ISSN: 1525-3066

Center for Policy Research
Working Paper No. 75

ON THE ESTIMATION AND INFERENCE OF A PANEL COINTEGRATION MODEL WITH CROSS-SECTIONAL DEPENDENCE

Jushan Bai and Chihwa Kao

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

December 2005

$5.00

CENTER FOR POLICY RESEARCH – Fall 2005

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director,
Budget and Administration

Douglas Holtz-Eakin
Professor of Economics
Associate Director, Center for Policy Research

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi .... Economics
Pablo Beramendi .... Political Science
Dan Black .... Economics
Lloyd Blanchard .... Public Administration
William Duncombe .... Public Administration
Gary Engelhardt .... Economics
Deborah Freund .... Public Administration
Madonna Harrington Meyer .... Sociology
Christine Himes .... Sociology
William C. Horrace .... Economics
Bernard Jump .... Public Administration
Duke Kao .... Economics
Eric Kingson .... Social Work
Thomas Kniesner .... Economics
Tom Krebs .... Economics
Jeff Kubik .... Economics
Andrew London .... Sociology
Len Lopoo .... Public Administration
Jerry Miner .... Economics
John Moran .... Economics
Jan Ondrich .... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder .... Health and Physical Education
Grant Reeher .... Political Science
Stuart Rosenthal .... Economics
Ross Rubenstein .... Public Administration
Margaret Usdansky .... Sociology
Michael Wasylenko .... Economics
Janet Wilmoth .... Sociology

GRADUATE ASSOCIATES

Javier Baez .... Economics
Sonali Ballal .... Public Administration
Dana Balter .... Public Administration
Jason Bladen .... Economics
Jesse Bricker .... Economics
Maria Brown .... Social Science
Yong Chen .... Economics
Ginnie Cronin .... Social Science
Ana Dammert .... Economics
Mike Eriksen .... Economics
Katie Fitzpatrick .... Economics
Alexandre Genest .... Public Administration
Julie Anna Golebiewski .... Economics
Nadia Greenhalgh-Stanley .... Economics
Yue Hu .... Economics
Becky Lafrancois .... Economics
Joseph Marchand .... Economics
Larry Miller .... Public Administration
Wendy Parker .... Sociology
Emily Pas .... Economics
Shawn Rohlin .... Economics
Cynthia Searcy .... Public Administration
Talisha Searcy .... Public Administration
Jeff Thompson .... Economics
Wen Wang .... Economics
Yoshikun Yamamoto .... Public Administration

STAFF

Kelly Bogart .... Administrative Secretary
Martha Bonney .... Publications/Events Coordinator
Karen Cimilluca .... Librarian/Office Coordinator
Kim Desmond .... Administrative Secretary
Kati Foley .... Administrative Assistant, LIS
Kitty Nasto .... Administrative Secretary
Candi Patterson .... Computer Consultant
Mary Santy .... Administrative Secretary

On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence

Jushan Bai
Department of Economics
New York University
New York, NY 10003
USA
and
Department of Economics
Tsinghua University
Beijing 10084
China
Jushan.Bai@nyu.edu

Chihwa Kao
Center for Policy Research
and
Department of Economics
Syracuse University
Syracuse, NY 13244-1020
USA
cdkao@maxwell.syr.edu

First Draft: March 2003
This Draft: May 2, 2005

Abstract

Most of the existing literature on panel data cointegration assumes cross-sectional independence, an assumption that is difficult to satisfy. This paper studies panel cointegration under cross-sectional dependence, which is characterized by a factor structure. We derive the limiting distribution of a fully modified estimator for the panel cointegrating coefficients. We also propose a continuous-updated fully modified (CUP-FM) estimator. Monte Carlo results show that the CUP-FM estimator has better small sample properties than the two-step FM (2S-FM) and OLS estimators.

1 Introduction

A convenient but difficult to justify assumption in panel cointegration analysis is cross-sectional independence. Left untreated, cross-sectional dependence causes bias and inconsistency estimation, as argued by Andrews (2003). In this paper, we use a factor structure to characterize cross-sectional dependence. Factors models are especially suited for this purpose. One major source of cross-section correlation in macroeconomic data is common shocks, e.g., oil price shocks and international financial crises. Common shocks drive the underlying comovement of economic variables. Factor models provide an effective way to extract the comovement and have been used in various studies.[1] Cross-sectional correlation exists even in micro level data because of herd behavior (fashions, fads, and imitation cascades) either at firm level or household level. The general state of an economy (recessions or booms) also affects household decision making. Factor models accommodate individual's different responses to common shocks through heterogeneous factor loadings.

Panel data models with correlated cross-sectional units are important due to increasing availability of large panel data sets and increasing interconnectedness of the economies. Despite the immense interest in testing for panel unit roots and cointegration,[2] not much attention has been paid to the issues of cross-sectional dependence. Studies using factor models for nonstationary data include Bai and Ng (2004), Bai (2004), Phillips and Sul (2003), and Moon and Perron (2004). Chang (2002) proposed to use a nonlinear IV estimation to construct a new panel unit root test. Hall et al (1999) considered a problem of determining the number of common trends. Baltagi et al. (2004) derived several Lagrange Multiplier tests for the panel data regression model with spatial error correlation. Robertson and Symons (2000), Coakley et al. (2002) and Pesaran (2004) proposed to use common factors to capture the cross-sectional dependence in stationary panel models. All these studies focus on either stationary data or panel unit root studies rather than panel cointegration.

This paper makes three contributions. First, it adds to the literature by suggesting a factor model for panel cointegrations. Second, it proposes a continuous-updated fully modified (CUP-FM) estimator. Third, it provides a comparison for the finite sample properties

[1] For example, Stock and Watson (2002), Gregory and Head (1999), Forni and Reichlin (1998) and Forni et al. (2000) to name a few.

[2] see Baltagi and Kao (2000) for a recent survey

of the OLS, two-step fully modified (2S-FM), CUP-FM estimators.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 presents assumptions. Sections 4 and 5 develop the asymptotic theory for the OLS and fully modified (FM) estimators. Section 6 discusses a feasible FM estimator and suggests a CUP-FM estimator. Section 7 makes some remarks on hypothesis testing. Section 8 presents Monte Carlo results to illustrate the finite sample properties of the OLS and FM estimators. Section 9 summarizes the findings. The appendix contains the proofs of Lemmas and Theorems.

The following notations are used in the paper. We write the integral $\int_0^1 W(s)ds$ as $\int W$ when there is no ambiguity over limits. We define $\Omega^{1/2}$ to be any matrix such that $\Omega = \left(\Omega^{1/2}\right)\left(\Omega^{1/2}\right)'$. We use $\|A\|$ to denote $\left\{tr\left(A'A\right)\right\}^{1/2}$, $|A|$ to denote the determinant of A, $\Rightarrow$ to denote weak convergence, $\xrightarrow{p}$ to denote convergence in probability, $[x]$ to denote the largest integer $\leq x$, $I(0)$ and $I(1)$ to signify a time-series that is integrated of order zero and one, respectively, and $BM(\Omega)$ to denote Brownian motion with the covariance matrix Ω. We let $M < \infty$ be a generic positive number, not depending on T or n.

2 The Model

Consider the following fixed effect panel regression:

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}, \; i = 1, ..., n, \; t = 1, ..., T, \tag{1}$$

where y_{it} is 1×1, β is a $1 \times k$ vector of the slope parameters, α_i is the intercept, and e_{it} is the stationary regression error. We assume that x_{it} is a $k \times 1$ integrated processes of order one for all i, where

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

Under these specifications, (1) describes a system of cointegrated regressions, i.e., y_{it} is cointegrated with x_{it}. The initialization of this system is $y_{i0} = x_{i0} = O_p(1)$ as $T \to \infty$ for all i. The individual constant term α_i can be extended into general deterministic time trends such as $\alpha_{0i} + \alpha_{1i}t +, ..., +\alpha_{pi}t$ or other deterministic component. To model the cross-sectional dependence we assume the error term, e_{it}, follows a factor model (e.g., Bai and Ng, 2002, 2004):

$$e_{it} = \lambda_i' F_t + u_{it}, \tag{2}$$

where F_t is a $r \times 1$ vector of common factors, λ_i is a $r \times 1$ vector of factor loadings and u_{it} is the idiosyncratic component of e_{it}, which means

$$E\left(e_{it}e_{jt}\right) = \lambda_i^{'} E\left(F_t F_t^{'}\right) \lambda_j$$

i.e., e_{it} and e_{jt} are correlated due to the common factors F_t.

Remark 1 *1. We could also allow ε_{it} to have a factor structure such that*

$$\varepsilon_{it} = \gamma_i^{'} F_t + \eta_{it}.$$

Then we can use $\triangle x_{it}$ to estimate F_t and γ_i. Or we can use e_{it} together with $\triangle x_{it}$ to estimate F_t, λ_i and γ_i. In general, ε_{it} can be of the form

$$\varepsilon_{it} = \gamma_i' F_t + \tau_i' G_t + \eta_{it},$$

where F_t and G_t are zero mean processes, and η_{it} are usually independent over i and t.

3 Assumptions

Our analysis is based on the following assumptions.

Assumption 1 *As $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n} \lambda_i \lambda_i^{'} \longrightarrow \Sigma_\lambda$, a $r \times r$ positive definite matrix.*

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

Assumption 3 *F_t and u_{it} are independent; u_{it} are independent across i.*

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_{it} \Rightarrow B\left(\Omega_i\right) \text{ as } T \to \infty \text{ for all } i, \tag{3}$$

where

$$B_i = \begin{bmatrix} B_F \\ B_{ui} \\ B_{\varepsilon i} \end{bmatrix}.$$

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

$$\begin{aligned} \Omega_i &= \sum_{j=-\infty}^{\infty} E\left(w_{i0}w_{ij}^{'}\right) \\ &= \Pi_i(1)\Sigma_v\Pi_i(1)^{'} \\ &= \Sigma_i + \Gamma_i + \Gamma_i^{'} \\ &= \begin{bmatrix} \Omega_{Fi} & \Omega_{Fui} & \Omega_{F\varepsilon i} \\ \Omega_{uFi} & \Omega_{ui} & \Omega_{u\varepsilon i} \\ \Omega_{\varepsilon Fi} & \Omega_{\varepsilon ui} & \Omega_{\varepsilon i} \end{bmatrix}, \end{aligned}$$

where

$$\Gamma_i = \sum_{j=1}^{\infty} E\left(w_{i0}w_{ij}^{'}\right) = \begin{bmatrix} \Gamma_{Fi} & \Gamma_{Fui} & \Gamma_{F\varepsilon i} \\ \Gamma_{uFi} & \Gamma_{ui} & \Gamma_{u\varepsilon i} \\ \Gamma_{\varepsilon Fi} & \Gamma_{\varepsilon ui} & \Gamma_{\varepsilon i} \end{bmatrix} \tag{4}$$

and

$$\Sigma_i = E\left(w_{i0}w_{i0}^{'}\right) = \begin{bmatrix} \Sigma_{Fi} & \Sigma_{Fui} & \Sigma_{F\varepsilon i} \\ \Sigma_{uFi} & \Sigma_{ui} & \Sigma_{u\varepsilon i} \\ \Sigma_{\varepsilon Fi} & \Sigma_{\varepsilon ui} & \Sigma_{\varepsilon i} \end{bmatrix}$$

are partitioned conformably with w_{it}. We denote

$$\Omega = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_i,$$

$$\Gamma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i,$$

and

$$\Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Sigma_i.$$

Assumption 4 $\Omega_{\varepsilon i}$ *is non-singular, i.e.,* $\{x_{it}\}$ *are not cointegrated.*

Define

$$\Omega_{bi} = \begin{bmatrix} \Omega_{Fi} & \Omega_{Fui} \\ \Omega_{uFi} & \Omega_{ui} \end{bmatrix}, \quad \Omega_{b\varepsilon i} = \begin{bmatrix} \Omega_{F\varepsilon i} \\ \Omega_{u\varepsilon i} \end{bmatrix}$$

and

$$\Omega_{b.\varepsilon i} = \Omega_{bi} - \Omega_{b\varepsilon i}\Omega_{\varepsilon i}^{-1}\Omega_{\varepsilon bi}.$$

Then, B_i can be rewritten as

$$B_i = \begin{bmatrix} B_{bi} \\ B_{\varepsilon i} \end{bmatrix} = \begin{bmatrix} \Omega_{b.\varepsilon i}^{1/2} & \Omega_{b\varepsilon i}\Omega_{\varepsilon i}^{-1/2} \\ 0 & \Omega_{\varepsilon i}^{1/2} \end{bmatrix} \begin{bmatrix} V_{bi} \\ W_i \end{bmatrix}, \tag{5}$$

where

$$B_{bi} = \begin{bmatrix} B_F \\ B_{ui} \end{bmatrix},$$

$$V_{bi} = \begin{bmatrix} V_F \\ V_{ui} \end{bmatrix},$$

and

$$\begin{bmatrix} V_{bi} \\ W_i \end{bmatrix} = BM\left(I\right)$$

is a standardized Brownian motion. Define the one-sided long-run covariance

$$\begin{aligned} \Delta_i &= \Sigma_i + \Gamma_i \\ &= \sum_{j=0}^{\infty} E\left(w_{i0}w_{ij}^{'}\right) \end{aligned}$$

with

$$\Delta_i = \begin{bmatrix} \Delta_{bi} & \Delta_{b\varepsilon i} \\ \Delta_{\varepsilon bi} & \Delta_{\varepsilon i} \end{bmatrix}.$$

Remark 2

1. *Assumption 1 is a standard assumption in factor models (e.g., Bai and Ng 2002, 2004) to ensure the factor structure is identifiable. We only consider nonrandom factor loadings for simplicity. Our results still hold when the $\lambda_i^{'}s$ are random, provided they are independent of the factors and idiosyncratic errors, and $E\left\|\lambda_i\right\|^4 \leq M$.*

2. *Assumption 2 assumes that the random factors, F_t, and idiosyncratic shocks $\left(u_{it}, \varepsilon_{it}^{'}\right)$ are stationary linear processes. Note that F_t and ε_{it} are allowed to be correlated. In particular, ε_{it} may have a factor structure as in Remark 1.*

3. *Assumption of independence made in Assumption 3 between F_t and u_{it} can be relaxed following Bai and Ng (2002). Nevertheless, independence is not a restricted assumption since cross-sectional correlations in the regression errors e_{it} are taken into account by the common factors.*

4 OLS

Let us first study the limiting distribution of the OLS estimator for equation (1). The OLS estimator of β is

$$\widehat{\beta}_{OLS}=\left[\sum_{i=1}^{n}\sum_{t=1}^{T}y_{it}\left(x_{it}-\overline{x}_{i}\right)^{'}\right]\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{'}\right]^{-1}. \tag{6}$$

Theorem 1 *Under Assumptions* $1-4$*, we have*
$\sqrt{n}T\left(\widehat{\beta}_{OLS}-\beta\right)-\sqrt{n}\delta_{nT}\Rightarrow N\left(0,6\Omega_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^{'}\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\right\}\Omega_{\varepsilon}^{-1}\right),$
as $(n,T\to\infty)$ *with* $\frac{n}{T}\to 0$ *where*

$$\begin{aligned}\delta_{nT} &= \frac{1}{n}\left[\sum_{i=1}^{n}\lambda_i^{'}\left(\Omega_{F\varepsilon i}\Omega_{\varepsilon i}^{1/2}\left(\int\widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon i}^{-1/2}+\Delta_{F\varepsilon i}\right)+\Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{1/2}\left(\int\widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon i}^{-1/2}+\Delta_{u\varepsilon i}\right]\\ &\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_{it}-x_{it}\right)\left(x_{it}-\overline{x}_i\right)^{'}\right]^{-1},\end{aligned}$$

$\widetilde{W}_i=W_i-\int W_i$ *and* $\Omega_{\varepsilon}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{\varepsilon i}$.

Remark 3 *It is also possible to construct the bias-corrected OLS by using the averages of the long run covariances. Note*

$$\begin{aligned}E\left[\delta_{nT}\right] &\simeq \frac{1}{n}\left[\sum_{i=1}^{n}\lambda_i^{'}\left(-\frac{1}{2}\Omega_{F\varepsilon i}+\Delta_{F\varepsilon i}\right)-\frac{1}{2}\Omega_{u\varepsilon i}+\Delta_{u\varepsilon i}\right]\left(\frac{1}{6}\Omega_{\varepsilon}\right)^{-1}\\ &= \frac{1}{n}\left[\sum_{i=1}^{n}\left(-\frac{1}{2}\right)\left(\lambda_i^{'}\Omega_{F\varepsilon i}+\Omega_{u\varepsilon i}\right)+\lambda_i^{'}\Delta_{F\varepsilon i}+\Delta_{u\varepsilon i}\right]\left(\frac{1}{6}\Omega_{\varepsilon}\right)^{-1}\\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(-\frac{1}{2}\right)\lambda_i^{'}\Omega_{F\varepsilon i}+\frac{1}{n}\sum_{i=1}^{n}\Omega_{u\varepsilon i}+\frac{1}{n}\sum_{i=1}^{n}\lambda_i^{'}\Delta_{F\varepsilon i}+\frac{1}{n}\sum_{i=1}^{n}\Delta_{u\varepsilon i}\right)\left(\frac{1}{6}\Omega_{\varepsilon}\right)^{-1}\end{aligned}$$

It can be shown by a central limit theorem that

$$\sqrt{n}\left(\delta_{nT} - E\left[\delta_{nT}\right]\right) \Rightarrow N\left(0, B\right)$$

for some B. Therefore,

$$\begin{aligned} & \sqrt{n}T\left(\widehat{\beta}_{OLS} - \beta\right) - \sqrt{n}E\left[\delta_{nT}\right] \\ = \; & \sqrt{n}T\left(\widehat{\beta}_{OLS} - \beta\right) - \sqrt{n}\delta_{nT} + \sqrt{n}\left(\delta_{nT} - E\left[\delta_{nT}\right]\right) \\ \Rightarrow \; & N\left(0, A\right) \end{aligned}$$

for some A.

5 FM Estimator

Next we examine the limiting distribution of the FM estimator, $\widehat{\beta}_{FM}$. The FM estimator was suggested by Phillips and Hansen (1990) in a different context (non-panel data). The FM estimator is constructed by making corrections for endogeneity and serial correlation to the OLS estimator $\widehat{\beta}_{OLS}$ in (6). The endogeneity correction is achieved by modifying the variable y_{it} in (1) with the transformation

$$y_{it}^{+} = y_{it} - \left(\lambda_i^{\prime}\Omega_{F\varepsilon i} + \Omega_{u\varepsilon i}\right)\Omega_{\varepsilon i}^{-1}\Delta x_{it}.$$

The serial correlation correction term has the form

$$\begin{aligned} \Delta_{b\varepsilon i}^{+} &= \Delta_{b\varepsilon i} - \Omega_{b\varepsilon i}\Omega_{\varepsilon i}^{-1}\Delta_{\varepsilon i}, \\ &= \begin{bmatrix} \Delta_{F\varepsilon i}^{+} \\ \Delta_{u\varepsilon i}^{+} \end{bmatrix}. \end{aligned}$$

Therefore, the infeasible FM estimator is

$$\widetilde{\beta}_{FM} = \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T} y_{it}^{+}\left(x_{it} - \overline{x}_i\right)^{\prime} - T\left(\lambda_i^{\prime}\Delta_{F\varepsilon i}^{+} + \Delta_{u\varepsilon i}^{+}\right)\right)\right]\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \overline{x}_i\right)\left(x_{it} - \overline{x}_i\right)^{\prime}\right]^{-1}. \tag{7}$$

Now, we state the limiting distribution of $\widetilde{\beta}_{FM}$.

Theorem 2 *Let Assumptions* $1-4$ *hold. Then as* $(n,T\to\infty)$ *with* $\frac{n}{T}\to 0$

$$\sqrt{n}T\left(\widetilde{\beta}_{FM}-\beta\right)\Rightarrow N\left(0,6\Omega_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^{'}\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\right\}\Omega_{\varepsilon}^{-1}\right).$$

Remark 4 *The asymptotic distribution in Theorem 2 is reduced to*

$$\sqrt{n}T\left(\widetilde{\beta}_{FM}-\beta\right)\Rightarrow \mathrm{N}\left(0,6\Omega_{\varepsilon}^{-1}\left(\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\lambda_i^2\right)\Omega_{F.\varepsilon}+\Omega_{u.\varepsilon}\right)\right)$$

if the long-run covariances are the same across the cross-sectional unit i *and* $r=1$.

6 Feasible FM

In this section we investigate the limiting distribution of the feasible FM. We will show that the limiting distribution of the feasible FM is not affected when λ_i, $\Omega_{\varepsilon i}$, $\Omega_{\varepsilon bi}$, $\Omega_{\varepsilon i}$, and $\Delta_{\varepsilon bi}$ are estimated. To estimate λ_i, we use the method of principal components used in Stock and Watson (2002). Let $\lambda=(\lambda_1,\lambda_2,...,\lambda_n)^{'}$ and $F=(F_1,F_2,...,F_T)^{'}$. The method of principal components of λ and F minimizes

$$V(r)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widehat{e}_{it}-\lambda_i^{'}F_t\right)^2$$

where

$$\begin{aligned}\widehat{e}_{it} &= y_{it}-\widehat{\alpha}_i-\widehat{\beta}x_{it}\\ &= (y_{it}-\overline{y}_i)-\widehat{\beta}(x_{it}-\overline{x}_i),\end{aligned}$$

with a consistent estimator $\widehat{\beta}$. Concentrating out λ and using the normalization that $F^{'}F/T=I_r$, the optimization problem is identical to maximizing $tr\left(F^{'}\left(ZZ^{'}\right)F\right)$, where $Z=(\widehat{e}_1,\widehat{e}_2,...,\widehat{e}_n)$ is $T\times n$ with $\widehat{e}_i=(\widehat{e}_{i1},\widehat{e}_{i2},...,\widehat{e}_{iT})^{'}$. The estimated factor matrix, denoted by $\widehat{F}$, a $T\times r$ matrix, is $\sqrt{T}$ times eigenvectors corresponding to the r largest eigenvalues of the $T\times T$ matrix $ZZ^{'}$, and

$$\begin{aligned}\widehat{\lambda}^{'} &= \left(\widehat{F}^{'}\widehat{F}\right)^{-1}\widehat{F}^{'}Z\\ &= \frac{\widehat{F}^{'}Z}{T}\end{aligned}$$

is the corresponding matrix of the estimated factor loadings. It is known that the solution to the above minimization problem is not unique, i.e., λ_i and F_t are not directly identifiable but they are identifiable up to a transformation H. For our setup, knowing $H\lambda_i$ is as good as knowing λ_i. For example in (7) using $\lambda_i^{'}\Delta_{F\varepsilon i}^{+}$ will give the same information as using $\lambda_i^{'}H^{'}H^{'-1}\Delta_{F\varepsilon i}^{+}$ since $\Delta_{F\varepsilon i}^{+}$ is also identifiable up to a transformation, i.e., $\lambda_i^{'}H^{'}H^{'-1}\Delta_{F\varepsilon i}^{+} = \lambda_i^{'}\Delta_{F\varepsilon i}^{+}$. Therefore, when assessing the properties of the estimates we only need to consider the differences in the space spanned by, say, between $\widehat{\lambda}_i$ and λ_i.

Define the feasible FM, $\widehat{\beta}_{FM}$, with $\widehat{\lambda}_i$, $\widehat{F}_t$, $\widehat{\Sigma}_i$, and $\widehat{\Omega}_i$ in place of λ_i, F_t, Σ_i, and Ω_i,

$$\widehat{\beta}_{FM} = \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\widehat{y}_{it}^{+}\left(x_{it}-\overline{x}_i\right)^{'} - T\left(\widehat{\lambda}_i^{'}\widehat{\Delta}_{F\varepsilon i}^{+} + \widehat{\Delta}_{u\varepsilon i}^{+}\right)\right)\right]\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)\left(x_{it}-\overline{x}_i\right)^{'}\right]^{-1},$$

where

$$\widehat{y}_{it}^{+} = y_{it} - \left(\widehat{\lambda}_i^{'}\widehat{\Omega}_{F\varepsilon i} + \widehat{\Omega}_{u\varepsilon i}\right)\widehat{\Omega}_{\varepsilon i}^{-1}\Delta x_{it}.$$

and $\widehat{\Delta}_{F\varepsilon i}^{+}$ and $\widehat{\Delta}_{u\varepsilon i}^{+}$ are defined similarly.

Assume that $\Omega_i = \Omega$ for all i. Let

$$e_{it}^{+} = e_{it} - \left(\lambda_i^{'}\Omega_{F\varepsilon} + \Omega_{u\varepsilon}\right)\Omega_{\varepsilon}^{-1}\Delta x_{it},$$

$$\widehat{\Delta}_{b\varepsilon n}^{+} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_{b\varepsilon i}^{+},$$

and

$$\Delta_{b\varepsilon n}^{+} = \frac{1}{n}\sum_{i=1}^{n}\Delta_{b\varepsilon i}^{+}.$$

Then

$$
\begin{aligned}
& \sqrt{n}T\left(\widehat{\beta}_{FM}-\widetilde{\beta}_{FM}\right) \\
= & \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left\{\left(\sum_{t=1}^{T}\widehat{e}_{it}^{+}\left(x_{it}-\overline{x}_{i}\right)^{\prime}-T\widehat{\Delta}_{b\varepsilon n}^{+}\right)-\left(\sum_{t=1}^{T}e_{it}^{+}\left(x_{it}-\overline{x}_{i}\right)^{\prime}-T\Delta_{b\varepsilon n}^{+}\right)\right\} \\
& \left[\frac{1}{nT^{2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{\prime}\right]^{-1} \\
= & \left[\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\left(\widehat{e}_{it}^{+}-e_{it}^{+}\right)\left(x_{it}-\overline{x}_{i}\right)^{\prime}-T\left(\widehat{\Delta}_{b\varepsilon n}^{+}-\Delta_{b\varepsilon n}^{+}\right)\right)\right] \\
& \left[\frac{1}{nT^{2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_{i}\right)\left(x_{it}-\overline{x}_{i}\right)^{\prime}\right]^{-1}.
\end{aligned}
$$

Before we prove Theorem 3 we need the following lemmas.

Lemma 1 *Under Assumptions 1-4* $\sqrt{n}\left(\widehat{\Delta}_{b\varepsilon n}^{+}-\Delta_{b\varepsilon n}^{+}\right)=o_{p}(1)$.

Lemma 1 can be proved similarly by following Phillips and Moon (1999) and Moon and Perron (2004).

Lemma 2 *Suppose Assumptions 1-4 hold. There exists an* H *with rank* r *such that as* $(n,T\rightarrow\infty)$
(i)

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\lambda}_{i}-H\lambda_{i}\right\|^{2}=O_{p}\left(\frac{1}{\delta_{nT}^{2}}\right)$$

(ii) let c_i $(i=1,2,...,n)$ *be a sequence of random matrices such that* $c_i=O_p(1)$ *and* $\frac{1}{n}\sum_{i=1}^{n}\|c_i\|^2=O_p(1)$ *then*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\lambda}_{i}-H\lambda_{i}\right)^{\prime}c_{i}=O_{p}\left(\frac{1}{\delta_{nT}^{2}}\right)$$

where $\delta_{nT}=\min\left\{\sqrt{n},\sqrt{T}\right\}$.

Bai and Ng (2002) showed that for a known $\widehat{e}_{it}$ that the average squared deviations between $\widehat{\lambda}_i$ and $H\lambda_i$ vanish as n and T both tend to infinity and the rate of convergence is the minimum of n and T. Lemma 2 can be proved similarly by following Bai and Ng (2002) that parameter estimation uncertainty for β has no impact on the null limit distribution of $\widehat{\lambda}_i$.

Lemma 3 *Under Assumptions 1-4*

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widehat{e}_{it}^{+}-e_{it}^{+}\right)\left(x_{it}-\overline{x}_i\right)'=o_p(1)$$

as $(n,T\to\infty)$ *and* $\frac{\sqrt{n}}{T}\to 0$.

Then we have the following theorem:

Theorem 3 *Under Assumptions 1-4 and* $(n,T\to\infty)$ *and* $\frac{\sqrt{n}}{T}\to 0$

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

In the literature, the FM-type estimators usually were computed with a two-step procedure, by assuming an initial consistent estimate of β, say $\widehat{\beta}_{OLS}$. Then, one constructs estimates of the long-run covariance matrix, $\widehat{\Omega}^{(1)}$, and loading, $\widehat{\lambda}_i^{(1)}$. The 2S-FM, denoted $\widehat{\beta}_{2S}^{(1)}$ is obtained using $\widehat{\Omega}^{(1)}$ and $\widehat{\lambda}_i^{(1)}$:

$$\begin{aligned}\widehat{\beta}_{2S}^{(1)} &= \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\widehat{y}_{it}^{+(1)}\left(x_{it}-\overline{x}_i\right)'-T\left(\widehat{\lambda}_i^{\prime(1)}\widehat{\Delta}_{F\varepsilon i}^{+(1)}+\widehat{\Delta}_{u\varepsilon i}^{+(1)}\right)\right)\right] \\ &\quad \left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)\left(x_{it}-\overline{x}_i\right)'\right]^{-1}. \end{aligned} \tag{8}$$

In this paper, we propose a CUP-FM estimator. The CUP-FM is constructed by estimating parameters and long-run covariance matrix and loading recursively. Thus $\widehat{\beta}_{FM}$, $\widehat{\Omega}$ and $\widehat{\lambda}_i$ are estimated repeatedly, until convergence is reached. In the Section 8, we find the CUP-FM has a superior small sample properties as compared with the 2S-FM, i.e., CUP-FM has smaller bias than the common 2S-FM estimator. The CUP-FM is defined as

$$\begin{aligned}\widehat{\beta}_{CUP} &= \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\widehat{y}_{it}^{+}\left(\widehat{\beta}_{CUP}\right)\left(x_{it}-\overline{x}_i\right)'-T\left(\widehat{\lambda}_i'\left(\widehat{\beta}_{CUP}\right)\widehat{\Delta}_{F\varepsilon i}^{+}\left(\widehat{\beta}_{CUP}\right)+\widehat{\Delta}_{u\varepsilon i}^{+}\left(\widehat{\beta}_{CUP}\right)\right)\right)\right] \\ &\quad \left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\overline{x}_i\right)\left(x_{it}-\overline{x}_i\right)'\right]^{-1}. \end{aligned} \tag{9}$$

Remark 5 *1. In this paper, we assume the number of factors, r, is known. Bai and Ng (2002) showed that the number of factors can be found by minimizing the following:*

$$IC(k) = \log\left(V\left(k\right)\right) + k\left(\frac{n+T}{nT}\right)\log\left(\frac{nT}{n+T}\right).$$

2. Once the estimates of w_{it}, $\widehat{w}_{it} = \left(\widehat{F}_t^{'},\ \widehat{u}_{it},\ \triangle x_{it}^{'}\right)^{'}$, were estimated, we used

$$\widehat{\Sigma} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{w}_{it}\widehat{w}_{it}^{'} \tag{10}$$

to estimate Σ, where

$$\widehat{u}_{it} = \widehat{e}_{it} - \widehat{\lambda}_i^{'}\widehat{F}_t.$$

Ω was estimated by

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T}\widehat{w}_{it}\widehat{w}_{it}^{'} + \frac{1}{T}\sum_{\tau=1}^{l}\varpi_{\tau l}\sum_{t=\tau+1}^{T}\left(\widehat{w}_{it}\widehat{w}_{it-\tau}^{'} + \widehat{w}_{it-\tau}\widehat{w}_{it}^{'}\right)\right\}, \tag{11}$$

where $\varpi_{\tau l}$ is a weight function or a kernel. Using Phillips and Moon (1999), $\widehat{\Sigma}_i$ and $\widehat{\Omega}_i$ can be shown to be consistent for Σ_i and Ω_i.

7 Hypothesis Testing

We now consider a linear hypothesis that involves the elements of the coefficient vector β. We show that hypothesis tests constructed using the FM estimator have asymptotic chi-squared distributions. The null hypothesis has the form:

$$H_0 : R\beta = r, \tag{12}$$

where r is a $m \times 1$ known vector and R is a known $m \times k$ matrix describing the restrictions. A natural test statistic of the Wald test using $\widehat{\beta}_{FM}$ is

$$W = \frac{1}{6}nT^2\left(R\widehat{\beta}_{FM} - r\right)^{'}\left[6\widehat{\Omega}_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\lambda}_i^{'}\widehat{\Omega}_{F.\varepsilon i}\widehat{\lambda}_i\widehat{\Omega}_{\varepsilon i} + \widehat{\Omega}_{u.\varepsilon i}\widehat{\Omega}_{\varepsilon i}\right)\right\}\widehat{\Omega}_{\varepsilon}^{-1}\right]^{-1}\left(R\widehat{\beta}_{FM} - r\right). \tag{13}$$

It is clear that W converges in distribution to a chi-squared random variable with k degrees of freedom, χ_k^2, as $(n, T \to \infty)$ under the null hypothesis. Hence, we establish the following theorem:

Theorem 4 *If Assumptions* 1–4 *hold, then under the null hypothesis (12), with* $(n, T \to \infty)$, $W \Rightarrow \chi_k^2$,

Remark 6 *1. One common application of the theorem 6 is the single-coefficient test: one of the coefficient is zero;* $\beta_j = \beta_0$,

$$R = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}$$

and $r = 0$. *We can construct a* t*-statistic*

$$t_j = \frac{\sqrt{n}T\left(\widehat{\beta}_{jFM} - \beta_0\right)}{s_j} \tag{14}$$

where

$$s_j^2 = \left[6\widehat{\Omega}_\varepsilon^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\lambda}_i^{'}\widehat{\Omega}_{F.\varepsilon i}\widehat{\lambda}_i\widehat{\Omega}_{\varepsilon i} + \widehat{\Omega}_{u.\varepsilon i}\widehat{\Omega}_{\varepsilon i}\right)\right\}\widehat{\Omega}_\varepsilon^{-1}\right]_{jj},$$

the j*th diagonal element of*

$$\left[6\widehat{\Omega}_\varepsilon^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\lambda}_i^{'}\widehat{\Omega}_{F.\varepsilon i}\widehat{\lambda}_i\widehat{\Omega}_{\varepsilon i} + \widehat{\Omega}_{u.\varepsilon i}\widehat{\Omega}_{\varepsilon i}\right)\right\}\widehat{\Omega}_\varepsilon^{-1}\right]$$

It follows that

$$t_j \Rightarrow N(0, 1). \tag{15}$$

2. General nonlinear parameter restriction such as $H_0 : h(\beta) = 0$, *where* $h(\cdot)$ *is* $k^* \times 1$ *vector of smooth functions such that* $\frac{\partial h}{\partial \beta'}$ *has full rank* k^* *can be conducted in a similar fashion as in Theorem 6. Thus, the Wald test has the following form*

$$W_h = nT^2 h\left(\widehat{\beta}_{FM}\right)^{'}\widehat{V}_h^{-1}h\left(\widehat{\beta}_{FM}\right)$$

where

$$\widehat{V}_h^{-1} = \left(\frac{\partial h\left(\widehat{\beta}_{FM}\right)}{\partial \beta^{'}}\right)\widehat{V}_\beta^{-1}\left(\frac{\partial h\left(\widehat{\beta}_{FM}^{'}\right)}{\partial \beta}\right)$$

and

$$\widehat{V}_{\beta}=6\widehat{\Omega}_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\lambda}_i^{'}\widehat{\Omega}_{F.\varepsilon i}\widehat{\lambda}\widehat{\Omega}_{\varepsilon ii}+\widehat{\Omega}_{u.\varepsilon i}\widehat{\Omega}_{\varepsilon i}\right)\right\}\widehat{\Omega}_{\varepsilon}^{-1}. \tag{16}$$

It follows that

$$W_h \Rightarrow \chi^2_{k^*}$$

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

8 Monte Carlo Simulations

In this section, we conduct Monte Carlo experiments to assess the finite sample properties of OLS and FM estimators. The simulations were performed by a Sun SparcServer 1000 and an Ultra Enterprise 3000. GAUSS 3.2.31 and COINT 2.0 were used to perform the simulations. Random numbers for error terms, $(F_t^*, u_{it}^*, \varepsilon_{it}^*)$ were generated by the GAUSS procedure RNDNS. At each replication, we generated an $n(T+1000)$ length of random numbers and then split it into n series so that each series had the same mean and variance. The first $1,000$ observations were discarded for each series. $\{F_t^*\}$, $\{u_{it}^*\}$ and $\{\varepsilon_{it}^*\}$ were constructed with $F_t^*=0$, $u_{i0}^*=0$ and $\varepsilon_{i0}^*=0$.

To compare the performance of the OLS and FM estimators we conducted Monte Carlo experiments based on a design which is similar to Kao and Chiang (2000)

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}$$

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

and

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

for $i=1,...,n$, $t=1,...,T$, where

$$\begin{pmatrix} F_t \\ u_{it} \\ \varepsilon_{it}\end{pmatrix}=\begin{pmatrix} F_t^* \\ u_{it}^* \\ \varepsilon_{it}^*\end{pmatrix}+\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.3 & -0.4 \\ \theta_{31} & \theta_{32} & 0.6\end{pmatrix}\begin{pmatrix} F_{t-1}^* \\ u_{it-1}^* \\ \varepsilon_{it-1}^*\end{pmatrix} \tag{17}$$

with

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

For this experiment, we have a single factor ($r = 1$) and λ_i are generated from i.i.d. $N(\mu_\lambda, 1)$. We let $\mu_\lambda = 0.1$. We generated α_i from a uniform distribution, $U[0, 10]$, and set $\beta = 2$. From Theorems 1-3 we know that the asymptotic results depend upon variances and covariances of $F_{t,}$ u_{it} and ε_{it}. Here we set $\sigma_{12} = 0$. The design in (17) is a good one since the endogeneity of the system is controlled by only four parameters, θ_{31}, θ_{32}, σ_{31} and σ_{32}. We choose $\theta_{31} = 0.8$, $\theta_{32} = 0.4$, $\sigma_{31} = -0.8$ and $\theta_{32} = 0.4$.

The estimate of the long-run covariance matrix in (11) was obtained by using the procedure KERNEL in COINT 2.0 with a Bartlett window. The lag truncation number was set arbitrarily at five. Results with other kernels, such as Parzen and quadratic spectral kernels, are not reported, because no essential differences were found for most cases.

Next, we recorded the results from our Monte Carlo experiments that examined the finite-sample properties of (a) the OLS estimator, $\widehat{\beta}_{OLS}$ in (6), (b) the 2S-FM estimator, $\widehat{\beta}_{2S}$, in (8), (c) the two-step naive FM estimator, $\widehat{\beta}_{FM}^{b}$, proposed by Kao and Chiang (2000) and Phillips and Moon (1999), (d) the CUP-FM estimator $\widehat{\beta}_{CUP}$, in (9) and (e) the CUP naive FM estimator $\widehat{\beta}_{FM}^{d}$ which is similar to the two-step naive FM except the iteration goes beyond two steps. The naive FM estimators are obtained assuming the cross-sectional independence. The maximum number of the iteration for CUP-FM estimators is set to 20. The results we report are based on $1,000$ replications and are summarized in Tables 1 - 4. All the FM estimators were obtained by using a Bartlett window of lag length five as in (11).

Table 1 reports the Monte Carlo means and standard deviations (in parentheses) of $\left(\widehat{\beta}_{OLS} - \beta\right)$, $\left(\widehat{\beta}_{2S} - \beta\right)$, $\left(\widehat{\beta}_{FM}^{b} - \beta\right)$, $\left(\widehat{\beta}_{CUP} - \beta\right)$,and $\left(\widehat{\beta}_{FM}^{d} - \beta\right)$ for sample sizes $T = n$ $= (20, 40, 60)$. The biases of the OLS estimator, $\widehat{\beta}_{OLS}$, decrease at a rate of T. For example, with $\sigma_\lambda = 1$ and $\sigma_F = 1$, the bias at $T = 20$ is -0.045, at $T = 40$ is -0.024, and at $T = 60$ is -0.015. Also, the biases stay the same for different values of σ_λ and σ_F.

TABLE 1 ABOUT HERE

While we expected the OLS estimator to be biased, we expected FM estimators to produce better estimates. However, it is noticeable that the 2S-FM estimator still has a downward bias for all values of σ_λ and σ_F, though the biases are smaller. In general, the 2S-FM estimator presents the same degree of difficulty with bias as does the OLS estimator. This is probably due to the failure of the nonparametric correction procedure.

In contrast, the results in Table 1 show that the CUP-FM, is distinctly superior to the OLS and 2S-FM estimators for all cases in terms of the mean biases. Clearly, the CUP-FM outperforms both the OLS and 2S-FM estimators.

TABLE 2 ABOUT HERE

It is important to know the effects of the variations in panel dimensions on the results, since the actual panel data have a wide variety of cross-section and time-series dimensions. Table 2 considers 16 different combinations for n and T, each ranging from 20 to 120 with $\sigma_{31} = -0.8$, $\sigma_{21} = -0.4$, $\theta_{31} = 0.8$, and $\theta_{21} = 0.4$. First, we notice that the cross-section dimension has no significant effect on the biases of all estimators. From this it seems that in practice the T dimension must exceed the n dimension, especially for the OLS and 2S-FM estimators, in order to get a good approximation of the limiting distributions of the estimators. For example, for OLS estimator in Table 2, the reported bias, -0.008, is substantially less for $(T = 120, n = 40)$ than it is for either $(T = 40, n = 40)$, (the bias is -0.024), or $(T = 40, n = 120)$, (the bias is -0.022). The results in Table 2 again confirm the superiority of the CUP-FM.

TABLE 3 ABOUT HERE

Monte Carlo means and standard deviations of the t-statistic, $t_{\beta=\beta_0}$, are given in Table 3. Here, the OLS t-statistic is the conventional t-statistic as printed by standard statistical packages. With all values of σ_λ and σ_F with the exception $\sigma_\lambda = \sqrt{10}$, the CUP-FM t-statistic is well approximated by a standard N(0,1) suggested from the asymptotic results. The CUP-FM t-statistic is much closer to the standard normal density than the OLS t-statistic and the 2S-FM t-statistic. The 2S-FM t-statistic is not well approximated by a standard N(0,1).

TABLE 4 ABOUT HERE

Table 4 shows that both the OLS t-statistic and the FM t-statistics become more negatively biased as the dimension of cross-section n increases. The heavily negative biases of the 2S-FM t-statistic in Tables 3-4 again indicate the poor performance of the 2S-FM estimator. For the CUP-FM, the biases decrease rapidly and the standard errors converge to 1.0 as T increases.

It is known that when the length of time series is short the estimate $\widehat{\Omega}$ in (11) may be sensitive to the length of the bandwidth. In Tables 2 and 4, we first investigate the sensitivity of the FM estimators with respect to the choice of length of the bandwidth. We extend the experiments by changing the lag length from 5 to other values for a Barlett window. Overall, the results (not reported here) show that changing the lag length from 5 to other values does not lead to substantial changes in biases for the FM estimators and their t-statistics.

9 Conclusion

A factor approach to panel models with cross-sectional dependence is useful when both the time series and cross-sectional dimensions are large. This approach also provides significant reduction in the number of variables that may cause the cross-sectional dependence in panel data. In this paper, we study the estimation and inference of a panel cointegration model with cross-sectional dependence. The paper contributes to the growing literature on panel data with cross-sectional dependence by (i) discussing limiting distributions for the OLS and FM estimators, (ii) suggesting a CUP-FM estimator and (iii) investigating the finite sample proprieties of the OLS, CUP-FM and 2S-FM estimators. It is found that the 2S-FM and OLS estimators have a non-negligible bias in finite samples, and that the CUP-FM estimator improves over the other two estimators.

Acknowledgements

We thank Badi Baltagi, Yu-Pin Hu, Giovanni Urga, Kamhon Kan, Chung-Ming Kuan, Hashem Pesaran, Lorenzo Trapani and Yongcheol Shin for helpful comments. We also thank seminar participants at Academia Sinica, National Taiwan University, Syracuse University, Workshop on Recent Developments in the Econometrics of Panel Data in London, March 2004 and the European Meeting of the Econometric Society in Madrid, August 2004 for helpful comments and suggestions. Jushan Bai acknowledges financial support from the NSF (grant SES-0137084).

Appendix

Let

$$B_{nT} = \left[\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)^{'} \right].$$

Note

$$\begin{aligned} &\sqrt{n}T\left(\widehat{\beta}_{OLS} - \beta\right) \\ &= \left[\sqrt{n}\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(\tfrac{1}{T}\sum_{t=1}^{T} e_{it}(x_{it} - \overline{x}_i)^{'}\right)\right]\left[\tfrac{1}{n}\tfrac{1}{T^2}B_{nT}\right]^{-1} \\ &= \left[\sqrt{n}\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\zeta_{1iT}\right]\left[\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{2iT}\right]^{-1} \\ &= \sqrt{n}\xi_{1nT}\left[\xi_{2nT}\right]^{-1}, \end{aligned}$$

where $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\zeta_{1iT} = \frac{1}{T}\sum_{t=1}^{T} e_{it}(x_{it} - \overline{x}_i)^{'}$, $\zeta_{2iT} = \frac{1}{T^2}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)^{'}$, $\xi_{1nT} = \frac{1}{n}\sum_{i=1}^{n}\zeta_{1iT}$, and $\xi_{2nT} = \frac{1}{n}\sum_{i=1}^{n}\zeta_{2iT}$. Before going into the next theorem, we need to consider some preliminary results.

Define $\Omega_\varepsilon = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{\varepsilon i}$ and

$$\theta^n = \frac{1}{n}\left[\sum_{i=1}^{n}\lambda_i^{'}\left(\Omega_{F.\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon i}^{1/2} + \Delta_{F\varepsilon i}\right) + \Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int \widetilde{W}_i dW_i^{'}\right)\Omega_{\varepsilon i}^{1/2} + \Delta_{u\varepsilon i}\right].$$

If Assumptions $1-4$ hold, then

Lemma 4 *(a) As* $(n, T\to\infty)$,

$$\frac{1}{n}\frac{1}{T^2}B_{nT} \xrightarrow{p} \frac{1}{6}\Omega_\varepsilon.$$

(b) As $(n, T\to\infty)$ *with* $\frac{n}{T}\to 0$,

$$\sqrt{n}\left(\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} e_{it}(x_{it} - \overline{x}_i)^{'} - \theta^n\right) \Rightarrow N\left(0, \frac{1}{6}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\lambda_i^{'}\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i} + \Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right\}\right)$$

Proof. *(a) and (b) can be shown easily by following the Theorem 8 in Phillips and Moon (1999).* ■

A Proof of Theorem 1

Proof.

Recall that

$$\begin{aligned}
&\sqrt{n}T\left(\widehat{\beta}_{OLS}-\beta\right)-\sqrt{n}\tfrac{1}{n}\left[\begin{array}{c}\sum_{i=1}^{n}\lambda_i'\left(\Omega_{F\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{F\varepsilon i}\right)\\ +\Omega_{\varepsilon u i}\Omega_{\varepsilon i}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{\varepsilon u i}\end{array}\right]\\
&\left[\tfrac{1}{n}\tfrac{1}{T^2}B_{nT}\right]^{-1}\\
&=\left[\sqrt{n}\tfrac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\zeta_{1iT}-\lambda_i'\left(\Omega_{F\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{F\varepsilon i}\right)\\ -\Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{u\varepsilon i}\end{array}\right\}\right]\left[\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{2iT}\right]^{-1}\\
&=\left[\sqrt{n}\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{1iT}^{*}\right]\left[\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{2iT}\right]^{-1}\\
&=\sqrt{n}\xi_{1nT}^{*}\left[\xi_{2nT}\right]^{-1},
\end{aligned}$$

where

$$\zeta_{1iT}^{*}=\zeta_{1iT}-\lambda_i'\left(\Omega_{F\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{F\varepsilon i}\right)-\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1/2}\left(\int\widetilde{W}_i dW_i'\right)\Omega_{\varepsilon}^{1/2}+\Delta_{u\varepsilon}$$

and

$$\xi_{1nT}^{*}=\frac{1}{n}\sum_{i=1}^{n}\zeta_{1iT}^{*}.$$

First, we note from Lemma 4 (b) that

$$\sqrt{n}\xi_{1nT}^{*}\Rightarrow N\left(0,\frac{1}{6}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\lambda_i'\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right\}\right)$$

as $(n,T\to\infty)$ and $\frac{n}{T}\to 0$. Using the Slutsky theorem and (a) from Lemma 4, we obtain

$$\sqrt{n}\xi_{1nT}^{*}\left[\xi_{2nT}\right]^{-1}\Rightarrow N\left(0,6\Omega_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i'\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\right\}\Omega_{\varepsilon}^{-1}\right).$$

Hence,

$$\begin{aligned}
&\sqrt{n}T\left(\widehat{\beta}_{OLS}-\beta\right)-\sqrt{n}\delta_{nT}\\
&\Rightarrow N\left(0,6\Omega_{\varepsilon}^{-1}\left\{\lim_{n\to\infty}\tfrac{1}{n}\sum_{i=1}^{n}\left(\lambda_i'\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\right\}\Omega_{\varepsilon}^{-1}\right),
\end{aligned}\tag{18}$$

proving the theorem, where

$$\delta_{nT} = \frac{1}{n}\left[\sum_{i=1}^{n}\lambda_i'\left(\Omega_{F.\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int \widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{F\varepsilon i}\right)+\Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{-1/2}\left(\int \widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{1/2}+\Delta_{u\varepsilon i}\right]\left[\frac{1}{n}\frac{1}{T^2}B_{nT}\right]^{-1}.$$

Therefore, we established Theorem 1. ■

B Proof of Theorem 2

Proof. Let

$$F_{it}^{+} = F_t - \Omega_{F\varepsilon i}\Omega_{\varepsilon i}^{-1}\varepsilon_{it},$$

and

$$u_{it}^{+} = u_{it} - \Omega_{u\varepsilon i}\Omega_{\varepsilon i}^{-1}\varepsilon_{it}.$$

The FM estimator of β can be rewritten as follows

$$\begin{aligned}\widetilde{\beta}_{FM} &= \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T} y_{it}^{+}\left(x_{it}-\overline{x}_i\right)' - T\left(\lambda_i'\Delta_{F\varepsilon i}^{+}+\Delta_{u\varepsilon i}^{+}\right)\right)\right]B_{nT}^{-1}\\ &= \beta + \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\left(\lambda_i'F_{it}^{+}+u_{it}^{+}\right)\left(x_{it}-\overline{x}_i\right)' - T\left(\lambda_i'\Delta_{F\varepsilon i}^{+}+\Delta_{u\varepsilon i}^{+}\right)\right)\right]B_{nT}^{-1}.\end{aligned} \tag{19}$$

First, we rescale $\left(\widetilde{\beta}_{FM}-\beta\right)$ by $\sqrt{n}T$

$$\begin{aligned}\sqrt{n}T\left(\widetilde{\beta}_{FM}-\beta\right) &= \sqrt{n}\tfrac{1}{n}\sum_{i=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T}\left[\left(\lambda_i'F_{it}^{+}+u_{it}^{+}\right)\left(x_{it}-\overline{x}_i\right)' - \lambda_i'\Delta_{F\varepsilon i}^{+} - \Delta_{u\varepsilon i}^{+}\right]\left[\tfrac{1}{n}\tfrac{1}{T^2}B_{nT}\right]^{-1}\\ &= \left[\sqrt{n}\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{1iT}^{**}\right]\left[\tfrac{1}{n}\sum_{i=1}^{n}\zeta_{2iT}\right]^{-1}\\ &= \sqrt{n}\xi_{1nT}^{**}\left[\xi_{2nT}\right]^{-1},\end{aligned} \tag{20}$$

where $\zeta_{1iT}^{**} = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\lambda_i'F_{it}^{+}+\widehat{u}_{it}^{+}\right)\left(x_{it}-\overline{x}_i\right)' - \lambda_i'\Delta_{F\varepsilon i}^{+} - \Delta_{u\varepsilon i}^{+}\right]$, and $\xi_{1nT}^{**} = \frac{1}{n}\sum_{i=1}^{n}\zeta_{1iT}^{**}$.

Modifying the Theorem 11 in Phillips and Moon (1999) and Kao and Chiang (2000) we can show that as $(n, T\to\infty)$ with $\frac{n}{T}\to 0$

$$\sqrt{n}\left(\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i'F_{it}^{+}\left(x_{it}-\overline{x}_i\right)' - \lambda_i'\Delta_{F\varepsilon i}^{+}\right)\right) \Rightarrow N\left(0, \frac{1}{6}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\lambda_i'\Omega_{F.\varepsilon i}\lambda_i\Omega_{\varepsilon i}\right)$$

and

$$\sqrt{n}\left(\frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widehat{u}_{it}^{+}\left(x_{it}-\overline{x}_{i}\right)^{\prime}-\Delta_{u\varepsilon i}^{+}\right)\right)\Rightarrow N\left(0,\frac{1}{6}\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)$$

and combing this with the Assumption 4 that F_t and u_{it} are independent and Lemma 4(a) yields

$$\sqrt{n}T\left(\widetilde{\beta}_{FM}-\beta\right)\Rightarrow N\left(0,6\Omega_{\varepsilon}^{-1}\left\{\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_{i}^{\prime}\Omega_{F.\varepsilon i}\lambda_{i}\Omega_{\varepsilon i}+\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\right\}\Omega_{\varepsilon}^{-1}\right)$$

as required. ■

C Proof of Lemma 3

Proof. We note that λ_i is estimating $H\lambda_i$, and $\widehat{\Omega}_{F\varepsilon}$ is estimating $H^{-1\prime}\widehat{\Omega}_{F\varepsilon}$. Thus $\widehat{\lambda}_{i}^{\prime}\widehat{\Omega}_{F\varepsilon}$ is estimating $\lambda_{i}^{\prime}\Omega_{F\varepsilon}$, which is the object of interest. For the purpose of notational simplicity, we shall assume H being a $r\times r$ identify matrix in our proof below. From

$$\widehat{e}_{it}^{+}=e_{it}-\left(\widehat{\lambda}_{i}^{\prime}\widehat{\Omega}_{F\varepsilon}+\widehat{\Omega}_{u\varepsilon}\right)\widehat{\Omega}_{\varepsilon}^{-1}\Delta x_{it}$$

and

$$e_{it}^{+}=e_{it}-\left(\lambda_{i}^{\prime}\Omega_{F\varepsilon}+\Omega_{u\varepsilon}\right)\Omega_{\varepsilon}^{-1}\Delta x_{it},$$

$$\begin{aligned}\widehat{e}_{it}^{+}-e_{it}^{+} &= -\left[\left\{\left(\widehat{\lambda}_{i}^{\prime}\widehat{\Omega}_{F\varepsilon}+\widehat{\Omega}_{u\varepsilon}\right)\widehat{\Omega}_{\varepsilon}^{-1}-\left(\lambda_{i}^{\prime}\Omega_{F\varepsilon}+\Omega_{u\varepsilon}\right)\Omega_{\varepsilon}^{-1}\right\}\Delta x_{it}\right] \\ &= -\left[\left\{\widehat{\lambda}_{i}^{\prime}\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}-\lambda_{i}^{\prime}\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1}+\widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}-\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1}\right\}\Delta x_{it}\right].\end{aligned}$$

Then,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}-\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it}-\overline{x}_{i}\right)^{\prime} \\ &= \left(\widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}-\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1}\right)\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta x_{it}\left(x_{it}-\overline{x}_{i}\right)^{\prime} \\ &= o_{p}(1)O_{p}(1) \\ &= o_{p}(1)\end{aligned}$$

because

$$\widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1} - \Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1} = o_p\left(1\right)$$

and

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta x_{it}\left(x_{it} - \overline{x}_i\right)' = O_p(1).$$

Thus

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widehat{e}_{it}^{+} - e_{it}^{+}\right)\left(x_{it} - \overline{x}_i\right)'$$

$$\begin{aligned}
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left[\left\{\left(\lambda_i'\Omega_{F\varepsilon} + \Omega_{u\varepsilon}\right)\Omega_{\varepsilon}^{-1} - \left(\widehat{\lambda}_i'\widehat{\Omega}_{F\varepsilon} + \widehat{\Omega}_{u\varepsilon}\right)\widehat{\Omega}_{\varepsilon}^{-1}\right\}\Delta x_{it}\right]\left(x_{it} - \overline{x}_i\right)' \\
&= \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i'\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1} - \widehat{\lambda}_i'\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it} - \overline{x}_i\right)' \\
&\quad + \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\Omega_{u\varepsilon}\Omega_{\varepsilon}^{-1} - \widehat{\Omega}_{u\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it} - \overline{x}_i\right)' \\
&= \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i'\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1} - \widehat{\lambda}_i'\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it} - \overline{x}_i\right)' + o_p(1).
\end{aligned}$$

The remainder of the proof needs to show that

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i'\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1} - \widehat{\lambda}_i'\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it} - \overline{x}_i\right)' = o_p\left(1\right).$$

We write A for $\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1}$ and $\widehat{A}$ for $\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}$ respectively and then

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i^{'}\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1}-\widehat{\lambda}_i^{'}\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'} \\
= & \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i^{'}A-\widehat{\lambda}_i^{'}\widehat{A}\right)\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'} \\
= & \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\lambda_i^{'}\left(A-\widehat{A}\right)+\left(\lambda_i^{'}-\widehat{\lambda}_i^{'}\right)\widehat{A}\right]\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'} \\
= & \frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\lambda_i^{'}\left(A-\widehat{A}\right)\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'} \\
& +\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i^{'}-\widehat{\lambda}_i^{'}\right)\widehat{A}\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'} \\
= & I+II, \text{ say.}
\end{aligned}$$

Term I is a row vector. Let I_j be the jth component of I. Let ℓ_j be the jth column of an identity matrix so that $\ell_j=(0,...,0,1,0,...0)'$. Left multiply I by ℓ_j to obtain the jth component, which is scalar and thus is equal to its trace. That is

$$\begin{aligned}
I_j &= tr\left[\left(A-\widehat{A}\right)\left(\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\lambda_i^{'}\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'}\ell_j\lambda_i\right)\right] \\
&= tr\left[(A-\widehat{A})O_p(1)\right] \\
&= o_p(1)O_p(1) \\
&= o_p(1)
\end{aligned}$$

because $\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\lambda_i^{'}\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'}\ell_j\lambda_i=O_p(1)$ and $A-\widehat{A}=o_p(1)$.

Next consider II. Let $c_i=\widehat{A}\frac{1}{T}\sum_{i=1}^{n}\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'}$. Then $c_i=O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\|c_i\|^2=$

$O_p(1)$, thus by Lemma 2 (ii), we have

$$\begin{aligned} II &\leq \sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i^{'}-\widehat{\lambda}_i^{'}\right)c_i\right| \\ &= \sqrt{n}O_p\left(\frac{1}{\delta_{nT}^2}\right) \\ &= \sqrt{n}O_p\left(\frac{1}{\min[n,T]}\right) \\ &= O_p\left(\frac{\sqrt{n}}{\min\{n,T\}}\right) \\ &= O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{n}}{T}\right) \\ &= o_p(1) \end{aligned}$$

since $(n, T \to \infty)$ and $\frac{\sqrt{n}}{T} \to 0$. This establishes

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\lambda_i^{'}\Omega_{F\varepsilon}\Omega_{\varepsilon}^{-1}-\widehat{\lambda}_i^{'}\widehat{\Omega}_{F\varepsilon}\widehat{\Omega}_{\varepsilon}^{-1}\right)\Delta x_{it}\left(x_{it}-\overline{x}_i\right)^{'}=o_p(1).$$

and proves Lemma 3. ■

References

[1] Andrews, D. W. K., 2003. Cross-section regression with common shocks. forthcoming in Econometrica.

[2] Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135-171.

[3] Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137-183.

[4] Bai, J., and Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

[5] Bai, J., and Ng, S., 2004. A panic attack on unit roots and cointegration. Econometrica 72, 1127-1177.

[6] Baltagi, B., and Kao, C., 2000. Nonstationary panels, cointegration in panels and dynamic panels: A survey. Advances in Econometrics 15, 7-51.

[7] Baltagi, B., Song, S. H., and Koh, W., 2004. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123-150.

[8] Chang, Y., 2002. Nonlinear IV unit root tests in panels with cross-sectional dependency. Journal of Econometrics 116, 261-292.

[9] Coakley, J., Fuerts, A., and Smith, R. P., 2002. A principal components approach to cross-section dependence in panels. Manuscript, Birckbeck College, University of London.

[10] Forni, M. and Reichlin, L., 1998. Let's get real: a factor-analytic approach to disaggregated business cycle dynamics. Review of Economic Studies 65, 453-473.

[11] Forni, M., Hallin, M., Lippi, M., and Reichlin, L., 2000. Reference cycles: The NBER methodology revisited. CEPR Discussion Paper 2400.

[12] Gregory, A., and Head, A., 1999. Common and country-specific fluctuations in productivity, investment, and the current account. Journal of Monetary Economics 44, 423-452.

[13] Hall, S. G., Lazarova, S., and Urga, G., 1999. A principle component analysis of common stochastic trends in heterogeneous panel data: Some monte carlo evidence. Oxford Bulletin of Economics and Statistics 61, 749-767.

[14] Kao, C., and Chiang, M. H., 2000. On the Estimation and Inference of a Cointegrated Regression in Panel Data. Advances in Econometrics 15, 179-222.

[15] Moon, H. R., and Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122, 81-126.

[16] Pesaran, H., 2004. Estimation and inference in large heterogenous panels with a multifactor error structure. Manuscript, Trinity College, Cambridge.

[17] 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.

[18] Phillips, P. C. B., and Moon, H., 1999. Linear regression limit theory for nonstationary panel data," Econometrica 67, 1057-1111.

[19] Phillips, P. C. B., and Sul, D., 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econometric Journal 6, 217-259.

[20] Robertson, D., and Symon, J., 2000. Factor residuals in SUR regressions: Estimating panels allowing for cross sectional correlation. Manuscript, Faculty of Economics and Politics, University of Cambridge.

[21] 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.

Table 1: Means Biases and Standard Deviation of OLS and FM Estimato

	$\sigma_\lambda = 1$					$\sigma_\lambda = \sqrt{10}$						
	OLS	FM^a	FM^b	FM^c	FM^d	OLS	FM^a	FM^b	FM^c	FM^d	OLS	F
$\sigma_F = 1$												
T=20	-0.045	-0.025	-0.029	-0.001	-0.006	-0.046	-0.025	-0.029	-0.001	-0.006	-0.045	-0
	(0.029)	(0.028)	(0.029)	(0.034)	(0.030)	(0.059)	(0.054)	(0.059)	(0.076)	(0.060)	(0.026)	(0.
T=40	-0.024	-0.008	-0.011	-0.002	-0.005	-0.024	-0.009	-0.012	-0.003	-0.005	-0.024	-0
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.020)	(0.019)	(0.019)	(0.021)	(0.018)	(0.009)	(0.
T=60	-0.015	-0.004	-0.005	-0.001	-0.003	-0.015	-0.003	-0.005	-0.001	-0.002	-0.015	-0
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.005)	(0.
$\sigma_F = \sqrt{10}$												
T=20	-0.054	-0.022	-0.036	0.011	-0.005	-0.057	-0.024	-0.038	0.013	-0.003	-0.054	-0
	(0.061)	(0.054)	(0.061)	(0.078)	(0.062)	(0.176)	(0.156)	(0.177)	(0.228)	(0.177)	(0.046)	(0.
T=40	-0.028	-0.007	-0.015	0.001	-0.007	-0.030	-0.009	-0.017	-0.001	-0.009	-0.028	-0
	(0.021)	(0.019)	(0.019)	(0.021)	(0.019)	(0.059)	(0.054)	(0.057)	(0.061)	(0.053)	(0.016)	(0.
T=60	-0.018	-0.002	-0.007	0.001	-0.004	-0.017	-0.001	-0.006	0.002	-0.003	-0.018	-0
	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.032)	(0.029)	(0.030)	(0.031)	(0.029)	(0.009)	(0.
$\sigma_F = \sqrt{0.5}$												
T=20	-0.044	-0.025	-0.028	-0.003	-0.006	-0.045	-0.026	-0.028	-0.002	-0.006	-0.044	-0
	(0.026)	(0.026)	(0.026)	(0.030)	(0.028)	(0.045)	(0.041)	(0.045)	(0.056)	(0.046)	(0.024)	(0.
T=40	-0.023	-0.009	-0.010	-0.003	-0.004	-0.023	-0.009	-0.011	-0.003	-0.005	-0.023	-0
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.016)	(0.015)	(0.015)	(0.016)	(0.014)	(0.009)	(0.
T=60	-0.015	-0.004	-0.005	-0.001	-0.003	-0.015	-0.004	-0.005	-0.001	-0.002	-0.015	-0
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.005)	(0.

Note: (a) FM^a is the 2S-FM, FM^b is the naive 2S-FM, FM^c is the CUP-FM and FM^d is the naive CUP-FM
(b) $\mu_\lambda = 0.1$, $\sigma_{31} = -0.8$, $\sigma_{21} = -0.4$, $\theta_{31} = 0.8$, and $\theta_{21} = 0.4$.

Table 2: Means Biases and Standard Deviation of OLS and FM Estimators for Different n and T

(n,T)	OLS	FM[a]	FM[b]	FM[c]	FM[d]
(20, 20)	-0.045	-0.019	-0.022	-0.001	-0.006
	(0.029)	(0.028)	(0.029)	(0.034)	(0.030)
(20, 40)	-0.024	-0.006	-0.009	-0.001	-0.004
	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)
(20, 60)	-0.017	-0.004	-0.006	-0.001	-0.003
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
(20, 120)	-0.008	-0.001	-0.002	-0.000	-0.001
	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
(40, 20)	-0.044	-0.018	-0.021	-0.002	-0.006
	(0.021)	(0.019)	(0.019)	(0.023)	(0.021)
(40, 40)	-0.024	-0.007	-0.009	-0.002	-0.004
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
(40, 60)	-0.015	-0.003	-0.005	-0.001	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
(40, 120)	-0.008	-0.001	-0.002	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
(60, 20)	-0.044	-0.018	-0.022	-0.002	-0.007
	(0.017)	(0.016)	(0.016)	(0.019)	(0.017)
(60, 40)	-0.022	-0.006	-0.008	-0.002	-0.004
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
(60, 60)	-0.015	-0.003	-0.005	-0.001	-0.003
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
(60, 120)	-0.008	-0.001	-0.002	-0.001	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
(120, 20)	-0.044	-0.018	-0.022	-0.002	-0.007
	(0.013)	(0.011)	(0.012)	(0.013)	(0.012)
(120, 40)	-0.022	-0.006	-0.008	-0.002	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
(120, 60)	-0.015	-0.003	-0.005	-0.001	-0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
(120, 120)	-0.008	-0.001	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

(a) $\mu_\lambda = 0.1$, $\sigma_{31} = -0.8$, $\sigma_{21} = -0.4$, $\theta_{31} = 0.8$, and $\theta_{21} = 0.4$.

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

			$\sigma_\lambda = 1$					$\sigma_\lambda = \sqrt{10}$				
	OLS	FM^a	FM^b	FM^c	FM^d	OLS	FM^a	FM^b	FM^c	FM^d	OLS	F
$\sigma_F = 1$												
T=20	-1.994	-1.155	-1.518	-0.056	-0.285	-0.929	-0.546	-0.813	-0.006	-0.122	-2.248	-1
	(1.205)	(1.267)	(1.484)	(1.283)	(1.341)	(1.149)	(1.059)	(1.495)	(1.205)	(1.254)	(1.219)	(1.
T=40	-2.915	-0.941	-1.363	-0.227	-0.559	-1.355	-0.465	-0.766	-0.128	-0.326	-3.288	-1
	(1.202)	(1.101)	(1.248)	(1.054)	(1.141)	(1.127)	(0.913)	(1.207)	(0.912)	(1.049)	(1.221)	(1.
T=60	-3.465	-0.709	-1.158	-0.195	-0.574	-1.552	-0.308	-0.568	-0.074	-0.261	-3.926	-0
	(1.227)	(1.041)	(1.177)	(0.996)	(1.100)	(1.146)	(0.868)	(1.113)	(0.851)	(1.016)	(1.244)	(1.
$\sigma_F = \sqrt{10}$												
T=20	-1.078	-0.484	-0.984	0.180	-0.096	-0.373	-0.154	-0.350	0.085	-0.006	-1.427	-0
	(1.147)	(1.063)	(1.501)	(1.220)	(1.271)	(1.119)	(0.987)	(1.508)	(1.194)	(1.223)	(1.163)	(1.
T=40	-1.575	-0.355	-0.963	0.042	-0.407	-0.561	-0.152	-0.397	-0.014	-0.190	-2.082	-0
	(1.131)	(0.917)	(1.214)	(0.926)	(1.063)	(1.097)	(0.844)	(1.179)	(0.871)	(1.008)	(1.154)	(0.
T=60	-1.809	-0.155	-0.776	0.111	-0.390	-0.588	-0.041	-0.247	0.049	-0.111	-2.424	-0
	(1.158)	(0.879)	(1.131)	(0.867)	(1.035)	(1.108)	(0.812)	(1.078)	(0.811)	(0.983)	(1.192)	(0.
$\sigma_F = \sqrt{0.5}$												
T=20	-2.196	-1.319	-1.606	-0.137	-0.327	-1.203	-0.734	-1.008	-0.054	-0.176	-2.367	-1
	(1.219)	(1.325)	(1.488)	(1.307)	(1.362)	(1.164)	(1.112)	(1.488)	(1.217)	(1.273)	(1.231)	(1.
T=40	-3.214	-1.093	-1.415	-0.311	-0.576	-1.752	-0.619	-0.922	-0.188	-0.385	-3.462	-1
	(1.226)	(1.057)	(1.155)	(1.104)	(1.169)	(1.148)	(0.962)	(1.222)	(0.944)	(1.087)	(1.236)	(1.
T=60	-3.839	-0.868	-1.217	-0.296	-0.602	-2.037	-0.446	-0.712	-0.139	-0.331	-4.149	-0
	(1.239)	(1.088)	(1.183)	(1.037)	(1.112)	(1.169)	(0.908)	(1.131)	(0.881)	(1.038)	(1.249)	(1.

Note: (a) FM^a is the 2S-FM, FM^b is the naive 2S-FM, FM^c is the CUP-FM and FM^d is the naive CUP- FM.
(b) $\mu_\lambda = 0.1$, $\sigma_{31} = -0.8$, $\sigma_{21} = -0.4$, $\theta_{31} = 0.8$, and $\theta_{21} = 0.4$.

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

(n,T)	OLS	FM[a]	FM[b]	FM[c]	FM[d]
(20, 20)	-1.994	-0.738	-1.032	-0.056	-0.286
	(1.205)	(1.098)	(1.291)	(1.283)	(1.341)
(20, 40)	-2.051	-0.465	-0.725	-0.105	-0.332
	(1.179)	(0.999)	(1.126)	(1.046)	(1.114)
(20, 60)	-2.129	-0.404	-0.684	-0.162	-0.421
	(1.221)	(0.963)	(1.278)	(0.983)	(1.111)
(20, 120)	-2.001	-0.213	-0.456	-0.095	-0.327
	(1.222)	(0.923)	(1.083)	(0.931)	(1.072)
(40, 20)	-2.759	-1.017	-1.404	-0.103	-0.402
	(1.237)	(1.116)	(1.291)	(1.235)	(1.307)
(40, 40)	-2.915	-0.699	-1.075	-0.227	-0.559
	(1.202)	(1.004)	(1.145)	(1.054)	(1.141)
(40, 60)	-2.859	-0.486	-0.835	-0.173	-0.493
	(1.278)	(0.998)	(1.171)	(1.014)	(1.154)
(40, 120)	-2.829	-0.336	-0.642	-0.181	-0.472
	(1.209)	(0.892)	(1.047)	(0.899)	(1.037)
(60, 20)	-3.403	-1.252	-1.740	-0.152	-0.534
	(1.215)	(1.145)	(1.279)	(1.289)	(1.328)
(60, 40)	-3.496	-0.807	-1.238	-0.255	-0.635
	(1.247)	(1.016)	(1.165)	(1.053)	(1.155)
(60, 60)	-3.465	-0.573	-0.987	-0.195	-0.574
	(1.227)	(0.974)	(1.111)	(0.996)	(1.100)
(60, 120)	-3.515	-0.435	-0.819	-0.243	-0.609
	(1.197)	(0.908)	(1.031)	(0.913)	(1.020)
(120, 20)	-4.829	-1.758	-2.450	-0.221	-0.760
	(1.345)	(1.162)	(1.327)	(1.223)	(1.308)
(120, 40)	-4.862	-1.080	-1.679	-0.307	-0.831
	(1.254)	(1.022)	(1.159)	(1.059)	(1.143)
(120, 60)	-4.901	-0.852	-1.419	-0.329	-0.846
	(1.239)	(0.964)	(1.097)	(0.978)	(1.077)
(120, 120)	-5.016	-0.622	-1.203	-0.352	-0.908
	(1.248)	(0.922)	(1.059)	(0.927)	(1.048)

(a) $\mu_\lambda = 0.1$, $\sigma_{31} = -0.8$, $\sigma_{21} = -0.4$, $\theta_{31} = 0.8$, and $\theta_{21} = 0.4$.