CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Panel Cointegration with Global Stochastic Trends

Jushan Bai, Chihwa Kao, and Serena Ng

CEA@Cass Working Paper Series

WP–CEA–01-2007

PANEL COINTEGRATION WITH GLOBAL STOCHASTIC TRENDS

Jushan Bai [*] Chihwa Kao[†] Serena Ng [‡]

This version: December 21, 2006

Abstract

This paper studies estimation of panel cointegration 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 as 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; Common shocks; Co-movements; Cross-sectional dependence; Factor analysis; Fully modified estimator.

[*]Department of Economics, New York University, New York, NY 10003,USA, and School of Economics and Management, Tsinghua University. Email: Jushan.Bai@nyu.edu.

[†]Center for Policy Research and Department of Economics, Syracuse University, Syracuse, NY 13244-1020, USA. Email: cdkao@maxwell.syr.edu.

[‡]Department of Economics, University of Michigan, Ann Arbor, MI 48109, USA, Email: serena.Ng@umich.edu.
We thank Yixiao Sun, Joon Park, Yoosoon Chang, and Joakim Westerlund for helpful discussions. We would also like to thank seminar and conference participants at NYU, Texas A&M, CESG 2006, Academia Sinica, National Taiwan University, Greater New York Area Econometrics Colloquium at Yale, and CIREQ Time Series Conference in Montreal for helpful comments. Bai and Ng gratefully acknowledge financial support from the NSF (grants SES-0551275 and SES-0549978).

1 Introduction

This paper is concerned with estimating panel cointegration models using a large panel of data. Our focus is on estimating the slope parameters of the non-stationary regressors when the cross sections share common sources of non-stationary variation in the form of global stochastic trends. The standard least squares estimator is either inconsistent or has slower convergence rate. 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 $\sqrt{n}T$ 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 hypotheses. Two dimensional variations 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.[1] 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, $(1\ -\beta')$, 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

[1]See, for example, Baltagi (2005), Hsiao (2003), Pesaran and Smith (1995), Kao (1999), and Moon and Phillips (2000, 2004) in the context of testing the unit root hypothesis using panel data. Stock and Watson (2002) suggest diffusion-index forecasting, while Bernanke and Boivin (2003) suggest new formulations of vector autoregressions to exploit the information in large panels.

on panel cointegration already exists[2] for modelling 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 common shocks can potentially invalidate estimation and inference of β.[3] In view of this, more recent work has allowed for cross-sectional dependence of e_{it} when testing for the null hypothesis of panel cointegration.[4] There is also a growing literature on panel unit root tests with cross-sectional dependence.[5] In this paper, we consider estimation and inference of parameters in a panel model with cross-sectional dependence in the form of common stochastic trends.

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 β when F_t is non-stationary. A regression of y_{it} on x_{it} will give a consistent estimator for β 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 β using an iterated procedure. The procedure is shown to yield a consistent estimator of β, 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 $\sqrt{n}T$ consistent for the common slope coefficient

[2]See, for example, Phillips and Moon (1999) and Kao (1999). Recent surveys can be found in Baltagi and Kao (2000) and Breitung and Pesaran (2005).

[3]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.

[4]See, for example, Phillips and Sul (2003), Gengenbach et al. (2005b), and Westerlund (2006).

[5]For example, Chang (2002,2004), Choi (2006), Moon and Perron (2004), Breitung and Das (2005), Gengenbach et al. (2005a), and Westerlund and Edgerton (2006). Breitung and Pesaran (2005) provide additional references in their survey.

vector, β. The estimators enable use of standard test statistics such as t, F, and χ^2 for inference. The estimators are robust to mixed I(1)/I(0) factors, as well as mixed I(1)/I(0) regressors. Thus, our approach is an alternative to the solution proposed in Bai and Kao (2006) 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 examines issues related to incidental trends, mixed I(0)/I(1) regressors and mixed I(0)/I(1) common shocks, and issues of testing cross-sectional independence. Section 5 presents Monte Carlo results to illustrate the finite sample properties of the proposed estimators. Section 6 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 $i = 1, ..., n$, $t = 1, ..., T$, y_{it} is a scalar,

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

is a set of k non-stationary regressors, β 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 β defined by

$$\hat{\beta}_{LS} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} x_{it}^{'} \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} y_{it} \tag{2}$$

is, in general, T consistent.[6] 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 ε_{it}. The exception is when x_{it} is strictly exogenous, in which case the estimator is $\sqrt{n}T$ 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 $\sqrt{n}T$ consistency and asymptotic normality.

[6] The estimator can be regarded as $\sqrt{n}T$ consistent but with a bias of order $O(\sqrt{n})$. Up to the bias, the estimator is also asymptotically normal.

The cross-section independence assumption is restrictive and difficult to justify when the data under investigation are economic time series. In view of comovements of economic variables and common shocks, we model the cross-section dependence by imposing a factor structure on e_{it}. That is,

$$e_{it} = \lambda_i' F_t + u_{it}$$

where F_t is a $r \times 1$ vector of latent common factors, λ_i is a $r \times 1$ vector of factor loadings and u_{it} is the idiosyncratic error. If both F_t and u_{it} are stationary, then e_{it} is also stationary. In this case, a consistent estimator of the regression coefficients can still be obtained even when the cross-section dependence is ignored, just like the fact that simultaneity bias is of second order in the fixed n cointegration framework. Using this property, Bai and Kao (2006) considered a two-step fully modified estimator (2sFM). In the first step, pooled OLS is used to obtain a consistent estimate of β. 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

$$y_{it} = x_{it}'\beta + \lambda_i' F_t + u_{it}. \tag{3}$$

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 β 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 I(0) regressors. Estimating β from (3) with F being I(0) was suggested in Bai and Kao (2006), but its theory was not explored.

When F_t is I(1), which is the primary focus of this paper, there is an important difference between estimating β from (3) versus pooled OLS in (2) because the latter 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}

$$x_{it} = \tau_i' F_t + \xi_{it} \tag{4}$$

with ξ_{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

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

If τ_i and λ_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} conintegrates with x_{it} *and* F_t with cointegrating vector $(1,\ -\beta',\ \lambda_i)$. Omitting F_t creates a spurious regression problem. It is worth noting that the cointegrating vector varies with i because the factor loading is unit specific. Estimation of the parameter of interest β involves a new methodology because F is unobservable.

In the rest of the paper, we will show how to obtain $\sqrt{n}T$ consistent and asymptotically normal estimates of β 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 $\Omega = \left(\Omega^{1/2}\right)\left(\Omega^{1/2}\right)'$, and $BM(\Omega)$ to denote Brownian motion with the covariance matrix Ω. 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 $(n, T) \to \infty$. We also define the matrix that projects onto the orthogonal space of z as $M_z = I_T - z\left(z'z\right)^{-1}z'$. We will use β^0, F_t^0, and λ_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) $E\left\|\lambda_i^0\right\|^4 \le M$. *As* $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n}\lambda_i^0\lambda_i^{0\prime} \xrightarrow{p} \Sigma_\lambda$, *a* $r \times r$ *positive definite matrix.*

(b) $E\left\|\eta_t\right\|^{4+\delta} \le M$ *for some* $\delta > 0$ *and for all* t; *As* $T \to \infty$, $\frac{1}{T^2}\sum_{i=1}^{n}F_t^0F_t^{0\prime} \xrightarrow{d} \int B_\eta B_\eta'$, *a* $r \times r$ *random matrix, where* B_η *is a vector of Brownian motions with covariance matrix* Ω_η, *which is a positive definite matrix.*

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

Assumption 3 *Weak cross-sectional correlation and heterokedasticity*

(a) $E(u_{it}u_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ *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)* $\frac{1}{nT}\sum_{i,j,t,s=1}|\sigma_{ij,ts}| \leq M$.

(b) For every (t,s), $E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[u_{is}u_{it} - E(u_{is}u_{it})]\right|^4 \leq M$.

(c) $\frac{1}{nT^2}\sum_{t,s,u,v}\sum_{i,j}|cov(u_{it}u_{is}, u_{ju}u_{jv})| \leq M$ *and* $\frac{1}{nT^2}\sum_{t,s}\sum_{i,j,k,l}|cov(u_{it}u_{js}, u_{ku}u_{ls})| \leq M$.

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 Ω_η 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:

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

where

$$B_i = \begin{bmatrix} B_{ui} & B_{\varepsilon i}' & B_\eta' \end{bmatrix}'.$$

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

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

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

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

For future reference, it will be convenient to group elements corresponding to ε_{it} and η_t taken together. Let

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

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 $\begin{bmatrix} V_i & W_i^{'} \end{bmatrix}' = BM(I)$ 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 $(\triangle x_{it}^{'},\ \triangle F_t^{0'})'$. Note that $\Omega_{bi} > 0$ since we assume that there is no cointegration relationship in $(x_{it}^{'},\ F_t^{0'})'$ in Assumption 4.

Finally, we state an additional assumption, which is needed when deriving the limiting distribution of various estimators.

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

It is noted that this assumption is not needed for consistency of the proposed estimators.

3 Estimation

In this section, we first consider the problem of estimating β when F is observed. We then consider two iterative procedures that jointly estimate β and F. The procedures yield two estimators that are $\sqrt{n}T$ 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 $\Lambda = (\lambda_1, .., \lambda_n)^{'}$ to be an 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

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

After concentrating out λ, the least squares estimator for β is then

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

The least squares estimator has the following properties.[7]

Proposition 1 *Under Assumptions 1-5, as* $(n, T) \to \infty$

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

where

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

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

with

$$D = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E \left(\int Q_i Q_i' \right)$$

$$Q_i = B_{\varepsilon i} - \left(\int B_{\varepsilon i} B_\eta' \right) \left(\int B_\eta B_\eta' \right)^{-1} B_\eta,$$

$$\theta_i^0 = \frac{1}{T} x_i' M_{F^0} \Delta b_i \Omega_{bi}^{-1} \Omega_{bui} + \left(\Delta_{\varepsilon ui}^+ - \delta_i^{0\prime} \Delta_{\eta u}^+ \right),$$

$$\delta_i^0 = (F^{0\prime} F^0)^{-1} F^{0\prime} x_i, \quad \Delta b_i = \begin{pmatrix} \Delta x_i & \Delta F^0 \end{pmatrix}$$

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

The estimator is $\sqrt{n}T$ 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 θ_i^0. 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 Ω_{bui} and Δ_{bui}.

[7] The limiting distribution for F being I(0) can also be obtained. Park and Phillips (1988) provide the limiting theory with mixed I(1) and I(0) regressors in a single equation framework.

To estimate the bias, we need to consistently estimate the nuisance parameters. We use a kernel estimator. Let

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

where $\hat{w}_{it} = (\hat{u}_{it}, \Delta x_{it}', \Delta F_t^{0\prime})'$. To state the asymptotic theory for the bias-corrected estimator, we need the following assumption, as used in Moon and Perron (2004):

Assumption 6 *(a)* $\liminf\limits_{n,T\to\infty} (\log T/\log n) > 1$.

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

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

Let

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

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

$$\tilde{\beta}_{LSBC} = \tilde{\beta}_{LS} - \frac{1}{T}\hat{\phi}_{nT}^{0}. \tag{9}$$

This estimator can alternatively be written as

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

where $\tilde{y}^{+}$ and $\tilde{\Delta}^{+}$ are consistent estimates of y^{+} and Δ^{+} etc, with

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

Viewed in this light, the bias-corrected estimator is also a panel fully-modified estimator in the spirit of Phillips and Hansen (1990), and is the reason why the estimator is also labeled $\hat{\beta}_{LSFM}$. It is not difficult to verify that $\hat{\beta}_{LSBC}$ and $\hat{\beta}_{LSFM}$ are identical. Panel fully modified estimators were also considered by Phillips and Moon (1999) and Bai and Kao (2006). Here, we extend those analysis to allow for common stochastic trends. By construction u_{it}^{+} has a zero long-run covariance with $\left(\Delta x_{it}' \quad \Delta F_t^{0\prime} \right)'$ and hence the endogeneity can be removed. Furthermore, nuisance parameters arising from the low frequency correlation of the errors are summarized in Δ_{bui}^{+}.

Proposition 2 *Let $\widetilde{\beta}_{LSFM}$ be defined by (10). Under Assumptions 1-6, as $(n,T) \to \infty$*

$$\sqrt{n}T(\widetilde{\beta}_{LSFM} - \beta^0) \xrightarrow{d} N\left(0, \Sigma\right).$$

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 $\sqrt{n}T$ 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 Unobserved F and 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 ignores F is, in general, inconsistent. In this section, we consider estimating F along with β and Λ by minimizing the objective function

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

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

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

Define

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

Notice that given β, w_i has a pure factor structure. Let $W = (w_i, ..., w_n)$ 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 $\Lambda = W'F\left(F'F\right)^{-1} = T^{-2}W'F$, we have the concentrated objective function:

$$tr\left(W'M_F W\right) = tr\left(W'W\right) - tr\left(F'WW'F/T^2\right). \tag{12}$$

Since the first term does not depend on F, minimizing (12) with respect to F is equivalent to maximizing $tr\left(T^{-2}F'WW'F\right)$ 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 β, and similarly, β 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 (β, F) is defined as

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

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

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

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

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 $\hat{\beta}$ and $\hat{F}$ using (13) and (14). It is a non-linear estimator even though linear least squares estimation is involved at each iteration. An estimate of Λ can be obtained as:

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

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

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

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

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

We now turn to the asymptotic representation of $\hat{\beta}_{Cup}$.

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

$$\sqrt{n}T\left(\widehat{\beta}_{Cup} - \beta^0\right) = D\left(F^0\right)^{-1}\left[\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i' M_{F^0} - \frac{1}{n}\sum_{k=1}^{n} a_{ik} x_i' M_{F^0}\right)u_i\right] + o_p(1),$$

where $a_{ik} = \lambda_i' \left(\frac{\Lambda'\Lambda}{n}\right)^{-1} \lambda_k$, $D\left(F^0\right) = \frac{1}{nT^2}\sum_{i=1}^n Z_i' Z_i$ *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, estimation of the stochastic trends clearly affects the limiting behavior of the estimator. The term involving a_{ik} is due to the estimation of F. This effect is carried over to the limiting distribution and to the asymptotic bias, as we now proceed to show. Let $\bar{w}_{it} = (u_{it},\ \Delta\bar{x}_i', \eta_t')'$ where $\bar{x}_i = x_i - \frac{1}{n}\sum_{k=1}^n x_k a_{ik}$. 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) generated from $\bar{w}_{it}$ instead of w_{it}. Thus, $\bar{\Omega}_i$ is the long run covariance matrix of $\bar{w}_{it}$ as in (5), and define $\bar{\Delta}_i$ is the one-sided covariance matrix of $\bar{w}_{it}$. These quantities depend on n, but this dependence is suppressed for notional simplicity.

Because the right hand side of the representation does not depend on estimated quantities, it is not difficult to derive the limiting distribution of $\hat{\beta}_{Cup}$, even allowing for cross-sectional correlation in u_{it}. However, estimating the resulting nuisance parameters would be more difficult. Thus, although consistency of the Cup estimator does not require the cross-section independence of u_{it}, our asymptotic distribution for $\hat{\beta}_{Cup}$ is derived with Assumption 5 imposed.

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

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

where

$$\phi_{nT} = \left[\frac{1}{nT^2}\sum_{i=1}^n Z_i' Z_i\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^n \theta_i\right)$$

$$\theta_i = \frac{1}{T} Z_i' \Delta\bar{b}_i \bar{\Omega}_{bi}^{-1} \bar{\Omega}_{bui} + \left(\bar{\Delta}_{\varepsilon ui}^{+} - \bar{\delta}_i' \bar{\Delta}_{\eta u}^{+}\right),$$

$$\Sigma = D_Z^{-1}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \bar{\Omega}_{u.bi} E\left(\int R_{ni} R_{ni}'\right)\right] D_Z^{-1},\ \ D_Z = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E\left(\int R_{ni} R_{ni}'\right), \tag{15}$$

$$R_{ni} = Q_i - \frac{1}{n}\sum_{k=1}^n Q_k a_{ik},$$

$$\Delta\bar{b}_i = \left(\Delta\bar{x}_i \quad \Delta F^0 \right),$$

$$\bar{x}_i = x_i - \frac{1}{n}\sum_{k=1}^n x_k a_{ik},$$

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

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 ϕ_{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

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

where

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

The bias-corrected Cup estimator is defined as

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

where

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

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

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

The CupBC is $\sqrt{n}T$ consistent with a limiting distribution that is centered at zero as long as $(n,T) \to \infty$ and $\frac{n}{T} \to 0$. 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 $\sqrt{n}T$ 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 LSBC estimator, which coincides with LSFM and whose asymptotic variance is Σ^0 instead of Σ. Thus, the estimation of F affects the asymptotic distribution of the estimator. As in the case when F is observed, the bias corrected estimator can be rewritten as a fully modified estimator. Such a fully-modified estimator is now discussed.

3.3 Fully Modified Cup Estimator

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

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

where $\widehat{\bar{\Omega}}_{ubi}$, $\widehat{\bar{\Omega}}_{bi}$, and $\widehat{\bar{\Delta}}_{bui}$ 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:

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

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

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 (β, 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 $(n,T) \to \infty$ with $n/T \to 0$,*

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

where Σ is given in (15)

The CupFM and CupBC have the same asymptotic distribution, but they are constructed differently. The estimator $\widehat{\beta}_{CupBC}$ does the bias correction only once, i.e., at the final stage of the iteration, and $\widehat{\beta}_{CupFM}$ does the correction at every iteration. 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

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

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

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

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

3.4 Estimated Global stochastic trends

While the focus is on estimating the slope parameters β, the global stochastic trends F are also of interest. Our procedure produces consistent estimates of F. We state this result as a proposition.

Proposition 5 *Let $\hat{F}$ be the solution of (17). Under assumptions of 1-4, we have*

$$\frac{1}{T}\sum_{t=1}^{T} \|\hat{F}_t - HF_t^0\|^2 = O_p(\frac{1}{n}) + O_p(\frac{1}{T^2}).$$

where H is an $r \times r$ invertible matrix.

Thus, we can estimate the true global stochastic trends up to a rotation. This is the same rate as in Bai (2004, Lemma B.1), where the regressor x_{it} is absent. Similarly, the factor loadings λ_i are estimated with the same rate of convergence as in Bai (2004).

4 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, and 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. Finally, we also discuss the issue of testing cross-sectional independence.

4.1 Incidental trends

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

$$y_{it} = \alpha_i + \rho_i t + x_{it}'\beta + \lambda_i' F_t + u_{it}. \tag{18}$$

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

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

where ι_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 ι_T and to the linear trend. Then

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

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_η in Q_i are replaced by the demeaned Brownian motions.

When linear trends are allowed, Δx_{it} is now replaced by demeaned version of Δx_{it}, i.e., $\hat{\varepsilon}_{it} = \Delta x_{it} - \overline{\Delta x_i}$, 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 $\Delta \hat{\dot{x}}_{it}$ in Section 3. Similarly, $\hat{F}$ should also be detrended. More specifically, we use $\hat{\eta}_t = \Delta \hat{F}_t - \overline{\Delta \hat{F}}$ in place of $\Delta \hat{F}_t$. Alternatively, since $\dot{x}_i$ is already a detrended series, and $\hat{F}$ is also asymptotically detrended (since it is estimating $\dot{F}$), $\Delta \dot{x}_{it}$ and $\Delta \hat{F}_t$ 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_η are replaced by the demeaned and detrended Brownian motions.

In either case, the test statistics (t and χ^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 β 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).

4.2 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

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

with $\delta_i = (F^{0\prime} F^0)^{-1} F^{0\prime} x_i$ 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 mixed I(1) and I(0) 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 $T^{-1} F^{0\prime} F^0 \xrightarrow{p} \Sigma_F = O_p(1)$. Thus

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

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 $T \to \infty$. 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 Ω_η 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 $\begin{pmatrix} \Delta_{\eta ui} & \Delta_{\eta i} \end{pmatrix}$ in $\begin{pmatrix} \Delta_{bui} & \Delta_{bi} \end{pmatrix}$ are also degenerate because the factors are I(0). Note that Ω_{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 ΔF_t in the FM construction.

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

$$y_{it} = x_{it}^{'}\beta + \lambda_{1i}^{'}F_{1t} + \lambda_{2i}^{'}F_{2t} + u_{it} \tag{19}$$

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 $\delta_i = \begin{bmatrix} \delta_{1i} & \delta_{2i} \end{bmatrix}'$. 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

$$E(x_{1it}u_{it}) = 0 \tag{20}$$

to rule out simultaneity bias with I(0) regressors. Otherwise β_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 β_1 is no longer the structural parameter. The dynamic least squares approach by adding Δx_{2it} is exactly based on this argument, with the purpose of more efficient estimation of β_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 β_1 and β_2. The estimated β_1 needs no correction and is asymptotically normal, and the estimated β_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 by 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 $\hat{\beta}_{Cup}$.

4.3 Test of Cross-Sectional Dependence

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

$$H_0 : \lambda_i = 0 \quad \text{for all i} \tag{21}$$

for all i in (3) against the alternative that [8]

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

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

$$\widehat{y}_{it} = x_{it}' \widehat{\beta}_{FM} + \widehat{w}_{it}$$

[8] We thank Joon Park for this suggestion.

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

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

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

$$J_i = \frac{RSS_{1i} - RSS_{2i}}{\widehat{\Omega}_{u.\varepsilon i}}. \tag{22}$$

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 χ^2 distribution as $T \to \infty$ with degree of freedom equal to r, under the null hypothesis of r common factors. The 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[9] that as $(n, T) \to \infty$

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

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

5 Monte Carlo Simulations

In this section, we conduct Monte Carlo experiments to assess the finite sample properties of the proposed CupBC and CupFM estimators. We also compare the performance of the proposed estimators with that of LSDV (least squares dummy variables, i.e., the within group estimator) and 2sFM (2-stage fully modified which is the CupFM estimator with only one iteration).

Data are generated based on the following design. For $i = 1, ..., n$, $t = 1, ..., T$,

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

where[10]

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

[9] See e.g., similar to Theorem 3 in Pesaran and Yamagata (2006)

[10] 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.

We assume a single factor, i.e., $r = 1$, λ_i and η_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 in the system is controlled by only two parameters, σ_{21} and σ_{31}. The parameter c controls the importance of the global stochastic trends. We consider $c = (5, 10)$, $\sigma_{32} = 0.4$, $\sigma_{21} = (0, 0.2, -0.2)$ and $\sigma_{31} = (0, 0.8, -0.8)$.

The long-run covariance matrix is estimated using the KERNEL procedure in COINT 2.0. We use the Bartlett window with the truncation set at five. Results for other kernels, such as Parzen and quadratic spectral kernels, are similar and hence not reported. The maximum number of the iteration for CupBC and CupFM estimators is set to 20.

Table 1 reports the means and standard deviations (in parentheses) of the estimators for sample sizes $T = n = (20, 40, 60, 120)$. The results are based on $10,000$ replications. The bias of the LSDV estimator does not decrease as (n, T) increases in general. In terms of mean bias, the CupBC and CupFM are distinctly superior to the LSDV and 2sFM estimators for all cases considered. The 2sFM estimator is less efficient than the CupBC and CupFM estimators, as seen by the 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 importance of the common shock is magnified as we increase c from 5 to 10. On the other hand, the CupBC and CupFM estimators are unaffected by the values of c. The results in Table 2 again indicate that the CupBC and CupFM perform 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 CupBC and CupFM t-statistics are much better approximated by a standard N(0,1). Interesting, the performance of CupBC is no worse than that of CupFM, even though CupBC does the full modification in the final stage of iteration.

Table 4 shows that, as n and T increases, the biases for the t-statistics associated with LSDV and 2sFM do not decrease. For CupBC and CupFM, the biases for the t-statistics becomes smaller (except for a small number of cases) as T increases for each fixed n. As n increases, no improvement in biases is found. The large standard deviations in the t-statistics associated with LSDV and 2sFM indicate their poor performance, especially as T increases. For the CupBC and CupFM, the standard errors converge to 1.0 as n and T (especially as T)increase.

6 Conclusion

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

Appendix

The proofs for Propositions 1 and 2 (with observable F) use standard arguments and are hence omitted. Propositions 3 and 4 are proved in the supplementary appendix of 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 ε_{it}.

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

Lemma A.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 $(n,T) \to \infty$

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

(b) If $\frac{n}{T} \to 0$, *and if* u_i *is uncorrelated with* (x_i, F^0) *for all* i, *then*

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n} Z_i' u_i \xrightarrow{d} N\left(0,\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ui}E\left(\int R_{ni}R_{ni}'\right)\right)$$

(c) If $\frac{n}{T} \to 0$, *and if* u_i *is possibly correlated with* (x_i, F^0), *then*

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n} Z_i' u_i - \theta^n \xrightarrow{d} N\left(0,\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{\Omega}_{u.bi}E\left(\int R_{ni}R_{ni}'\right)\right)$$

where

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

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

Proof of (a). Recall

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

where

$$\delta_i = \left(F^{0\prime}F^0\right)^{-1}F^{0\prime}x_i = \left(\frac{F^{0\prime}F^0}{T^2}\right)^{-1}\frac{1}{T^2}\sum_{t=1}^{T}F_t^0 x_{it}^{\prime} \xrightarrow{d} \left(\int B_\eta B_\eta^{\prime}\right)^{-1}\int B_\eta B_{\varepsilon i}^{\prime} = \pi_i$$

is a $r \times k$ matrix as $T \to \infty$. 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

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

be a $T \times k$ matrix. Hence

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

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

$$\frac{\widetilde{x}_{it}}{\sqrt{T}} = \frac{x_{it}}{\sqrt{T}} - \delta_i^{\prime}\frac{F_t^0}{\sqrt{T}} \xrightarrow{d} B_{\varepsilon i} - \left[\left(\int B_\eta B_\eta^{\prime}\right)^{-1}\int B_\eta B_{\varepsilon i}^{\prime}\right]^{\prime}B_\eta = Q_i$$

a $k \times 1$ vector, e.g., Phillips and Ouliaris (1990), p. 169 as $T \to \infty$. 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 $n \to \infty$,

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

Let

$$\xi_{iT} = \frac{1}{T^2}\sum_{t=1}^{T}Z_{it}Z_{it}^{\prime}$$

Then as $T \to \infty$,

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

It can be shown that $\|\xi_{iT}\|$ is uniformly integrable in T for all i.[11] Apply Theorem 1 in Phillips and Moon (1999) we have

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

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

Proof of part (b). Notice that

$$\begin{aligned}
\frac{1}{\sqrt{n}T} \sum_{i=1}^{n} Z_i^{'} u_i &= \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(M_{F^0} x_i - \frac{1}{n} \sum_{k=1}^{n} M_{F^0} x_k a_{ik} \right)^{'} u_i \\
&= \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(M_{F^0} x_i \right)^{'} u_i - \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{k=1}^{n} M_{F^0} x_k a_{ik} \right)^{'} u_i \\
&= I_b + II_b.
\end{aligned}$$

Consider I_b.

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

Note

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

as $T \to \infty$. Let

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

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

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

[11] If $\xi_{iT} \xrightarrow{d} \xi_i$ as $T \to \infty$, the uniformly integrability of $\|\xi_{iT}\|$ is equivalent to $E\|\xi_{iT}\| \to E\|\xi_i\|$ as $T \to \infty$.

as $T \to \infty$. The ζ_{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),

$$\frac{1}{\sqrt{n}T} \sum_{i=1}^{n} (M_{F^0} x_i)' u_i \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Omega_{ui} E\left(\int Q_i Q_i'\right)\right)$$

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

Similarly, for II_b, we have

$$\frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{k=1}^{n} a_{ik} M_{F^0} x_k\right)' u_i \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Omega_{ui} E C_{ni}\right)$$

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 convergences imply part (b).

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

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

as $T \to \infty$ (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

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

Note

$$
\begin{aligned}
& \frac{1}{T} x_i^{'} M_{F^0} \left(\begin{array}{cc} \Delta x_i & \Delta F \end{array} \right) \Omega_{bi}^{-1} \Omega_{bui} \\
= & \frac{1}{T} \widetilde{x}_i^{'} \left(\begin{array}{cc} \Delta x_i & \Delta F \end{array} \right) \Omega_{bi}^{-1} \Omega_{bui} \\
= & \left(\begin{array}{cc} I_k & -\delta_i^{'} \end{array} \right) \frac{1}{T} \sum_{t=1}^{T} \left(\begin{array}{c} x_{it} \\ F_t^0 \end{array} \right) \Omega_{bi}^{-1} \Omega_{bui} \left(\begin{array}{c} \Delta x_{it} \\ \Delta F_t^0 \end{array} \right) \\
& \xrightarrow{d} \left(\begin{array}{cc} I_k & -\pi_i^{'} \end{array} \right) \left[\int \left(\begin{array}{c} B_{\varepsilon i} \\ B_\eta \end{array} \right) dB_{bi}^{'} \Omega_{bi}^{-1} \Omega_{bui} + \Delta_{bi} \Omega_{bi}^{-1} \Omega_{bui} \right].
\end{aligned}
$$

Therefore

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

where

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

Let

$$
\theta_1^n = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{T} x_i^{'} M_{F^0} \left(\begin{array}{cc} \Delta x_i & \Delta F \end{array} \right) \Omega_{bi}^{-1} \Omega_{bui} + \left(\begin{array}{cc} I_k & -\delta_i^{'} \end{array} \right) \Delta_{bui}^{+} \right].
$$

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

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

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

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

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

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

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

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

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

Clearly

$$\begin{aligned}
\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}Z_i^{'}u_i - \theta^n &= \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(\widetilde{x}_i - \frac{1}{n}\sum_{k=1}^{n}\widetilde{x}_k a_{ik}\right)^{'}u_i \\
&\xrightarrow{d} N\left(0,\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{\Omega}_{u.bi}E\left(\int R_{ni}R_{ni}^{'}\right)\right)
\end{aligned}$$

as $(n, T\to\infty)$ 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 A.1 as $(n,T)\to\infty$ when $\frac{n}{T}\to 0$

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

where $D_Z=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\int R_{ni}R_{ni}^{'}\right)$ 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 $\begin{pmatrix}\Delta\bar{x}_{it}^{'} & \Delta F_t^{0\prime}\end{pmatrix}^{'}$ 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 $\bar{\delta}_i=\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\bar{x}_i$.

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

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

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

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

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

where

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

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 A.1 that

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

as $T \to \infty$. 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,

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

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

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

Therefore,

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

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

$$
\begin{aligned}
&\sqrt{n}T\left(\widetilde{\beta}_{CupFM}-\beta^0\right) \\
&\xrightarrow{d}N\left(0,D_Z^{-1}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{u.bi}E\left(\int R_{ni}R_{ni}'\right)D_Z^{-1}\right)
\end{aligned}
$$

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

To show $\sqrt{n}T\left(\widehat{\beta}_{CupFM}-\widetilde{\beta}_{CupFM}\right)=o_p(1)$, we need the following lemma.

Lemma A.3 *Under Assumptions of 1-6, we have*

(a) $\sqrt{n}\left(\widehat{\Delta}_{\varepsilon un}^{+}-\Delta_{\varepsilon un}^{+}\right)=o_p(1)$,

(b) $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta_i'\widehat{\Delta}_{\eta u}^{+}-\delta_i'\Delta_{\eta u}^{+}\right)=o_p(1)$,

(c) $\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}\widehat{u}_i^{+}-x_i'M_{F^0}u_i^{+}\right)=o_p(1)$

where $\widehat{u}_{it}^{+}=u_{it}-\widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1}\left(\begin{array}{c}\Delta x_{it}\\ \Delta\widehat{F}_t\end{array}\right)$, $\widehat{\Delta}_{\varepsilon un}^{+}=\frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_{\varepsilon ui}^{+}$ *and* $\Delta_{\varepsilon un}^{+}=\frac{1}{n}\sum_{i=1}^{n}\Delta_{\varepsilon ui}^{+}$.

Note that the lemma holds when the long run variances are replaced by the bar versions. Since the proofs are basically the same (as demonstrated in the proof of Theorem 1), the proof is focused on the variances without the bar.

Proof. 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 Ω_{bi}^{-1}. Following the same lines the proofs of Theorems 9 and 10 of Hannan (1970) (also see similar result of Moon and Perron (2004)), 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. $K\backsim n^{b}$. Then

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

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

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

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

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta_{i}^{\prime}\widehat{\Delta}_{\eta u}^{+}-\delta_{i}^{\prime}\Delta_{\eta u}^{+}\right) &= \left(\frac{1}{n}\sum_{i=1}^{n}\delta_{i}^{\prime}\right)\sqrt{n}\left(\widehat{\Delta}_{\eta u}^{+}-\Delta_{\eta u}^{+}\right)\\
&= O_{p}(1)\,O_{p}\left(\max\left\{\sqrt{\frac{K}{T}},\sqrt{\frac{n}{K^{2q}}}\right\}\right)=o_{p}(1)
\end{aligned}$$

as required for part (b).

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

$$\begin{aligned}
&\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}\widehat{u}_i^{+}-x_i^{'}M_{F^0}u_i^{+}\right)\\
=\;&\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}\widehat{u}_i^{+}-x_i^{'}M_{\widehat{F}}\widetilde{u}_i^{+}+x_i^{'}M_{\widehat{F}}\widetilde{u}_i^{+}-x_i^{'}M_{\widehat{F}}u_i^{+}+x_i^{'}M_{\widehat{F}}u_i^{+}-x_i^{'}M_{F^0}u_i^{+}\right)\\
=\;&\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}\widetilde{u}_i^{+}-x_i^{'}M_{\widehat{F}}u_i^{+}\right)+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}u_i^{+}-x_i^{'}M_{F^0}u_i^{+}\right)+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}\widehat{u}_i^{+}-x_i^{'}M_{\widehat{F}}\widetilde{u}_i^{+}\right)\\
=\;&\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}M_{\widehat{F}}\left(\widetilde{u}_i^{+}-u_i^{+}\right)+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(x_i^{'}M_{\widehat{F}}-x_i^{'}M_{F^0}\right)u_i^{+}+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}M_{\widehat{F}}\left(\widehat{u}_i^{+}-\widetilde{u}_i^{+}\right)\\
=\;&\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}M_{\widehat{F}}\left(\widetilde{u}_i^{+}-u_i^{+}\right)+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}\left(M_{\widehat{F}}-M_{F^0}\right)u_i^{+}+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}M_{\widehat{F}}\left(\widehat{u}_i^{+}-\widetilde{u}_i^{+}\right)\\
=\;&I+II+III.
\end{aligned}$$

From the proof of Proposition 4 in the supplementary appendix,

$$II=\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}x_i^{'}\left(M_{\widehat{F}}-M_{F^0}\right)u_i^{+}=o_p(1)$$

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.

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

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

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

Then we have

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

and

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

For $I_{c.}$, by the Cauchy Schwarz inequality,

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

Similarly,

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

Combining $I_{c.}$ and II_c, we have

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

Recall $K \backsim n^b$ and $\liminf \frac{\log T}{\log n} > 1$ from Assumption 6. It follows that, as in Moon and Perron (2004)

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

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

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

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

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

Let

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

Note that

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

Consider III.

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

We use Lemma 12.3 in Bai (2005) to get

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

It follows that

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

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

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

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

Proof. To save the notations, we only show that results with x_i in place of $\bar{x}_i$ and δ_i in place of of $\bar{\delta}_i$ since the steps are basically the same. In the supplementary appendix, it is shown that (see the proof of Proposition 4)

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

Then

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

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 A.3,

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

Proof of Theorem 3: This follows directly from Proposition A.1. ■

Proof of Proposition 5: In the supplementary appendix, it is shown that

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t-HF_t^0\|^2=T\,O_p(\|\hat{\beta}-\beta^0\|^2)+O_p(\frac{1}{n})+O_p(\frac{1}{T^2}).$$

From $\sqrt{n}T(\hat{\beta}-\beta^0)=O_p(1)$, the first term on the right hand side is $O_p(1/(nT))$, which is dominated by $O(1/n)$. ■

References

[1] Andrews, D. W. K. (2005), "Cross-section Regression with Common Shocks," *Econometrica*, 73, 1551-1585.

[2] Bai, J. (2004), "Estimating Cross-section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics,* 122, 137-183.

[3] Bai, J. (2005), "Panel Data Models with Interactive Fixed Effects", working paper.

[4] Bai, J., and Kao, C. (2006), "On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence," *Contributions to Economic Analysis,* edited by Badi Baltagi, Elsevier, 3-30.

[5] Bai, J., Kao, C., and Ng, S. (2006). "Panel Cointegration with Global Stochastic Trends: Supplementary Appendix. "

[6] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica,* 70, 191-221.

[7] Bai, J., and Ng, S. (2004), "A Panic Attack on Unit Roots and Cointegration," *Econometrica,* 72, 1127-1177.

[8] Baltagi, B. (2005), *Econometric Analysis of Panel Data*, New York, Wiley.

[9] Baltagi, B., and Kao, C. (2000), Nonstationary panels, Cointegration in Panels and Dynamic Panels: A Survey, *Advances in Econometrics*, 15, 7-51.

[10] Bernanke, B. S., and Boivin, J. (2003), "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics*, 50, 525-546

[11] Breitung, J., and Das, S. (2005), "Testing for Unit Roots in Panels with a Factor Structure," forthcoming in *Econometric Theory.*

[12] Breitung, J., and Pesaran, M. H. (2005), "Unit Roots and Cointegration in Panels," forthcoming in L. Matyas, and P. Sevestre, *The Econometrics of Panel Data* (Third Edition), Kluwer Academic Publishers.

[13] Chang, Y. (2002), "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 110, 261-292.

[14] Chang, Y. (2004), "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 120, 263-293.

[15] 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

[16] Gengenbach, C., Palm, F. C., and Urbain, J.-P. (2005a), "Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications of Modelling," working paper.

[17] Gengenbach, C., Palm, F. C., and Urbain, J.-P. (2005b), "Panel Cointegration Testing in the Presence of Common Factors," working paper.

[18] Hannan, E. (1970), *Multiple Time Series*, New York, Wiley.

[19] Hansen, B.E. (1992), Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends. *Journal of Econometrics*, 53, 87-121.

[20] Hsiao, C., (2003), *Analysis of Panel Data*, Cambridge, Cambridge University Press.

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

[22] Moon, H. R., and Perron, B. (2004), "Testing for a Unit Root in Panels with Dynamic Factor," *Journal of Econometrics*, 122, 81-126.

[23] Moon, H. R., and Phillips, P. C. B. (2000), "Estimation of Autoregressive Roots of Near Unit Using Panel Data," *Econometric Theory*, 927-997.

[24] Moon, H. R., and Phillips, P. C. B. (2004), "GMM Estimation of Autoregressive Roots Near Unit with Panel Data," *Econometric Theory*, 927-997.

[25] Park, J. (1990), "Testing for Unit Roots and Cointegration By Variable Addition," *Advances in Econometrics*, ed. by Fomby and Rhodes, 107-133, JAI Press.

[26] Park, J. and Phillips, P. C. B. (1988), "Regressions with Integrated Regressors: Part I," *Econometric Theory*, 468-498.

[27] Pesaran, M. H., and Smith, R. (1995), "Estimation of Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79-114.

[28] Pesaran, M. H., and Yamagata, T. (2006), "Testing Slope Homogeneity in Large Panels," Working paper.

[29] Phillips, P. C. B. (1995), "Fully-modified least squares and autoregression," *Econometrica*, 63, 1023-1079.

[30] Phillips, P. C. B., and Durlauf, S. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.

[31] Phillips, P. C. B., and Hansen, B. E. (1990), "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-125.

[32] Phillips, P. C. B., and Moon, H. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica,* 67, 1057-1111.

[33] Phillips, P. C. B., and Ouliaris, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-193.

[34] Phillips, P. C. B., and Sul, D. (2003), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *Econometrics Journal*, 6, 217-259.

[35] Stock, J. H., and Watson, M. W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

[36] Westerlund, J. (2006), "On the Estimation of Cointegrated Panel Regressions with Common Factors," working paper.

[37] Westerlund, J., and Edgerton, D. (2005), Testing the Barro-Gordon Model in Breaking and Dependent Panels: Evidence from the OECD Countries, working paper.

Table 1: Mean Bias and Standard Deviation of Estimators

	$\sigma_{31} = 0$				$\sigma_{31} = 0.8$				$\sigma_{31} = -0.8$			
	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM
$\sigma_{21} = 0$												
n,T=20	1.352	0.349	0.030	0.030	-0.712	0.257	0.000	0.000	2.216	-0.086	0.030	0.030
	(1.559)	(0.387)	(0.030)	(0.029)	(1.505)	(0.372)	(0.030)	(0.029)	(1.524)	(0.394)	(0.029)	(0.029)
n,T=40	3.371	-0.719	-0.000	-0.000	2.761	-0.246	-0.000	-0.000	1.010	-0.371	-0.000	-0.000
	(1.139)	(0.225)	(0.009)	(0.009)	(1.529)	(0.227)	(0.010)	(0.009)	(1.124)	(0.217)	(0.009)	(0.009)
n,T=60	-2.006	0.094	-0.000	-0.000	-1.393	0.038	-0.000	-0.000	-1.073	0.199	-0.000	-0.000
	(0.920)	(0.138)	(0.005)	(0.005)	(0.915)	(0.139)	(0.005)	(0.005)	(0.929)	(0.138)	(0.005)	(0.005)
n,T=120	0.204	-0.064	-0.000	-0.000	0.548	-0.062	-0.020	0.015	-0.163	-0.061	0.018	-0.000
	(0.645)	(0.056)	(0.018)	(0.002)	(0.646)	(0.056)	(0.002)	(0.002)	(0.643)	(0.056)	(0.002)	(0.002)
$\sigma_{21} = 0.2$												
n,T=20	4.333	0.317	-0.119	0.332	2.258	0.129	-0.158	0.293	4.903	-0.220	-0.117	0.322
	(1.584)	(0.385)	(0.030)	(0.029)	(1.529)	(0.382)	(0.031)	(0.029)	(1.614)	(0.396)	(0.030)	(0.028)
n,T=40	4.567	-0.768	-0.113	0.100	4.051	-0.333	-0.117	0.101	1.964	-0.376	-0.115	0.102
	(1.133)	(0.223)	(0.010)	(0.009)	(1.153)	(0.227)	(0.010)	(0.009)	(1.120)	(0.216)	(0.010)	(0.009)
n,T=60	-1.100	0.109	-0.071	0.045	-0.337	0.082	-0.067	0.049	0.032	0.150	-0.065	0.051
	(0.923)	(0.138)	(0.005)	(0.005)	(0.925)	(0.139)	(0.005)	(0.005)	(0.938)	(0.140)	(0.005)	(0.005)
n,T=120	0.696	-0.059	0.000	0.178	1.161	-0.070	-0.017	0.017	0.151	-0.026	0.017	-0.017
	(0.648)	(0.055)	0.018	(0.002)	(0.649)	(0.055)	(0.002)	(0.002)	(0.646)	(0.055)	(0.002)	(0.002)
$\sigma_{21} = -0.2$												
n,T=20	-1.600	0.376	0.179	-0.274	-3.763	0.331	0.151	-0.291	-0.754	-0.049	0.169	-0.274
	(1.588)	(0.393)	(0.031)	(0.029)	(1.593)	(0.345)	(0.031)	(0.029)	(1.603)	(0.394)	(0.031)	(0.029)
n,T=40	2.086	-0.653	0.105	-0.108	0.812	-0.077	0.101	-0.113	-0.353	-0.313	0.096	-0.112
	(1.144)	(0.225)	(0.010)	(0.009)	(1.141)	(0.223)	(0.010)	(0.009)	(1.128)	(0.218)	(0.010)	(0.009)
n,T=60	-2.850	0.008	0.055	-0.062	-2.178	-0.018	0.058	-0.058	-1.872	0.236	0.056	-0.060
	(0.917)	(0.142)	(0.005)	(0.005)	(0.905)	(0.136)	(0.005)	(0.005)	(0.921)	(0.138)	(0.005)	(0.005)
n,T=120	-0.501	0.000	0.000	0.000	-0.175	-0.000	-0.018	0.017	-0.839	0.029	0.000	-0.000
	(0.650)	(0.057)	(0.002)	(0.018)	(0.646)	(0.057)	(0.002)	(0.002)	(0.654)	(0.058)	(0.002)	(0.002)

Note: (a) The Mean biases here have been multiplied by 100.
(b) $c = 5$, $\sigma_{32} = 0.4$.

Table 2: Mean Bias and Standard Deviation of Estimators for Different n and T

	$c=5$				$c=10$			
(n,T)	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM
(20, 20)	2.258	0.129	-0.158	0.293	1.538	0.275	-0.158	0.294
	(1.594)	(0.382)	(0.031)	(0.028)	(3.186)	(0.771)	(0.031)	(0.029)
(20, 40)	4.832	-0.426	-0.067	0.107	8.141	-0.006	-0.067	0.106
	(1.692)	(0.288)	(0.014)	(0.014)	(3.186)	(0.566)	(0.014)	(0.014)
(20, 60)	0.460	0.282	-0.019	-0.058	-0.105	0.0561	-0.186	0.058
	(1.560)	(0.206)	(0.009)	(0.009)	(3.121)	(0.412)	(0.009)	(0.009)
(20, 120)	3.018	0.040	0.010	0.021	-6.550	0.067	0.010	0.021
	(1.572)	(0.123)	(0.005	(0.005)	(3.144)	(0.245)	(0.005)	(0.004)
(40, 20)	4.012	-0.566	-0.225	0.320	5.092	-1.087	-0.226	0.320
	(1.126)	(0.280)	(0.0218)	(0.019)	(2.252)	(0.593)	(0.021)	(0.019)
(40, 40)	4.051	-0.332	-0.117	0.101	6.616	-0.622	-0.117	0.101
	(1.153)	(0.227)	(0.010)	(0.009)	(2.305)	(0.454)	(0.010)	(0.009)
(40, 60)	1.818	0.114	-0.055	0.051	2.628	0.248	-0.055	0.051
	(1.098)	(0.158)	(0.007)	(0.006)	(2.196)	(0.317)	(0.007)	(0.006)
(40, 120)	1.905	-0.090	-0.010	0.015	3.303	-0.178	-0.010	0.015
	(1.111)	(0.087)	(0.003)	(0.003)	(2.243)	(0.187)	(0.003)	(0.003)
(60, 20)	3.934	-0.317	-0.294	0.295	4.989	-0.544	-0.294	0.295
	(0.921)	(0.249)	(0.018)	(0.017)	(1.841)	(0.497)	(0.014)	(0.016)
(60, 40)	2.023	0.110	-0.125	0.108	2.573	0.267	-0.125	0.109
	(0.923)	(0.187)	(0.009)	(0.008)	(1.296)	(0.027)	(0.009)	(0.008)
(60, 60)	-0.337	0.082	-0.067	0.049	-1.666	0.191	-0.067	0.049
	(0.925)	(0.139)	(0.005)	(0.005)	(1.850)	(0.279)	(0.005)	(0.005)
(60, 120)	-1.168	0.109	-0.015	0.015	-2.839	-0.223	-0.014	0.015
	(0.923)	(0.075)	(0.003)	(0.003)	(1.847)	(0.151)	(0.003)	(0.003)
(120, 20)	2.548	-0.151	-0.304	0.294	2.236	-0.203	-0.304	0.294
	(0.651)	(0.182)	(0.014)	(0.011)	(1.303)	(0.362)	(0.014)	(0.011)
(120, 40)	1.579	-0.026	-0.013	0.001	1.678	0.000	-0.133	0.112
	(0.661)	(0.137)	(0.006)	(0.005)	(1.321)	(0.279)	(0.006)	(0.005)
(120, 60)	0.764	0.004	-0.077	0.013	0.539	0.061	-0.077	0.048
	(0.634)	(0.100)	(0.004)	(0.004)	(1.267)	(0.199)	(0.004)	(0.004)
(120, 120)	1.161	-0.070	-0.017	0.017	1.823	-0.134	-0.017	0.018
	(0.649)	(0.055)	(0.002)	(0.002)	(1.298)	(0.111)	(0.002)	(0.002)

(a) The Mean biases here have been multiplied by 100.
(b) $\sigma_{21} = 0.2$, $\sigma_{31} = 0.8$, and $\sigma_{32} = 0.4$.

Table 3: Mean Bias and Standard Deviation of t-statistics

	$\sigma_{31} = 0$				$\sigma_{31} = 0.8$				$\sigma_{31} = -0.8$			
	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM
$\sigma_{21} = 0$												
n,T=20	0.036	0.006	0.016	0.016	0.006	0.0224	0.001	0.001	0.041	-0.001	0.019	0.019
	(2.414)	(2.445)	(1.531)	(1.502)	(2.527)	(2.449)	(1.529)	(1.503)	(2.534)	(2.455)	(1.515)	(1.491)
n,T=40	0.092	-0.036	-0.007	-0.006	0.074	-0.052	-0.012	-0.011	0.019	0.008	-0.006	-0.005
	(3.576)	(2.589)	(1.276)	(1.256)	(3.592)	(2.618)	(1.273)	(1.254)	(3.588)	(2.581)	(1.278)	(1.217)
n,T=60	-0.098	0.016	-0.019	-0.019	-0.036	-0.016	-0.011	-0.011	-0.060	0.045	-0.009	-0.009
	(4.346)	(2.647)	(1.182)	(1.169)	(4.325)	(2.640)	(1.189)	(1.178)	(4.315)	(2.644)	(1.182)	(1.169)
n,T=120	0.046	-0.019	-0.003	-0.003	0.099	-0.019	-0.075	0.102	-0.088	-0.040	0.068	-0.011
	(6.093)	(2.696)	(1.101)	(1.096)	(6.089)	(2.661)	(1.118)	(1.094)	(6.095)	(2.705)	(1.120)	(1.095)
$\sigma_{21} = 0.2$												
n,T=20	0.104	0.040	0.001	0.185	0.070	0.037	-0.013	0.188	0.105	0.033	0.004	0.181
	(2.508)	(2.454)	(1.558)	(1.497)	(2.529)	(2.453)	(1.561)	(1.442)	(2.539)	(2.465)	(1.543)	(1.483)
n,T=40	0.149	-0.013	-0.081	0.140	0.134	-0.022	-0.085	0.142	0.059	-0.003	-0.081	0.143
	(3.563)	(2.597)	(1.304)	(1.252)	(3.578)	(2.639)	(1.307)	(1.252)	(3.578)	(2.612)	(1.314)	(1.258)
n,T=60	-0.032	0.039	-0.100	0.115	0.027	0.013	-0.094	0.123	0.011	0.038	-0.087	0.127
	(4.357)	(2.651)	(1.209)	(1.167)	(4.357)	(2.647)	(1.215)	(1.174)	(4.325)	(2.646)	(1.204)	(1.162)
n,T=120	0.049	-0.016	0.003	0.002	0.097	-0.019	-0.059	0.114	0.012	-0.029	0.062	-0.109
	(6.060)	(2.640)	(1.096)	(1.092)	(6.084)	(2.645)	(1.115)	(1.093)	(6.043)	(2.635)	(1.111)	(1.089)
$\sigma_{21} = -0.2$												
n,T=20	-0.031	-0.013	0.029	-0.155	-0.064	0.005	0.125	-0.166	-0.031	-0.029	0.027	-0.152
	(2.519)	(2.456)	(1.559)	(1.497)	(2.528)	(2.439)	(1.556)	(1.498)	(2.538)	(2.458)	(1.556)	(1.498)
n,T=40	0.033	-0.068	0.067	-0.153	-0.005	-0.071	0.061	-0.162	-0.035	-0.021	0.058	-0.159
	(3.586)	(2.593)	(1.312)	(1.255)	(3.597)	(2.618)	(1.305)	(1.248)	(3.588)	(2.574)	(1.305)	(1.252)
n,T=60	-0.162	0.002	0.062	-0.154	-0.093	-0.035	0.067	-0.146	-0.114	0.028	0.067	-0.147
	(4.335)	(2.657)	(1.212)	(1.169)	(4.283)	(2.633)	(1.210)	(1.168)	(4.308)	(2.643)	(1.206)	(1.166)
n,T=120	-0.066	0.001	0.007	0.007	-0.010	0.022	-0.062	0.117	-0.111	-0.004	0.077	-0.104
	(6.098)	(2.679)	(1.106)	(1.106)	(6.152)	(2.577)	(1.116)	(1.092)	(6.119)	(2.691)	(1.125)	(1.101)

Note: (a) $c = 5$, $\sigma_{32} = 0.4$.

Table 4: Mean Bias and Standard Deviation of t-statistics for Different n and T

	$c=5$				$c=10$			
(n,T)	LSDV	2sFM	CupBC	CupFM	LSDV	2sFM	CupBC	CupFM
$(20,20)$	0.070	0.037	-0.013	0.169	0.036	0.030	-0.013	0.169
	(2.529)	(2.453)	(1.561)	(1.497)	(2.532)	(2.562)	(1.560)	(1.496)
$(20,40)$	0.130	-0.007	-0.009	0.110	0.106	-0.011	-0.009	0.110
	(3.539)	(1.863)	(1.313)	(1.286)	(3.541)	(1.896)	(1.313)	(1.286)
$(20,60)$	0.029	0.009	0.015	0.085	0.009	0.003	0.016	0.085
	(4.303)	(1.553)	(1.253)	(1.239)	(4.305)	(1.569)	(1.253)	(1.239)
$(20,120)$	-0.090	0.015	0.057	0.064	-0.105	0.013	0.057	0.064
	(6.131)	(1.222)	(1.156)	(1.151)	(6.132)	(1.220)	(1.156)	(1.151)
$(40,20)$	0.119	-0.015	-0.086	0.242	0.073	-0.019	-0.086	0.241
	(2.518)	(3.376)	(1.549)	(1.443)	(2.520)	(3.610)	(1.549)	(1.443)
$(40,40)$	0.134	-0.022	-0.085	0.142	0.100	-0.026	-0.085	0.142
	(3.578)	(2.639)	(1.307)	(1.252)	(3.580)	(2.739)	(1.307)	(1.252)
$(40,60)$	0.113	0.012	-0.048	0.109	0.085	0.008	-0.047	0.109
	(4.328)	(2.164)	(1.209)	(1.177)	(4.329)	(2.222)	(1.209)	(1.176)
$(40,120)$	0.133	-0.014	-0.007	0.059	0.113	-0.019	-0.007	0.059
	(6.097)	(1.519)	(1.131)	(1.123)	(6.098)	(1.535)	(1.131)	(1.123)
$(60,20)$	0.123	0.005	-0.161	0.276	0.067	-0.002	-0.160	0.276
	(2.521)	(4.042)	(1.579)	(1.424)	(2.524)	(4.409)	(1.579)	(1.425)
$(60,40)$	0.100	0.069	-0.109	0.192	0.059	0.065	-0.109	0.192
	(3.532)	(3.206)	(1.352)	(1.272)	(3.534)	(3.375)	(1.352)	(1.272)
$(60,60)$	0.027	0.013	-0.094	0.123	-0.006	0.010	-0.094	0.122
	(4.426)	(2.613)	(1.215)	(1.174)	(4.359)	(2.751)	(1.215)	(1.174)
$(60,120)$	-0.020	0.031	-0.024	0.077	-0.044	0.030	-0.025	0.077
	(6.131)	(1.866)	(1.118)	(1.104)	(6.132)	(1.902)	(1.118)	(1.104)
$(120,20)$	0.139	0.044	-0.243	0.386	0.060	0.063	-0.243	0.386
	(2.478)	(5.269)	(1.681)	(1.404)	(2.479)	(5.969)	(1.681)	(1.404)
$(120,40)$	0.135	0.037	-0.186	0.268	0.078	0.040	-0.186	0.268
	(3.588)	(4.369)	(1.366)	(1.233)	(3.589)	(4.706)	(1.366)	(1.233)
$(120,60)$	0.099	0.011	-0.162	0.174	0.052	0.004	-0.162	0.174
	(4.272)	(3.683)	(1.249)	(1.166)	(4.273)	(3.902)	(1.249)	(1.167)
$(120,120)$	0.097	-0.189	-0.589	0.114	0.063	-0.027	-0.059	0.114
	(6.084)	(2.645)	(1.115)	(1.093)	(6.086)	(2.741)	(1.115)	(1.093)

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