-0.015 (0.945)	-0.002 (0.082)	0.000 (0.003)	-0.009 (1.843)	-0.006 (0.153)	0.000 (0.002)
(120, 20)	-0.002 (0.640)	0.001 (0.176)	0.003 (0.012)	-0.034 (1.280)	0.002 (0.352)	0.003 (0.012)	-0.002 (0.640)	0.001 (0.176)	0.003 (0.012)	-0.034 (1.280)	0.002 (0.352)	0.003 (0.012)
(120, 40)	-0.005 (0.689)	0.004 (0.137)	0.001 (0.005)	-0.024 (1.379)	0.008 (0.275)	0.001 (0.005)	-0.005 (0.689)	0.004 (0.137)	0.001 (0.005)	-0.024 (1.379)	0.008 (0.275)	0.001 (0.005)
(120, 60)	0.009 (0.629)	-0.004 (0.101)	0.001 (0.003)	0.009 (1.258)	-0.008 (0.202)	0.001 (0.004)	0.009 (0.629)	-0.004 (0.101)	0.001 (0.003)	0.009 (1.258)	-0.008 (0.202)	0.001 (0.004)
(120, 120)	0.032 (0.623)	-0.001 (0.056)	0.000 (0.002)	0.059 (1.247)	-0.003 (0.111)	-0.000 (0.002)	0.032 (0.623)	-0.001 (0.056)	0.000 (0.002)	0.059 (1.247)	-0.003 (0.111)	-0.000 (0.002)

(a)  $\sigma_{21} = 0.2$ ,  $\sigma_{31} = 0.8$ , and  $\sigma_{32} = 0.4$ .

**Table 3: Means Biases and Standard Deviation of t-statistics**

	$\sigma_{31} = 0$			$\sigma_{31} = 0.8$			$\sigma_{31} = -0.8$		
	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM
$\sigma_{21} = 0$									
T=20	0.036 (2.414)	0.092 (2.445)	0.045 (1.469)	0.096 (2.419)	-0.059 (2.449)	0.001 (1.459)	-0.074 (2.471)	0.109 (2.446)	0.023 (1.461)
T=40	0.148 (3.674)	-0.051 (2.627)	-0.029 (1.236)	0.179 (3.644)	-0.057 (2.658)	-0.048 (1.223)	0.025 (3.623)	-0.009 (2.602)	-0.048 (1.217)
T=60	-0.317 (4.283)	0.090 (2.655)	-0.065 (1.188)	-0.071 (4.383)	-0.017 (2.599)	-0.061 (1.184)	-0.236 (4.361)	0.016 (2.596)	-0.059 (1.162)
T=120	0.206 (5.986)	0.086 (2.696)	-0.038 (1.133)	0.269 (6.101)	0.023 (2.618)	-0.095 (1.123)	-0.088 (5.987)	0.051 (2.705)	-0.011 (1.122)
$\sigma_{21} = 0.2$									
T=20	0.097 (2.426)	0.075 (2.448)	0.228 (1.470)	0.160 (2.415)	-0.042 (2.426)	0.188 (1.442)	-0.024 (2.468)	0.122 (2.456)	0.191 (1.443)
T=40	0.185 (3.664)	0.018 (2.684)	0.133 (1.226)	0.201 (3.618)	-0.043 (2.700)	0.108 (1.218)	0.046 (3.664)	0.017 (2.602)	0.117 (1.218)
T=60	-0.284 (4.309)	0.091 (2.647)	0.032 (1.183)	-0.066 (4.445)	0.019 (2.606)	0.081 (1.168)	-0.192 (4.397)	0.021 (2.595)	0.040 (1.159)
T=120	0.262 (5.998)	0.099 (2.692)	0.075 (1.133)	0.279 (6.095)	-0.012 (2.633)	0.063 (1.110)	-0.076 (6.029)	0.045 (2.750)	0.077 (1.122)
$\sigma_{21} = -0.2$									
T=20	-0.067 (2.528)	0.117 (2.478)	-0.083 (1.471)	0.051 (2.417)	-0.055 (2.464)	-0.118 (1.464)	0.051 (2.417)	-0.055 (2.464)	-0.118 (1.464)
T=40	0.008 (3.623)	-0.052 (2.615)	-0.195 (1.245)	0.131 (3.678)	-0.073 (2.682)	-0.158 (1.215)	0.131 (3.648)	-0.173 (2.659)	-0.158 (1.228)
T=60	-0.376 (4.296)	0.094 (2.691)	-0.203 (1.163)	-0.015 (4.296)	-0.089 (2.626)	-0.198 (1.164)	-0.118 (4.365)	-0.048 (2.618)	-0.161 (1.184)
T=120	0.057 (6.141)	-0.026 (2.750)	-0.131 (1.132)	0.287 (6.152)	0.027 (2.577)	-0.160 (1.131)	-0.084 (6.045)	0.089 (2.722)	-0.091 (1.121)

Note:

(a)  $\sigma_{32} = 0.4$ .

**Table 4: Means Biases and Standard Deviation  
of t-statistics for Different n and T**

(n, T)	c = 5						c = 10					
	LSDV	2sFM	CupFM	LSDV	2sFM	CupFM	LSDV	2sFM	2sFM	LSDV	2sFM	CupFM
(20, 20)	0.160 (2.415)	-0.042 (2.426)	0.188 (1.443)	0.127 (2.351)	-0.067 (2.541)	0.188 (1.442)	0.160 (2.415)	-0.042 (2.426)	0.188 (1.443)	0.127 (2.351)	-0.067 (2.541)	0.188 (1.442)
(20, 40)	0.148 (3.511)	-0.006 (1.850)	0.092 (1.289)	0.124 (3.475)	-0.009 (1.684)	0.092 (1.303)	0.148 (3.511)	-0.006 (1.850)	0.092 (1.289)	0.124 (3.475)	-0.009 (1.684)	0.092 (1.303)
(20, 60)	0.207 (4.469)	-0.071 (1.589)	0.007 (1.272)	0.189 (4.472)	-0.067 (1.597)	0.007 (1.272)	0.207 (4.469)	-0.071 (1.589)	0.007 (1.272)	0.189 (4.472)	-0.067 (1.597)	0.007 (1.272)
(20, 120)	0.003 (6.069)	0.019 (1.186)	0.047 (1.146)	0.017 (6.069)	0.017 (1.181)	0.047 (1.146)	0.003 (6.069)	0.019 (1.186)	0.047 (1.146)	0.017 (6.069)	0.017 (1.181)	0.047 (1.146)
(40, 20)	-0.055 (2.486)	0.009 (3.412)	0.269 (1.437)	-0.103 (2.489)	-0.028 (3.664)	0.269 (1.438)	-0.055 (2.486)	0.009 (3.412)	0.269 (1.437)	-0.103 (2.489)	-0.028 (3.664)	0.269 (1.438)
(40, 40)	0.107 (3.592)	-0.037 (2.676)	0.103 (1.219)	0.073 (3.594)	-0.051 (2.780)	0.103 (1.219)	0.107 (3.592)	-0.037 (2.676)	0.103 (1.219)	0.073 (3.594)	-0.051 (2.780)	0.103 (1.219)
(40, 60)	0.069 (4.158)	-0.101 (2.104)	0.081 (1.159)	0.045 (4.166)	-0.106 (2.166)	0.081 (1.162)	0.069 (4.158)	-0.101 (2.104)	0.081 (1.159)	0.045 (4.166)	-0.106 (2.166)	0.081 (1.162)
(40, 120)	-0.394 (5.998)	0.034 (1.490)	0.046 (1.075)	-0.414 (5.999)	0.0316 (1.507)	0.046 (1.074)	-0.394 (5.998)	0.034 (1.490)	0.046 (1.075)	-0.414 (5.999)	0.0316 (1.507)	0.046 (1.074)
(60, 20)	0.031 (2.533)	0.228 (3.959)	0.330 (1.479)	-0.151 (2.414)	0.012 (2.602)	-0.288 (1.492)	0.031 (2.533)	0.228 (3.959)	0.330 (1.479)	-0.151 (2.414)	0.012 (2.602)	-0.288 (1.492)
(60, 40)	0.050 (3.570)	-0.109 (3.293)	0.144 (1.290)	-0.054 (3.530)	-0.165 (3.376)	-0.224 (1.259)	0.050 (3.570)	-0.109 (3.293)	0.144 (1.290)	-0.054 (3.530)	-0.165 (3.376)	-0.224 (1.259)
(60, 60)	-0.051 (4.426)	0.022 (2.613)	0.092 (1.152)	-0.187 (4.380)	-0.001 (2.731)	-0.206 (1.175)	-0.051 (4.426)	0.022 (2.613)	0.092 (1.152)	-0.187 (4.380)	-0.001 (2.731)	-0.206 (1.175)
(60, 120)	-0.165 (6.292)	-0.069 (1.932)	0.041 (1.100)	-0.090 (6.166)	-0.096 (1.949)	-0.104 (1.099)	-0.165 (6.292)	-0.069 (1.932)	0.041 (1.100)	-0.090 (6.166)	-0.096 (1.949)	-0.104 (1.099)
(120, 20)	0.059 (2.466)	0.077 (5.338)	0.396 (1.428)	-0.021 (2.468)	0.108 (6.038)	0.396 (1.428)	0.059 (2.466)	0.077 (5.338)	0.396 (1.428)	-0.021 (2.468)	0.108 (6.038)	0.396 (1.428)
(120, 40)	0.086 (3.686)	0.059 (4.288)	0.274 (1.201)	0.027 (3.686)	0.041 (4.643)	0.274 (1.201)	0.086 (3.686)	0.059 (4.288)	0.274 (1.201)	0.027 (3.686)	0.041 (4.643)	0.274 (1.201)
(120, 60)	0.133 (4.318)	-0.281 (3.731)	0.202 (1.161)	0.058 (4.315)	-0.272 (3.933)	0.149 (1.159)	0.133 (4.318)	-0.281 (3.731)	0.202 (1.161)	0.058 (4.315)	-0.272 (3.933)	0.149 (1.159)
(120, 120)	0.280 (5.953)	-0.021 (2.690)	0.069 (1.120)	0.247 (5.954)	-0.023 (2.788)	0.068 (1.120)	0.280 (5.953)	-0.021 (2.690)	0.069 (1.120)	0.247 (5.954)	-0.023 (2.788)	0.068 (1.120)

(a)  $\sigma_{21} = 0.2$ ,  $\sigma_{31} = 0.8$ , and  $\sigma_{32} = 0.4$ .