Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Econometrics 149 (2009) 82–99

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Panel cointegration with global stochastic trends[☆]

Jushan Bai[a,b,*], Chihwa Kao[c], Serena Ng[d]

[a] *Department of Economics, New York University, New York, NY 10003, USA*
[b] *School of Economics and Management, Tsinghua University, Beijing, China*
[c] *Center for Policy Research and Department of Economics, Syracuse University, Syracuse, NY 13244-1020, USA*
[d] *Department of Economics, Columbia University, 440 W. 118 Street, New York, NY 10027, USA*

ARTICLE INFO

Article history:
Received 21 December 2006
Received in revised form
2 July 2008
Accepted 20 October 2008
Available online 24 November 2008

JEL classification:
C13
C33

Keywords:
Panel data
Common shocks
Co-movements
Cross-sectional dependence
Factor analysis
Fully-modified estimator

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 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 (mixed) 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.

© 2008 Elsevier B.V. All rights reserved.

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 a slow 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 mixed 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 observed regressors are stationary.

Panel data have long been used to study and test economic hypotheses. Panel data bring in information from two dimensions 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

☆ 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).

* Corresponding author at: Department of Economics, New York University, New York, NY 10003, USA. Tel.: +1 202 998 8960.
E-mail addresses: Jushan.Bai@nyu.edu (J. Bai), cdkao@maxwell.syr.edu (C. Kao), serena.Ng@columbia.edu (S. Ng).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.10.012

[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.

Author's personal copy

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 (1987) means that there exists a cointegrating vector, $(1\ -\beta')$, such that the linear combinations $y_t - x_t'\beta$ are stationary, or are 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 already exists[2] for modelling panel cointegration when e_{it} is cross-sectionally independent. A serious drawback of panel cointegration models with cross-section independence is that there is no role for common shocks, which, in theory should be the underlying source of co-movement in the cross-section units. 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. A regression of y_{it} on x_{it} will give a consistent estimator for β when F_t is I(0). We focus on estimation and inference about β when F_t is non-stationary. Empirical studies suggest the relevance of such a setup. Holly et al. (2006) analyzed the relationship between real housing prices and real income at the state level, allowing for unobservable common factors. They found evidence of cointegration after controlling for common factors and additional spatial correlations. Some economic models lead naturally to this set up. Consider a panel of industry data on output and factor inputs such as capital, and labor. Neoclassical production function suggests that log output y_{it} is linear in log factor inputs x_{it} and log productivity e_{it}. Decomposing the latent e_{it} into the industry wide component F_t and an industry specific component u_{it} and assuming that F_t is the source of non-stationarity leads to a model with latent common trends. In such a case, 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 vector, β. The estimators enable the 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. Section 5 presents Monte Carlo results to illustrate the finite sample properties of the proposed estimators. Section 6 provides a brief conclusion. Appendix A contains the technical materials. A detailed technical appendix is given in Bai et al. (2006).

2. The model

Consider the model

$$y_{it} = x_{it}'\beta + e_{it},$$

where for $i = 1, \ldots, n$, $t = 1, \ldots, T$, 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

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

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.

The cross-section independence assumption is restrictive and difficult to justify when the data under investigation are economic time series. In view of co-movements 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 an $r \times 1$ vector of latent common factors, λ_i is an $r \times 1$ vector of factor loadings and u_{it} is the idiosyncratic error. If F_t and u_{it} are both stationary, then e_{it} is also stationary. In this

---

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 (2008), Gengenbach et al. (2005a), Westerlund and Edgerton (2005), Bai and Ng (2004), and Pesaran and Yamagata (2006). Breitung and Pesaran (2005) provide additional references in their survey.

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 mixed normal.

Author's personal copy

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

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

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 the 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). More precisely, if

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

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

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

with 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

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

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. As F is not observable, consistent estimation of the parameter of interest β thus involve a new methodology.

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 $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$, and $BM(\Omega)$ to denote Brownian motion with the covariance matrix Ω. We use $\|A\|$ to denote $(\text{tr}(A'A))^{1/2}$, $\xrightarrow{d}$ 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. We also define the matrix that projects onto the orthogonal space of z as $M_z = I_T - z(z'z)^{-1}z'$. We will use 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. Denote $(n, T) \to \infty$ as the joint limit. Denote $(n, T)_{seq} \to \infty$ as the sequential limit, i.e., $T \to \infty$ first and $n \to \infty$ later. We use $MN(0, V)$ to denote a mixed normal distribution with variance V.

Our analysis is based on the following assumptions.

Assumption 1. Factors and Loadings:

(a) $E\|\lambda_i^0\|^4 \le M$. As $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n}\lambda_i^0\lambda_i^{0\prime} \xrightarrow{p} \Sigma_\lambda$, an $r \times r$ diagonal matrix.
(b) $E\|\eta_t\|^{4+\delta} \le M$ for some $\delta > 0$ and for all t. As $T \to \infty$, $\frac{1}{T^2}\sum_{i=1}^{n} F_t^0 F_t^{0\prime} \xrightarrow{d} \int B_\eta B_\eta'$, an $r \times r$ 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 $w_{it} = (u_{it}, \varepsilon_{it}', \eta_t')'$. For each i, $w_{it} = \Pi_i(L)v_{it} = \sum_{j=0}^{\infty}\Pi_{ij}v_{it-j}$ where v_{it} are i.i.d. over t, $\sum_{j=0}^{\infty} j^2\|\Pi_{ij}\| \le M$, and $|\Pi_i(1)| > c > 0$ for all i. In addition, $E(v_{it}) = 0$, $E(v_{it}v_{it}') = I > 0$, and $E\|v_{it}\|^8 \le M < \infty$; v_{it} are independent of λ_j for all i, j, t.

Partition $\Pi_i(L)$ and v_{it} conformably with w_{it},

$$\Pi_i(L) = \begin{bmatrix} \Pi_i^{uu}(L) & \Pi_i^{u\varepsilon}(L) & \Pi_i^{u\eta}(L) \\ \Pi_i^{\varepsilon u}(L) & \Pi_i^{\varepsilon\varepsilon}(L) & \Pi_i^{\varepsilon\eta}(L) \\ \Pi_i^{\eta u}(L) & \Pi_i^{\eta\varepsilon}(L) & \Pi_i^{\eta\eta}(L) \end{bmatrix}, \quad v_{it} = \begin{bmatrix} v_{it}^u \\ v_{it}^\varepsilon \\ v_t^\eta \end{bmatrix}.$$

We assume, throughout, that v_{it}^u, v_{it}^ε, and v_t^η are mutually independent, and $(v_{it}^\eta, v_{it}^\varepsilon)$ are also independent across i. Since η_t does not depend on i, it must be the case that $\Pi_i^{\eta u}(L) = 0$ and $\Pi_i^{\eta\varepsilon}(L) = 0$. Moreover, $\Pi_i^{\eta\eta}(L) = \Pi^{\eta\eta}(L)$ does not depend on i. The entry $\Pi_i^{u\eta}(L)$ links the regression error u_{it} and the common shocks η_t. To prevent the regression error u_{it} to have strong cross-section correlation, one may assume $\Pi_i^{u\eta}(L) = 0$, implying cross-sectional independence for u_{it}. However, cross-sectional independence for u_{it} is not necessary. Write $\Pi_i^{u\eta}(L) = \sum_{j=0}^{\infty}\Pi_{ij}^{u\eta}L^j$. We shall assume that the coefficients $\Pi_{ij}^{u\eta}$ are iid zero-mean random variables across i. This assumption is similar to that of Phillips and Moon (1999), who assume, in our notation, the coefficients of every polynomial in the matrix $\Pi_i(L)$ are iid random variables across i, but they do not consider common shocks. The zero-mean assumption is sufficient for our purpose, and at the same time, still permitting u_{it} to be influenced by the common shocks. For example, consider

$$u_{it} = a_i\eta_t + b_{it},$$

where a_i are iid zero mean and b_{it} are iid zero-mean random variables across i and over t. Then $\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it} = (n^{-1/2}\sum_{i=1}^{n}a_i)(T^{-1/2}\sum_{t=1}^{T}\eta_t) + \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}b_{it} = O_p(1)$. Moreover, $\frac{1}{n}\sum_{i=1}^{n}u_{it} \to 0$ in probability; its limit is not zero if $E(a_i) \neq 0$. These properties are needed in our argument. Note that zero-mean assumption is not made for other entries of $\Pi_i(L)$. In particular, x_{it} can be strongly correlated with the common factors.

Assumption 3. One of the following holds:

(a) $\Pi_i^{u\eta}(L) = 0$ so that u_{it} are cross-sectionally independent.
(b) The coefficients $\Pi_{ij}^{u\eta}$ in $\Pi_i^{u\eta}(L)$ are iid across i for all j, and they are independent of all other random variables of the model; $E(\Pi_{ij}^{u\eta}) = 0$, $E\|\Pi_{ij}^{u\eta}\|^4 \le M$, and $E[\sum_{j=0}^{\infty} j^2\|\Pi_{ij}^{u\eta}\|] \le M$ for all i.

Author's personal copy

Because $\Pi_{ij}^{u\eta}$ are random variables that are independent of all other random variables, they can be treated as constants when convenient by invoking the conditional argument. Note that u_{it} are cross-sectionally independent conditional on $\{\eta_t\}$. The same is true for x_{it}. But unconditionally, they are cross-sectionally dependent under Assumption 3(b).

Remark. When $\Pi_i^{u\eta}(L) \neq 0$ and their coefficients are not zero mean, the problem can be solved by including I(0) common factors. For example, suppose that $\Pi_i^{u\eta}(L) = a_{i1}L + \cdots a_{ip}L^p$. Define $F_t^* = (F_t', v_t^{\eta\prime}, \ldots, v_{t-p}^{\eta\prime})'$ and $\lambda_i^* = (\lambda_i', a_{i1}', \ldots, a_{ip}')'$, then $\lambda_i'F_t + u_{it} = \lambda_i^{*\prime}F_t^* + u_{it}^*$, where $u_{it}^* = \Pi_i^{uu}(L)v_{it}^u + \Pi_i^{u\varepsilon}(L)v_{it}^\varepsilon$, which are cross-sectionally independent, and F_t^* now includes I(0) factors. However, p must be finite in order to obtain good estimate with principal components. This is the approach employed by Bai (2004) in the absence of regressors. Assumption 3 allows $\Pi_i^{u\eta}(L)$ to have infinite number of lags. Throughout our analysis, F_t is assumed to be I(1) for simplicity. Extension to I(0) common factors is discussed in Section 4.

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

Assumptions 2 and 3 imply 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:

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

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

$$\Omega_i = \sum_{j=-\infty}^{\infty} E_c(w_{i0}w_{ij}') = \Pi_i(1)\Pi_i(1)' = \begin{bmatrix} \Omega_{ui} & \Omega_{u\varepsilon i} & \Omega_{u\eta i} \\ \Omega_{\varepsilon ui} & \Omega_{\varepsilon i} & \Omega_{\varepsilon\eta i} \\ \Omega_{\eta ui} & \Omega_{\eta\varepsilon i} & \Omega_\eta \end{bmatrix} \quad (5)$$

partitioned conformably with w_{it}, where E_c is the conditional expectation, conditional on $\{\Pi_{ij}^{u\eta}\}$, see Lemma 3 of Phillips and Moon (1999). Define the one-sided long-run covariance

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

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

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

Then B_i can be rewritten as

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

where $\begin{bmatrix} V_i & W_i' \end{bmatrix}' = BM(I)$ 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 $(\Delta x_{it}', \Delta F_t^{0\prime})'$. Note that $\Omega_{bi} > 0$ since we assume that there is no cointegration relationship in $(x_{it}', F_t^{0\prime})'$ in Assumption 4. Once again, we emphasize that u_{it} and x_{it} are cross-sectionally independent conditional on the common shocks $\{\eta_t\}$.

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 Sections 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

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad x_i = \begin{bmatrix} x_{i1}' \\ x_{i2}' \\ \vdots \\ x_{iT}' \end{bmatrix}, \quad F = \begin{bmatrix} F_1' \\ F_2' \\ \vdots \\ F_T' \end{bmatrix}, \quad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

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

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

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

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

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

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

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

Proposition 1. *Under Assumptions 1–4, as $(n, T)_{seq} \to \infty$*

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

where

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

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

and with C being the σ-field generated by $\{F_t\}$,

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

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

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

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

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

The estimator is $\sqrt{n}T$ consistent if $\phi_{nT}^0 = 0$, which occurs when x_{it} and F_t are 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

---

7 The limiting distribution when F is 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.

Author's personal copy

regression error and the innovations of the I(1) regressors as given by terms such as Ω_{bui} and Δ_{bui}.

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

$$\widehat{\Omega}_i = \sum_{j=T+1}^{T-1} \omega\left(\frac{j}{K}\right)\widehat{\Gamma}_i(j),$$

$$\widehat{\Delta}_i = \sum_{j=0}^{T-1} \omega\left(\frac{j}{K}\right)\widehat{\Gamma}_i(j)$$

$$\widehat{\Gamma}_i(j) = \frac{1}{T}\sum_{t=1}^{T-j} \widehat{w}_{it+j}\widehat{w}'_{it},$$

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, taken from Moon and Perron (2004):

Assumption 5. (a) $\lim\inf_{n,T\to\infty}(\log T/\log n) > 1$.
(b) The kernel function $\omega(\cdot) : \mathrm{R} \to [-1, 1]$ satisfies (i) $\omega(0) = 1, \omega(x) = \omega(-x)$, (ii) $\int_{-1}^{1} \omega(x)^2\,\mathrm{d}x < \infty$ 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 $\frac{1}{2q} < b < \lim\inf \frac{\log T}{\log n} - 1$.

Let

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

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

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

This estimator can alternatively be written as

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

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

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

The bias-corrected estimator can be seen as a panel fully-modified estimator in the spirit of Phillips and Hansen (1990), and is the reason why the estimator is also labeled $\widetilde{\beta}_{\mathrm{LSFM}}$. It is not difficult to verify that $\widetilde{\beta}_{\mathrm{LSBC}}$ and $\widetilde{\beta}_{\mathrm{LSFM}}$ 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 $\widetilde{\beta}_{\mathrm{LSFM}}$ be defined by (10). Under Assumptions 1–5, as $(n, T)_{seq} \to \infty$*

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

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^0 is not observed. This problem is considered in the next two subsections.

3.2. Unobserved F^0 and the Cup estimator

The LSFM considered above is a linear estimator and can be obtained if F^0 is observed. When F^0 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

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

subject to the constraint $T^{-2}F'F = I_r$ and $\Lambda'\Lambda$ is positive definite. 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

$$w_i = y_i - x_i\beta$$
$$= F\lambda_i + u_i.$$

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

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

Since the first term does not depend on F, minimizing (12) with respect to F is equivalent to maximizing $\mathrm{tr}(T^{-2}F'WW'F)$ 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 continuously-updated estimator (Cup) for (β, F) is defined as

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

More precisely, $(\widehat{\beta}_{\mathrm{Cup}}, \widehat{F}_{\mathrm{Cup}})$ 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 nonlinear estimator even though linear least squares estimation is involved at each iteration. An estimate of Λ can be obtained as:

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

Author's personal copy

The triplet $(\widehat{\beta}, \widehat{F}, \widehat{\Lambda})$ jointly minimizes the objective function (11). The estimator $\hat{\beta}_{\text{Cup}}$ is consistent for β. We state this result in the following proposition.

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

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

The next proposition concerns the asymptotic representation of $\hat{\beta}_{\text{Cup}}$.

Proposition 4. *Suppose Assumptions 1–4 hold and $(n, T) \to \infty$ with $\sqrt{n}/T \to 0$. Then*

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

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

The term involving a_{ik} is due to the estimation of F. Thus 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. 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 $\hat{\beta}_{\text{Cup}}$, even allowing for cross-sectional correlation in u_{it}. However, u_{it} are cross-sectionally independent conditional on the σ-field generated by $\{\eta_t\}$ or equivalently generated by $\{v_t^{\eta}\}$ by Assumptions 2 and 3. This conditional independence together with bias-correction is sufficient for the mixture normality.

Theorem 1. *Suppose that Assumptions 1–4 hold. Let $\hat{\beta}_{\text{Cup}}$ be obtained by iteratively updating (13) and (14). As $(n, T)_{seq} \to \infty$, we have*

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

where

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

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

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

$$D_Z = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E\left(\int R_{ni} R_{ni}' | C\right), \quad (15)$$

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

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

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

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

Theorem 1 establishes the large sample properties of the Cup estimator. The Cup estimator is $\sqrt{n}T$ consistent provided that $\phi_{nT} = 0$, which occurs when x_{it} and F_t are exogenous. Since $\phi_{nT} = O_p(1)$, the Cup estimator is at least 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. There is an extra bias term (the term involving a_{ik}) that 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.

We now consider 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 $\hat{\beta}_{\text{Cup}}$, and is denoted by $\hat{\beta}_{\text{CupBC}}$. The second one will be considered in the next subsection, where correction is made during each iteration, and will be denoted by $\hat{\beta}_{\text{CupFM}}$. Let

$$\widehat{\bar{\Omega}}_i = \sum_{j=T+1}^{T-1}\omega\left(\frac{j}{K}\right)\widehat{\Gamma}_i(j),$$

$$\widehat{\bar{\Delta}}_i = \sum_{j=0}^{T-1}\omega\left(\frac{j}{K}\right)\widehat{\Gamma}_i(j)$$

$$\widehat{\Gamma}_i(j) = \frac{1}{T}\sum_{t=1}^{T-j}\widehat{\bar{w}}_{it+j}\widehat{\bar{w}}_{it}',$$

where

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

The bias-corrected Cup estimator is defined as

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

where

$$\widehat{\phi}_{nT} = \left[\frac{1}{nT^2}\sum_{i=1}^{n}\widehat{Z}_i'\widehat{Z}_i\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i\right),$$

$$\hat{\theta}_i = \widehat{Z}_i'\Delta\widehat{\bar{b}}_i\hat{\bar{\Omega}}_{bi}^{-1}\hat{\bar{\Omega}}_{bui} + \left(\widehat{\bar{\Delta}}_{\varepsilon ui}^{+} \quad -\hat{\bar{\delta}}_i' \quad \widehat{\bar{\Delta}}_{\eta u}^{+}\right),$$

$$\hat{\bar{\delta}}_i = (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\widehat{\bar{x}}_i \quad \Delta\widehat{\bar{b}}_i = \left(\Delta\hat{\bar{x}}_i \quad \Delta\widehat{F}\right),$$

$$\widehat{\bar{x}}_i = x_i - \frac{1}{n}\sum_{k=1}^{n} x_{kt}\widehat{a}_{ik}, \quad \widehat{a}_{ik} = \widehat{\lambda}_i'\left(\widehat{\Lambda}'\widehat{\Lambda}/n\right)^{-1}\widehat{\lambda}_k.$$

Theorem 2. *Suppose Assumptions 1–5 hold. Then as $(n, T)_{seq} \to \infty$,*

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

The CupBC is $\sqrt{n}T$ consistent with a limiting distribution that is centered at zero. This type of bias-correction approach is also used in Hahn and Kuersteniner (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,

Author's personal copy

the bias-corrected estimator can be rewritten as a fully-modified estimator. Such a fully-modified estimator is now discussed.

3.3. A fully-modified Cup estimator

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

$$y_{it}^{+} = y_{it} - \widehat{\widehat{\Omega}}_{ubi}\widehat{\widehat{\Omega}}_{bi}^{-1}\begin{pmatrix}\Delta\widehat{\widehat{x}}_{it}\\ \Delta\widehat{F}_{t}\end{pmatrix}$$

$$\hat{\hat{\delta}}_i = (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\widehat{\widehat{x}}_i,$$

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

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

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

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

Theorem 3. *Suppose Assumptions 1–5 hold. Then as $(n, T)_{seq} \to \infty$,*

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

where Σ is given in (15).

The CupFM and CupBC have the same asymptotic distribution, but they are constructed differently. The estimator $\widehat{\beta}_{\text{CupBC}}$ does the bias-correction only once, i.e., at the final stage of the iteration, and $\widehat{\beta}_{\text{CupFM}}$ 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. The equivalence breaks down because of the need to iterate. Again, because of the mixture of normality, hypothesis testing on β can proceed with the usual t or chi square distributions.

Kapetanios et al. (2006) suggest an alternative estimation procedure based on Pesaran (2006). The model is augmented with additional regressors $\bar{y}_t$ and $\bar{x}_t$, which are cross-sectional averages of y_{it} and x_{it}. These averages are used as proxy for F_t. The estimator for the slope parameter β is shown to be $\sqrt{n}$ consistent, but a fully-modified estimator is not considered.

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

Proposition 5. *Suppose that F_t is I(1) and that Assumptions 1, 2 and 4 hold. Let $\hat{F}$ be the solution of (17). Then*

(i) *under Assumption 3 (b), we have $\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - HF_t^0\|^2 = O_p(\frac{1}{n}) + O_p(\frac{1}{T})$,*

(ii) *under Assumption 3 (a), 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. Part (ii) has the same rate as in Bai (2004, Lemma B.1) under the cross-sectional independence of u_{it} but without regressors x_{it}. Similarly, the factor loadings λ_i are estimated with the same rate of convergence as in Bai (2004). Without Assumption 3, Bai (2004) shows that the rate on the right-hand side is only $O_p(1)$.

Thus far, our analysis assumes 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

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

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

$$IC(r) = \log\widehat{\sigma}^2(r) + rg_{nT},$$

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 $a_{nT} = \frac{nT}{n+T}$. Then $P(\hat{r} = r) \to 1$ as $n, T \to \infty$. This criterion estimates the total number of factors, including I(0) factors. To estimate the number of I(1) factors only, the criterion in Bai (2004) can be used. Our theory allows us to group the I(0) common factors as part of the error process u_{it}, as shown in Assumptions 2 and 3. If u_{it} are assumed to be cross-sectionally independent, then I(0) factors must be considered as part of F_t.

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 mixed normality and the construction of test statistics (and their limiting distribution) do not depend on the convergence rate.

4.1. Incidental trends

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

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

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 $\hat{\varepsilon}_{it} = \Delta x_{it} - \overline{\Delta x_i}$, which is detrended residual of x_{it}. But since $\dot{x}_i$ is already

Author's personal copy

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 Theorem 2 and consequently in Theorem 3 is modified upon replacing the random processes $B_{\varepsilon i}$ and B_η by the demeaned and detrended Brownian motions.[8] 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 factors are I(1). There are no stationary regressors or stationary common shocks. The above results should be 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 $\hat{\beta}_{\text{LS}} = (\sum_{i=1}^n x_i' M_{F^0} x_i)^{-1} \sum_{i=1}^n x_i' M_{F^0} y_i$. The term

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

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

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

We now consider this term under 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

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

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 $\left(\Delta_{\eta ui} \quad \Delta_{\eta i}\right)$ in $\left(\Delta_{bui} \quad \Delta_{bi}\right)$ 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), one can assume u_{it} to be independent of common factors so that $\Omega_{u\eta i}$ and $\Delta_{u\eta i}$ are set to zero in the FM construction. For example, if $u_{it} = a_i\eta_t + b_{it}$, then $a_i\eta_t$ should be treated as part of the common factors, leaving b_{it} as the regression errors, which are assumed to be independent of η_t.

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 in assuming that x_{1it} are mean zero. 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 the 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 the 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)

---

8 Alternatively, we can 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, we use $\hat{\eta}_t = \Delta\hat{F}_t - \overline{\Delta\hat{F}}$ in place of $\Delta\hat{F}_t$.

Author's personal copy

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	2s FM	CupBC	CupFM	LSDV	2sFM	CupBC	Cup FM
$\sigma_{21}=0$												
$n, T=20$	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) $c=5$, $\sigma_{32}=0.4$.

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. In practice, 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 $\hat{\beta}_{\text{Cup}}$.

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 $i=1,\ldots,n$, $t=1,\ldots,T$,

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

where[9]

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

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.

9 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.

Author's personal copy

Table 2
Mean bias and standard deviation of estimators for different n and T.

(n, T)	$c = 5$				$c = 10$			
	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	2s FM	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 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 become smaller (except for a small number of cases) as T increases for each

Author's personal copy

Table 4
Mean bias and standard deviation of t-statistics for different n and T.

(n, T)	$c = 5$				$c = 10$			
	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 $\sigma_{32} = 0.4$.

fixed n. As n increases, no improvement in bias 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. In contrast, the proposed Cup estimator is shown to be consistent (at least T-consistent). In the absence of endogeneity, the Cup estimator is also $\sqrt{n}T$ consistent. Because we allow the regressors and the unobservable trends to be endogenous, an asymptotic bias exists for the Cup estimator. We further consider two bias-corrected estimators, CupBC and CupFM, and derive their rate of convergence and their limiting distributions. We show that these estimators are $\sqrt{n}T$ consistent and this holds in spite of endogeneity and 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 A

Throughout we use $(n, T)_{seq} \to \infty$ to denote the sequential limit, i.e., $T \to \infty$ first and followed by $n \to \infty$. We use $MN(0, V)$ to denote a mixed normal distribution with variance V. Let C be the σ-field generated by $\{F_t^0\}$. The first lemma assumes u_i is uncorrelated with (x_i, F^0) for every i. This assumption is relaxed in Lemma A.2.

Lemma A.1. *Suppose that Assumptions 1–4 hold and that u_i is uncorrelated with (x_i, F^0), then as $(n, T)_{seq} \to \infty$*

(a)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}x_i'M_{F^0}x_i \xrightarrow{d} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\int Q_iQ_i'|C\right),$$

(b)

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

Proof. Note that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}x_i'M_{F^0}x_i = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}x_i'M_{F^0}M_{F^0}x_i = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}',$$

where $\widetilde{x}_{it} = x_{it} - \delta_i'F_t^0$ and

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

see, e.g., Phillips and Ouliaris (1990). Thus

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

Author's personal copy

By the continuous mapping theorem

$$\frac{1}{T^2}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}' \xrightarrow{d} \int Q_i Q_i' = \zeta_{1i}$$

as $T \to \infty$. The variable ζ_{1i} is independent across i conditional on C, which is an invariant σ-field. Thus conditioning on C, the law of large numbers for independent random variables gives,

$$\frac{1}{n}\sum_{i=1}^{n}\zeta_{1i} \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E(\zeta_{1i}|C) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\int Q_iQ_i'|C\right).$$

Thus, the sequential limit is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}x_i'M_{F^0}x_i \xrightarrow{d} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\int Q_iQ_i'|C\right).$$

This proves part (a).

Consider (b). Rewrite

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}x_i'M_{F^0}u_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it},$$

where $\widetilde{x}_{it} = x_{it} - \delta_i'F_t^0$ as before. By assumption, u_{it} is I(0) and is uncorrelated with $\widetilde{x}_{it}$. It follows that

$$\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it} \xrightarrow{d} \int Q_i dB_{ui} = \xi_{2i} \sim \Omega_{ui}^{1/2}\left(\int Q_iQ_i'\right)^{1/2}\times Z,$$

where $Z \sim N(0, I_k)$ as $T \to \infty$ for a fixed n. The variable ξ_{2i} is independent across i conditional on C, which is an invariant σ-field. Thus conditioning on C,

$$\frac{1}{n}\sum_{i=1}^{n}\xi_{2i}\xi_{2i}' \xrightarrow{p} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'|C\right)$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ui}\int E\left(Q_iQ_i'|C\right). \tag{22}$$

Let I_i be the σ field generated by $\{F_t^0\}$ and $(\xi_{21},\ldots,\xi_{2i})$. Then $\{\xi_{2i}, I_i; i \geq 1\}$ is a martingale difference sequence (MDS) because $\{\xi_{2i}\}$ are independent across i conditional on C and

$E(\xi_{2i}|I_{i-1}) = E(\xi_{2i}|C) = 0.$

From $\sum_{i=1}^{n}\xi_{2i}\xi_{2i}' = O_p(n)$, the conditional Lindeberg condition in Corollary 3.1 of Hall and Heyde (1980) can be written as

$$\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|I_{i-1}\right) \xrightarrow{p} 0 \tag{23}$$

for all $\delta > 0$. To see (23), notice that

$$\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|I_{i-1}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|C\right).$$

Without loss of generality we assume that ξ_{2i} is a scalar to save notations. By the Cauchy–Schwarz inequality

$$E\left(\xi_{2i}^2 1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|C\right)$$

$$\leq \left\{E\left(\xi_{2i}^4|C\right)\right\}^{1/2}\left\{E\left[1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|C\right]\right\}^{1/2}.$$

Furthermore,

$$E\left[1\left(\|\xi_{2i}\| > \sqrt{n}\delta\right)|C\right] \leq \frac{E\left(\xi_{2i}^2|C\right)}{n\delta^2}.$$

It follows that

$$\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}^2 1\left(\|\xi_i\| > \sqrt{n}\delta\right)|C\right)$$

$$\leq \frac{1}{\sqrt{n}\delta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[E\left(\xi_{2i}^4|C\right)E\left(\xi_{2i}^2|C\right)\right]^{1/2}\right] = O_p(n^{-1/2})$$

in view of

$$\frac{1}{n}\sum_{i=1}^{n}[E\left(\xi_{2i}^4|C\right)E(\xi_{2i}^2|C)]^{1/2} = O_p(1).$$

This proves (23). The central limit theorem for martingale difference sequence, e.g., Corollary 3.1 of Hall and Heyde (1980), implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{2i} \xrightarrow{d} \left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'|C\right)\right]^{1/2}\times Z, \tag{24}$$

where $Z \sim N(0, I)$ and Z is independent of $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'|C\right)$. Note that

$$\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\xi_{2i}\xi_{2i}'|C\right)\right]^{1/2}$$

$$= \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ui}E\left(\int Q_iQ_i'|C\right)\right)^{1/2}.$$

Thus, as $(n, T)_{seq} \to \infty$, we have

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it} \xrightarrow{d} \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ui}E\left(\int Q_iQ_i'|C\right)\right)^{1/2}\times Z$$

which is a mixed normal. The above can be rewritten as

$$\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it} \xrightarrow{d} MN\left(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ui}E\left(\int Q_iQ_i'|C\right)\right).$$

This proves part (b). ■

The convergence in parts (a) and (b) of Lemma A.1 holds jointly. The proofs for Propositions 1 and 2 (with observable F) follow immediately from Lemma A.1. Propositions 3 and 4 are proved in the supplementary appendix of Bai et al. (2006).

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

Lemma A.2. *Suppose Assumptions 1–5 hold. Let* $Z_i = M_{F^0}x_i - \frac{1}{n}\sum_{k=1}^{n}M_{F^0}x_k a_{ik}$. *Then as* $(n, T)_{seq} \to \infty$

(a)

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

(b) *If* u_i *is uncorrelated with* (x_i, F^0) *for all* i, *then*

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

(c) *If* u_i *is possibly correlated with* (x_i, F^0), *then*

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}Z_i'u_i - \sqrt{n}\,\theta^n$$

$$\xrightarrow{d} MN\left(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{\Omega}_{u.bi}E\left(\int R_{ni}R_{ni}'|C\right)\right),$$

Author's personal copy

where

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

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

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

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

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

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

where

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

is an $r\times k$ matrix as $T\to\infty$. Write $\widetilde{x}_i = M_{F^0}x_i = x_i - F^0\delta_i$, a $T\times k$ matrix. Hence

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

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

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

a $k\times 1$ vector, 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 using similar steps in part (a) in Lemma A.1 as $n\to\infty$,

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

Hence

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

as $(n,T)_{seq}\to\infty$, showing (a).

Proof of part (b). Notice that

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

I_b is proved in Lemma A.1, as $(n,T)_{seq}\to\infty$,

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

if $\widetilde{x}_{it}$ and u_{it} are uncorrelated. Similarly, for II_b, we have

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

where $C_{ni} = \frac{1}{n}\sum_{k=1}^{n}a_{ik}\int Q_kQ_k'$ 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

$$\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it} = \frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \delta_i'F_t^0\right)u_{it} = \frac{1}{T}\sum_{t=1}^{T}\left(I_k \quad -\delta_i'\right)\begin{pmatrix}x_{it}\\ F_t^0\end{pmatrix}u_{it}$$
$$= \left(I_k \quad -\delta_i'\right)\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}x_{it}\\ F_t^0\end{pmatrix}u_{it}$$
$$\xrightarrow{d} \left(I_k \quad -\pi_i'\right)\left[\int\begin{pmatrix}B_{\varepsilon i}dB_{ui}\\ B_\eta dB_{ui}\end{pmatrix} + \begin{pmatrix}\Delta_{\varepsilon ui}\\ \Delta_{\eta u}\end{pmatrix}\right]$$
$$= \int Q_i dB_{ui} + \left(I_k \quad -\pi_i'\right)\begin{pmatrix}\Delta_{\varepsilon ui}\\ \Delta_{\eta u}\end{pmatrix} \tag{25}$$

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

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

such that

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

Note that

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

Therefore

$$\frac{1}{T}\widetilde{x}_i'u_i - \left[\frac{1}{T}x_i'M_{F^0}\left(\Delta x_i \quad \Delta F^0\right)\Omega_{bi}^{-1}\Omega_{bui}\right.$$
$$\left. + \left(I_k \quad -\delta_i'\right)\left[\Delta_{bui} - \Delta_{bi}\Omega_{bi}^{-1}\Omega_{bui}\right]\right]$$

Author's personal copy

$$= \frac{1}{T}\widetilde{x}_i' u_i - \left[\frac{1}{T}x_i' M_{F^0}\left(\Delta x_i \quad \Delta F^0\right)\Omega_{bi}^{-1}\Omega_{bui} + \left(I_k \quad -\delta_i'\right)\Delta_{bui}^{+}\right]$$

$$\xrightarrow{d} \Omega_{u.bi}^{1/2}\int Q_i dV_i \sim \left[\Omega_{u.bi}^{1/2}\int Q_i Q_i'\right]^{1/2} \times N\left(0, I_k\right), \qquad (26)$$

where

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

Let

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

Then we use similar steps in part (b) in Lemma A.1 to get

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\widetilde{x}_i' u_i - \sqrt{n}\theta_1^n = \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{it}u_{it} - \sqrt{n}\theta_1^n$$

$$\xrightarrow{d} MN\left(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{u.bi}E\left(\int Q_i Q_i'|C\right)\right)$$

as $(n, T)_{seq} \to \infty$.

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

$$Z_i = \widetilde{x}_i - \frac{1}{n}\sum_{k=1}^{n}\widetilde{x}_k a_{ik}$$

$$= \left(x_i - F^0\delta_i\right) - \frac{1}{n}\sum_{k=1}^{n}\left(x_k - F^0\delta_k\right)a_{ik}$$

$$= \left(x_i - \frac{1}{n}\sum_{k=1}^{n}x_k a_{ik}\right) - F^0\left(\delta_i - \frac{1}{n}\sum_{k=1}^{n}\delta_k a_{ik}\right)'$$

$$= \bar{x}_i - F^0\bar{\delta}_i',$$

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 (25) as

$$\frac{1}{T}\sum_{t=1}^{T}Z_{it}u_{it} = \frac{1}{T}\sum_{t=1}^{T}\left(\bar{x}_{it} - \bar{\delta}_i' F_t^0\right)u_{it}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(I_k \quad -\bar{\delta}_i'\right)\begin{pmatrix}\bar{x}_{it}\\ F_t^0\end{pmatrix}u_{it}$$

$$= \left(I_k \quad -\bar{\delta}_i'\right)\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\bar{x}_{it}\\ F_t^0\end{pmatrix}u_{it}$$

$$\xrightarrow{d} \left(I_k \quad -\bar{\pi}_i'\right)\left[\int\begin{pmatrix}\bar{B}_{\varepsilon i}dB_{ui}\\ B_\eta dB_{ui}\end{pmatrix} + \begin{pmatrix}\bar{\Delta}_{\varepsilon ui}\\ \bar{\Delta}_{\eta u}\end{pmatrix}\right]$$

$$= \int R_{ni}dB_{ui} + \left(I_k \quad -\bar{\pi}_i'\right)\begin{pmatrix}\bar{\Delta}_{\varepsilon ui}\\ \bar{\Delta}_{\eta u}\end{pmatrix}, \qquad (27)$$

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 appear in the last line in (27) because

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

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

$$- \frac{1}{n}\sum_{k=1}^{n}\left\{B_{\varepsilon k} - \left[\left(\int B_\eta B_\eta'\right)^{-1}\int B_\eta B_{\varepsilon k}'\right]B_\eta\right\}a_{ik}$$

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

Let

$$\theta^n = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}Z_i'\left(\Delta\bar{x}_i \quad \Delta F^0\right)\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui} + \left(I_k \quad -\bar{\delta}_i'\right)\bar{\Delta}_{bui}^{+}\right].$$

Clearly

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}Z_i' u_i - \sqrt{n}\,\theta^n = \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\left(\widetilde{x}_i - \frac{1}{n}\sum_{k=1}^{n}\widetilde{x}_k a_{ik}\right)' u_i$$

$$\xrightarrow{d} MN\left(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{\Omega}_{u.bi}E\left(\int R_{ni}R_{ni}'|C\right)\right)$$

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

$$\sqrt{n}T\left(\widehat{\beta}_{\text{Cup}} - \beta\right) - \sqrt{n}\phi_{nT}$$

$$\xrightarrow{d} MN\left(0, D_Z^{-1}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{\Omega}_{u.bi}E\left(\int R_{ni}R_{ni}'|C\right)\right]D_Z^{-1}\right),$$

where

$$D_Z = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(\int R_{ni}R_{ni}'|C\right)$$

and

$$\phi_{nT} = \left[\frac{1}{nT^2}\sum_{i=1}^{n}Z_i' Z_i\right]^{-1}\theta^n. \quad ■$$

Proof of Theorem 2 and 3. The proof for Theorem 2 is similar to that of Theorem 3, thus omitted. To prove Theorem 3, we need some preliminary results. First we examine the limiting distribution of the infeasible FM estimator, $\widetilde{\beta}_{\text{CupFM}}$. 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 $\left(\Delta\bar{x}_{it}' \quad \Delta F_t^{0\prime}\right)'$ and hence the endogeneity can be removed. The serial correlation correction term has the form

$$\bar{\Delta}_{bui}^{+} = \begin{pmatrix}\bar{\Delta}_{\varepsilon ui}^{+}\\ \bar{\Delta}_{\eta u}^{+}\end{pmatrix} = \left(\bar{\Delta}_{bui} \quad \bar{\Delta}_{bi}\right)\begin{pmatrix}I_k\\ -\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui}\end{pmatrix}$$

$$= \bar{\Delta}_{bui} - \bar{\Delta}_{bi}\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui},$$

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

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

Author's personal copy

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 $\widetilde{\beta}_{\text{CupFM}}$.

Lemma A.3. *Suppose assumptions in Theorem 1 hold. Then as* $(n, T)_{\text{seq}} \to \infty$

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

Proof. Let $w_{it}^+ = \left(u_{it}^+ \quad \varepsilon_{it}' \quad \eta'\right)'$ and we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} w_{it}^+ \xrightarrow{d} \begin{bmatrix} B_{ui}^+ \\ B_{\varepsilon i} \\ B_\eta \end{bmatrix} = \begin{bmatrix} B_{ui}^+ \\ B_{bi} \end{bmatrix} = BM\left(\Omega_i^+\right) \quad \text{as } T \to \infty, \tag{28}$$

where

$$B_{bi} = \begin{bmatrix} B_{\varepsilon i} \\ B_\eta \end{bmatrix}, \quad \Omega_{u.bi} = \Omega_{ui} - \Omega_{ubi}\Omega_{bi}^{-1}\Omega_{bui},$$

$$\Omega_i^+ = \begin{bmatrix} \Omega_{u.bi} & 0 \\ 0 & \Omega_{bi} \end{bmatrix} = \begin{bmatrix} \Omega_{u.bi} & 0 & 0 \\ 0 & \Omega_{\varepsilon i} & \Omega_{\varepsilon\eta i} \\ 0 & \Omega_{\eta\varepsilon i} & \Omega_\eta \end{bmatrix}$$

$$= \Sigma^+ + \Gamma^+ + \Gamma^{+\prime},$$

$$\begin{bmatrix} B_{ui}^+ \\ B_{bi} \end{bmatrix} = \begin{bmatrix} I & -\Omega_{ubi}\Omega_{bi}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} B_{ui} \\ B_{bi} \end{bmatrix}.$$

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 (26) in Lemma A.2 that

$$\zeta_{1iT}^+ = \frac{1}{T}\sum_{t=1}^T \widetilde{x}_{it}u_{1it}^+ = \left(I_k \quad -\delta_i'\right)\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{1it}^+$$

$$= \left(I_k \quad -\delta_i'\right)\left[\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix} u_{it} - \frac{1}{T}\sum_{t=1}^T \begin{pmatrix} x_{it} \\ F_t^0 \end{pmatrix}\Omega_{ubi}\Omega_{bi}^{-1}\begin{pmatrix} \Delta x_{it} \\ \Delta F_t^0 \end{pmatrix}\right]$$

$$\xrightarrow{d} \Omega_{u.bi}^{1/2}\int Q_i dV_i + \left(\Delta_{\varepsilon ui}^+ - \pi_i'\Delta_{\eta u}^+\right) \tag{29}$$

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,

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^n \left(x_i'M_{F^0}u_{1i}^+ - T\left(\Delta_{\varepsilon ui}^+ - \delta_i'\Delta_{\eta u}^+\right)\right)$$

$$= \frac{1}{\sqrt{n}T}\sum_{i=1}^n\left(\sum_{t=1}^T \widetilde{x}_{it}u_{1it}^+ - T\left(\Delta_{\varepsilon ui}^+ - \delta_i'\Delta_{\eta u}^+\right)\right)$$

$$\xrightarrow{d} MN\left(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Omega_{u.bi}E\left(\int Q_iQ_i'|C\right)\right)$$

as $(n, T)_{\text{seq}} \to \infty$. Next, we modify (29).

$$\frac{1}{T}\sum_{t=1}^T Z_{it}u_{it}^+ = \frac{1}{T}\sum_{t=1}^T\left(\bar{x}_{it} - \bar{\delta}_i'F_t^0\right)u_{it}^+$$

$$= \left(I_k \quad -\bar{\delta}_i'\right)\left[\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix} u_{it}^+ - \frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \bar{x}_{it} \\ F_t^0 \end{pmatrix}\Omega_{ubi}\Omega_{bi}^{-1}\begin{pmatrix} \Delta\bar{x}_{it} \\ \Delta F_t^0 \end{pmatrix}\right]$$

$$\xrightarrow{d} \left(I_k \quad -\bar{\pi}_i'\right)\left\{\int\begin{pmatrix} \bar{B}_{\varepsilon i} \\ B_\eta \end{pmatrix} dB_{ui} + \begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} - \left[\int\begin{pmatrix} \bar{B}_{\varepsilon i} \\ B_\eta \end{pmatrix} dB_{bi}'\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui} + \bar{\Delta}_{bi}\right]\right\}$$

$$= \int R_{ni}dB_{ui} + \left(I_k \quad -\bar{\pi}_i'\right)\begin{pmatrix} \bar{\Delta}_{\varepsilon ui} \\ \bar{\Delta}_{\eta u} \end{pmatrix} - \int\left[R_{ni}dB_{bi}'\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui} + \left(I_k \quad -\bar{\pi}_i'\right)\begin{pmatrix} \bar{\Delta}_{\varepsilon i} \\ \bar{\Delta}_\eta \end{pmatrix}\bar{\Omega}_{bi}^{-1}\bar{\Omega}_{bui}\right]$$

$$= \bar{\Omega}_{u.bi}^{1/2}\int R_{ni}dV_i + \left(\bar{\Delta}_{\varepsilon ui}^+ - \bar{\pi}_i'\bar{\Delta}_{\eta u}^+\right).$$

Therefore,

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

as $(n, T)_{\text{seq}} \to \infty$. Then

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

as $(n, T)_{\text{seq}} \to \infty$. This proves the theorem. ■

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

Lemma A.4. *Under Assumptions 1–5, 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 $\widehat{u}_{it}^+ = u_{it} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1}\begin{pmatrix}\Delta x_{it} \\ \Delta \widehat{F}_t\end{pmatrix}$, $\widehat{\Delta}_{\varepsilon un}^+ = \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_{\varepsilon ui}^+$ *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), the proof is focused on the variances without the bar.

Proof. First, note that

$$\Delta_{bui}^+ = \begin{pmatrix} \Delta_{\varepsilon ui}^+ \\ \Delta_{\eta u}^+ \end{pmatrix} = \left(\Delta_{bui} \quad \Delta_{bi}\right)\begin{pmatrix} 1 \\ -\Omega_{bi}^{-1}\Omega_{bui} \end{pmatrix}$$

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

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 arguments as in the proofs of Theorems 9 and 10 of Hannan (1970) (also see similar result of Moon and Perron (2004)), we have

$$E\left\|\sqrt{n}\left(\widehat{\Delta}_{\varepsilon un}^+ - \Delta_{\varepsilon un}^+\right)\right\|^2 \le \sup_i E\left\|\widehat{\Delta}_{\varepsilon ui}^+ - E\widehat{\Delta}_{\varepsilon ui}^+\right\|^2 + n\sup_i\left\|E\widehat{\Delta}_{\varepsilon ui}^+ - \Delta_{\varepsilon ui}^+\right\|^2$$

$$= O\left(\frac{K}{T}\right) + O\left(\frac{n}{K^{2q}}\right).$$

Author's personal copy

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 5, $K \backsim n^b$. Then

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

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

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

if $b < \liminf \frac{\log T}{\log n}$ by Assumption 5. Then

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

as required. This proves (a).

To establish (b), we note

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

as required for part (b).

Let $\widetilde{u}^{+}_{it}=u_{it}-\widehat{\Omega}_{ubi}\widehat{\Omega}^{-1}_{bi}\begin{pmatrix}\Delta x_{it}\\ \Delta F^0_t\end{pmatrix}$. Next,

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}\widehat{u}^{+}_i-x_i'M_{F^0}u^{+}_i\right)$$
$$=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}\widehat{u}^{+}_i-x_i'M_{\widehat{F}}\widetilde{u}^{+}_i\right.$$
$$\left.+x_i'M_{\widehat{F}}\widetilde{u}^{+}_i-x_i'M_{\widehat{F}}u^{+}_i+x_i'M_{\widehat{F}}u^{+}_i-x_i'M_{F^0}u^{+}_i\right)$$
$$=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}\widetilde{u}^{+}_i-x_i'M_{\widehat{F}}u^{+}_i\right)$$
$$+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}u^{+}_i-x_i'M_{F^0}u^{+}_i\right)$$
$$+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}\widehat{u}^{+}_i-x_i'M_{\widehat{F}}\widetilde{u}^{+}_i\right)$$
$$=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'M_{\widehat{F}}\left(\widetilde{u}^{+}_i-u^{+}_i\right)+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(x_i'M_{\widehat{F}}-x_i'M_{F^0}\right)u^{+}_i$$
$$+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'M_{\widehat{F}}\left(\widehat{u}^{+}_i-\widetilde{u}^{+}_i\right)$$
$$=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'M_{\widehat{F}}\left(\widetilde{u}^{+}_i-u^{+}_i\right)+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'\left(M_{\widehat{F}}-M_{F^0}\right)u^{+}_i$$
$$+\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'M_{\widehat{F}}\left(\widehat{u}^{+}_i-\widetilde{u}^{+}_i\right)$$
$$=I+II+III.$$

From the proof of Proposition 4 in the supplementary appendix,

$$II=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}x_i'\left(M_{\widehat{F}}-M_{F^0}\right)u^{+}_i=o_p(1)$$

if we replace u_i by u^{+}_i. Let $\Delta b_i=\begin{pmatrix}\Delta x_i & \Delta F^0\end{pmatrix}$ be a $T\times(k+r)$ matrix. Consider I.

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

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

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

Then we have

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

and

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

For I_c, by the Cauchy–Schwarz inequality,

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

Author's personal copy

Similarly,

$$\|II_{c.}\| = \left\| \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} x_i' \frac{\widehat{F}\widehat{F}'}{T^2} \left(\Delta b_i \left(\Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1}\right)\right) \right\|$$

$$= \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i'\widehat{F}}{T^2} \frac{\widehat{F}'\Delta b_i}{T} \left(\Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1}\right) \right\|$$

$$\leq \left\| \left(\sqrt{n}\frac{1}{n} \sum_{i=1}^{n} \left\| \frac{x_i'\widehat{F}}{T^2} \frac{\widehat{F}'\Delta b_i}{T} \right\|^2 \right)^{1/2} \times \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\| \Omega_{ubi}\Omega_{bi}^{-1} - \widehat{\Omega}_{ubi}\widehat{\Omega}_{bi}^{-1} \right\|^2 \right)^{1/2} \right\|$$

$$= O_p\left(\sqrt{n}\right) O_p\left(\text{Max}\left\{\sqrt{\frac{K}{T}}, \sqrt{\frac{1}{K^{2q}}}\right\}\right).$$

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

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

$$= O_p\left(\text{Max}\left\{\sqrt{\frac{nK}{T}}, \sqrt{\frac{n}{K^{2q}}}\right\}\right).$$

Recall $K \backsim n^b$ and $\liminf \frac{\log T}{\log n} > 1$ from Assumption 5. It follows that, as in Moon and Perron (2004)

$$\frac{nK}{T} \backsim \frac{n^{b+1}}{T} = \exp\left(\log\left(\frac{n^{b+1}}{T}\right)\right) = \exp\left(b + 1 - \frac{\log T}{\log n}\right)\log n$$

$$= n^{b+1-\frac{\log T}{\log n}} \leq n^{b+1-\liminf \frac{\log T}{\log n}} \to 0$$

by Assumption 5 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 5 and $\frac{1}{2q} < b$. Therefore

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

Let

$$\Delta\widehat{b}_i = \left(\Delta x_i \quad \Delta\widehat{F}\right).$$

Note that

$$\Delta b_i - \Delta\widehat{b}_i = \left(\Delta x_i \quad \Delta F^0\right) - \left(\Delta x_i \quad \Delta\widehat{F}\right) = \left(0 \quad \Delta F^0 - \Delta\widehat{F}\right).$$

Consider III.

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

$$= \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} x_i' M_{\widehat{F}} \left(u_i - \Delta\widehat{b}_i\widehat{\Omega}_{bi}^{-1}\widehat{\Omega}_{bui} - u_i + \Delta b_i\widehat{\Omega}_{bi}^{-1}\widehat{\Omega}_{bui}\right)$$

$$= \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} x_i' M_{\widehat{F}} \left(\Delta b_i - \Delta\widehat{b}_i\right) \widehat{\Omega}_{bi}^{-1}\widehat{\Omega}_{bui}$$

$$= \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} x_i' M_{\widehat{F}} \left(\Delta F^0 - \Delta\widehat{F}\right) \widehat{\Omega}_{bi}^{-1}\widehat{\Omega}_{bui}.$$

We use Lemma 12.3 in Bai (2005) to get

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

It follows that

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

$$= \sqrt{n}\left[O_p\left(\widehat{\beta} - \beta^0\right) + O_p\left(\frac{1}{\min(n, T)}\right)\right]$$

$$= \sqrt{n}O_p\left(\frac{1}{T}\right) + O_p\left(\frac{\sqrt{n}}{\min(n, T)}\right) = o_p(1)$$

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

Proposition A.1. *Under Assumptions 1–5,*

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

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)

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

Then

$$\sqrt{n}T\left(\widehat{\beta}_{\text{CupFM}} - \widetilde{\beta}_{\text{CupFM}}\right)$$

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

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

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

$$+ o_p(1),$$

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.4,

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

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

Proof of Proposition 5. Consider (i). In the supplementary appendix, it is shown that

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

Author's personal copy

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 $O(1/n)+O_p(1/T)$. The proof of (ii) is similar. ■

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jeconom.2008.10.012.

References

Andrews, D.W.K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551–1585.

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

Bai, J., 2005. Panel data models with interactive fixed effects. Working paper.

Bai, J., Kao, C., 2006. On the estimation inference of a panel cointegration model with cross-sectional dependence. In: Baltagi, Badi (Ed.), Contributions to Economic Analysis. Elsevier, pp. 3–30.

Bai, J., Kao, C., Ng, S., 2006. Panel cointegration with global stochastic trends: Supplementary appendix.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2004. A panic attack on unit roots cointegration. Econometrica 72, 1127–1177.

Baltagi, B., 2005. Econometric Analysis of Panel Data. Wiley, New York.

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

Bernanke, B.S., Boivin, J., 2003. Monetary policy in a data-rich environment. Journal of Monetary Economics 50, 525–546.

Breitung, J., Das, S., 2008. Testing for unit roots in panels with a factor structure. Econometric Theory 24, 88–108.

Breitung, J., Pesaran, M.H., 2005. Unit roots and cointegration in panels. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, third ed. Kluwer Academic Publishers, pp. 236–265.

Chang, Y., 2002. Nonlinear IV unit root tests in panels with cross-sectional dependence. Journal of Econometrics 110, 261–292.

Chang, Y., 2004. Bootstrap unit root tests in panels with cross-sectional dependence. Journal of Econometrics 120, 263–293.

Choi, I., 2006. Combination unit root tests for cross-sectionally correlated panels. In: Econometric Theory Practice: Frontiers of Analysis Applied Research: Essays in Honor of Peter C.B. Phillips. Cambridge University Press, Cambridge.

Engle, R.F., Granger, C.W.J., 1987. Cointegration and error-correction: Representation, estimation, and testing. Econometrica 55, 251–276.

Gengenbach, C., Palm, F.C., Urbain, J.-P., 2005a. Panel unit root tests in the presence of cross-sectional dependencies: Comparison implications of modelling. Working paper.

Gengenbach, C., Palm, F.C., Urbain, J.-P., 2005b. Panel cointegration testing in the presence of common factors. Working paper.

Hahn, J., Kuersteniner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects. Econometrica 70, 1639–1657.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Applications. Academic Press, Inc., New York.

Hannan, E., 1970. Multiple Time Series. Wiley, New York.

Hansen, B.E., 1992. Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. Journal of Econometrics 53, 87–121.

Holly, S., Pesaran, M.H., Yamagata, T., 2006. A spatio-temporal model of house prices in the US. IZA Discussion Paper No. 2338, University of Cambridge.

Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, Cambridge.

Kao, C., 1999. Spurious regression residual-based tests for cointegration in panel data. Journal of Econometrics 90, 1–44.

Kapetanios, G., Pesaran, H.H., Yamagata, T., 2006. Panels with nonstationary multifactor error structures. Discussion paper No. 2243. Quenn Mary, University of London.

Moon, H.R., Perron, B., 2004. Testing for a unit root in panels with dynamic factor. Journal of Econometrics 122, 81–126.

Moon, H.R., Phillips, P.C.B., 2000. Estimation of autoregressive roots of near unit using panel data. Econometric Theory 927–997.

Moon, H.R., Phillips, P.C.B., 2004. GMM estimation of autoregressive roots near unit with panel data. Econometric Theory 927–997.

Park, J., Phillips, P.C.B., 1988. Regressions with integrated regressors: Part I. Econometric Theory 468–498.

Pesaran, M.H., 2006. Estimation inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

Pesaran, M.H., Smith, R., 1995. Estimation of long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79–114.

Pesaran, M.H., Yamagata, T., 2006. Testing slope homogeneity in large panels. Working paper.

Phillips, P.C.B., 1995. Fully-modified least squares and autoregression. Econometrica 63, 1023–1079.

Phillips, P.C.B., Durlauf, S., 1986. Multiple time series regression with integrated processes. Review of Economic Studies 53, 473–495.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I 1) processes. Review of Economic Studies 57, 99–125.

Phillips, P.C.B., Moon, H., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057–1111.

Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. Econometrica 58, 165–193.

Phillips, P.C.B., Sul, D., 2003. Dynamic panel estimation homogeneity testing under cross section dependence. Econometrics Journal 6, 217–259.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Westerlund, J., 2006. On the estimation of cointegrated panel regressions with common factors. Working paper.

Westerlund, J., Edgerton, D., 2005. Testing the Barro–Gordon model in breaking dependent panels: Evidence from the OECD countries. Working paper.