Maastricht University

Joakim Westerlund,
Jean-Pierre Urbain

**Cross sectional averages or
principal components?**

RM/11/053

METEOR

CROSS-SECTIONAL AVERAGES OR PRINCIPAL COMPONENTS?[*]

Joakim Westerlund[†]
University of Gothenburg
Sweden

Jean-Pierre Urbain
Maastricht University
The Netherlands

November 11, 2011

Abstract

In spite of the increased use of factor-augmented regressions in recent years, little is known regarding the relative merits of the two main approaches to estimation and inference, namely, the cross-sectional average and principal components estimators. As a response to this, the current paper offers an in-dept theoretical analysis of the issue.

JEL Classification: C12; C13; C33.

Keywords: Factor-augmented panel regressions; common factor models; principal components; cross-sectional averages; cross-sectional dependence.

1 Introduction

Recently, there has been increased interest in analysis of panel data models in which the standard assumption that the regression errors are cross-sectionally uncorrelated is violated. When the regression errors are cross-sectionally correlated standard estimation methods do not necessarily produce consistent estimates of the coefficients of interest, and much effort has therefore gone into the development of robust methods. In particular, the use of factor-augmented regressions has recently become very popular.

The key assumption in factor-augmented regressions is that the cross-section dependence can be represented by means of a small number of common factors, which can then be included as additional regressors. For factors that are observed, such as interest rates or oil prices, this is of course very easy. However, many factors are unobserved and lack good proxies. The most common approach to deal with the presence of such latent factors is to use estimates in their stead. In this paper we focus on two estimators of the factor-augmented regression; (1) the principal components (PC) estimator considered by for example Bai (2009) and Greenaway-McGrevy *et al.* (2010), and (2) the cross-sectional average (CA) estimator of Pesaran (2006). The main reason for this is that in spite of their popularity, the relative merits of these estimators are not well understood. In fact, most practitioners seem to use them quite interchangeably, as though their properties were the same. In Pesaran (2006) the CA estimator is termed the common correlated effects (CCE) estimator. However, since Pesaran (2006) only proposes CCE as an estimator of the coefficient of the factor-augmented regression and not of the factors themselves, in order to keep the distinction, in this paper we use CA to refer to the estimator based on cross-sectional averages.

Stock and Watson (2002), and Bai (2003) study the PC estimator of the factors in the context of a conventional factor model. They show that the PC estimates are consistent for the space spanned by the true factors instead of the factors themselves. Fortunately, in factor-augmented regressions consistent estimation of the factors is not necessary, as the factor estimates are there merely to control for the cross-section dependence. This has been shown in recent works of Bai (2009) and Greenaway-McGrevy *et al.* (2010), who study the theoretical properties of the estimated PC factor-augmented regression. According to their results, in spite of the generated regressor problem caused by the use of the estimated factors, normal inference is usually possible with the estimated coefficients converging to their true values at the rate $\sqrt{NT}$, where T and N denote the number of time series and cross-sectional observations, respectively.

Pesaran (2006) is not interested in the estimation of the factors *per se* and only considers the estimation of the factor-augmented regression. The CA approach consists of approximating the common component of the data by the cross-sectional averages of the dependent and explanatory variables, and then augmenting the panel regression with these averages. As Pesaran (2006) shows, the CA estimator of the regression coefficients is asymptotically normal with the rate of consistency again given by $\sqrt{NT}$.

However, while the inferential theory is there, little is known regarding the relative merits of the PC and CA estimators. Indeed, as Pesaran (2006, page 1001) concludes:

> It is also of interest to compare the approach proposed in this paper with the alternative procedure that proxies the unobserved common factors by principal components (PC) of y_{it} and x_{it}. This alternative is considered in a series of Monte Carlo experiments in Kapetanios and Pesaran (2006) and Coakley, Fuertes, and Smith (2006). Kapetanios and Pesaran's experiments allow for up to four regressors and factors, and find that the PC procedure does not perform as well as the CCE approach and leads to substantial size distortions even if, when using the PC procedure, the true number of unobserved factors is assumed to be known.

As the quotation suggests the knowledge regarding the relative merits of the two estimators is limited, at best, and where evidence exists is it based exclusively on Monte Carlo simulation, which need not be informative of any theoretical differences. In Kapetanios and Pesaran (2006, page 13) it is concluded that: "Overall, it appears that even if one knows the factors, the small sample bias in the model selection aspect of the PC augmented procedure is important enough to adversely affect the performance of the estimators for moderate values of T, even if one abstracts from the small sample bias in estimation of the unobserved factors." Thus, while the results seem to suggest that the CA estimator performs best, the reason for this is largely unknown.

In this paper we offer a theoretical explanation for the previously obtained simulation results. The paper is organized as follows. Section 2 introduces the model of interest, and discusses how it relates to the models considered by Bai (2009), Greenaway-McGrevy *et al.* (2010) and Pesaran (2006). Section 3 presents the asymptotic results. Since Pesaran (2006) does not consider the estimation of the factors themselves, nothing is known regarding the CA factor estimates. Therefore, since any differences in the regression estimates can only be due to the augmentation, we start by studying the relative properties of the PC and CA factor estimates. Two of the results that emerges are that (i) the PC factor estimator is relatively more efficient, and (ii) the PC and CA estimators of the factor-augmented regression can be biased, even asymptotically, with the size of the bias depending on the variance of the estimated factors. Thus, given the efficiency of the PC factor estimator, one would expect the associated PC regression estimator to perform best. However, this is not what we find. In fact, according to our results the CA estimator generally performs best, a finding that is

verified in small-samples in Section 4. Section 5 concludes. Proofs of important results are provided in Appendix.

A word on notation. The symbols $\rightarrow_d$ and $\rightarrow_p$ will be used to signify convergence in distribution and convergence in probability, respectively. As usual, $y_T = O_p(T^r)$ will be used to signify that y_T is at most order T^r in probability, while $y_T = o_p(T^r)$ will be used in case y_T is of smaller order in probability than T^r.[1] For a $m \times n$ matrix A, $\text{tr}(A)$, $\text{rk}(A)$, A^- and $||A|| = \sqrt{\text{tr}(A'A)}$ will be used to denote its trace, rank, generalized inverse and Euclidean norm, respectively.

2 Model

Consider the scalar and m-dimensional vector of observable panel data variables y_{it} and x_{it}, where $i = 1, ..., N$ and $t = 1, ..., T$ indexes the cross-sectional and time series dimensions, respectively. The data generating process of the vector of stacked observations on y_{it}, $y_i = (y_{i1}, ..., y_{iT})$, is given by

$$y_i = x_i \beta + e_i, \tag{1}$$

$$e_i = F \lambda_i + \epsilon_i, \tag{2}$$

where $x_i = (x'_{i1}, ..., x'_{iT})'$ stacks x_{it}, β is a m-dimensional vector of slope coefficients, $F = (F'_1, ..., F'_T)'$ is a $T \times r$ matrix of common factors with λ_i being the associated r-dimensional vector of factor loadings, which is assumed to be non-random, and $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$ is a T-dimensional vector of idiosyncratic errors that are independent of F and x_i. Although ϵ_i can in principle be cross-section correlated to some extent, in this paper we assume it to be independent with mean zero and positive definite covariance matrix $\sigma^2_{\epsilon i} I_T$.

The above model is the prototypical pooled panel regression with a factor error structure, in which ϵ_i is independent of x_i. If F is also independent of x_i, then (1) is nothing but a static panel data regression with exogenous regressors, which can be estimated consistently using least squares, although efficiency will be gained by using weighted least squares based on the factor error structure. If, however, x_i is correlated with F, then consistency will be lost. To allow for this possibility, we assume that

$$x_i = F \Lambda'_i + \varepsilon_i, \tag{3}$$

---

[1] If y_T is deterministic, then $O_p(T^r)$ and $o_p(T^r)$ are replaced by $O(T^r)$ and $o(T^r)$, respectively.

where Λ_i is a $m \times r$ loading matrix and $\varepsilon_i = (\varepsilon'_{i1}, ..., \varepsilon'_{iT})'$ is a $T \times m$ matrix of idiosyncratic errors that are independent of F. Analogous to ϵ_i, ε_i is assumed to be independent across i, having zero mean and positive definite covariance matrix $I_T \otimes \Sigma_{\varepsilon i}$. The assumptions placed on ϵ_i, ε_i and F may be summarized in the following way.

Assumption 1.

(a) $\epsilon_i \sim \text{iid}(0, \sigma^2_{\epsilon i} I_T)$ and $\varepsilon_i \sim \text{iid}(0, I_T \otimes \Sigma_{\varepsilon i})$ with $\sigma^2_{\epsilon i} < \infty$ and $\Sigma_{\varepsilon i} > 0$;

(b) F_t is covariance stationary such that $E(||F_t||^4) < \infty$ and $E(F_t F'_t) = \Sigma_F > 0$;

(c) λ_i and Λ_i are non-random such that $||\lambda_i|| < \infty$ and $||\Lambda_i|| < \infty$;

(d) ϵ_{it}, ε_{it} and F_t are mutually independent.

Consider the $T \times (m+1)$ matrix $z_i = (y_i, x_i)$. By combining (1)–(3),

$$z_i = FC_i + u_i, \tag{4}$$

where $C_i = (\Lambda'_i \beta + \lambda_i, \Lambda'_i)$ is $r \times (m+1)$ and $u_i = (u'_{i1}, ..., u'_{iT})' = (\varepsilon_i \beta + \epsilon_i, \varepsilon_i)$ is $T \times (m+1)$ with covariance matrix

$$E(u_{it} u'_{it}) = \Sigma_{ui} = \begin{bmatrix} \beta' \Sigma_{\varepsilon i} \beta + \sigma^2_{\epsilon i} & \beta' \Sigma_{\varepsilon i} \\ \Sigma_{\varepsilon i} \beta & \Sigma_{\varepsilon i} \end{bmatrix}.$$

Thus, (1)–(3) can be rewritten equivalently as a static factor model for z_i. At times it will be useful to write this model in matrix notation. Let us therefore introduce the $T \times N(m+1)$ matrix $z = ((z'_{1t}, ..., z'_{Nt})', ..., (z'_{1T}, ..., z'_{NT})')'$ with a similar definition of u, and the $N(m+1) \times r$ matrix $C = (C_1, ..., C_N)'$. In this notation, (4) becomes

$$z = FC' + u.$$

Define $A_i = C_i C'_i$, $B_i = C_i \Sigma_{ui} C'_i$ and $D_i = \sigma^2_{\epsilon i} \Sigma_{\varepsilon i}$, and let $\bar{A}$ denote the cross-sectional average of A_i with a similar definition of $\bar{B}$, $\bar{C}$, $\bar{D}$, $\bar{\Sigma}_\varepsilon$ and $\bar{\Sigma}_u$. In what follows the rank of the limits of these quantities is going to be important, and we therefore make the following assumption.

Assumption 2.

(a) $\bar{A}$, $\bar{B}$, and $\bar{\Sigma}_\varepsilon$ are positive definite for all N, including $N \to \infty$;

(b) $\text{rk}(\bar{C}) = r \leq m + 1$ *for all* N, *including* $N \to \infty$.

Remarks.

1. Many of the conditions stated in Assumption 1 are not strictly necessary, but are retained in order to simplify the comparison between the two estimators. The assumption that ϵ_i and ε_i are iid across i can, for example, be relaxed in the usual manner, by simply replacing all contemporaneous variances with their corresponding long-run variances. Thus, in case of u_{it}, $\Sigma_{ui} = E(u_{it}u'_{it})$ is replaced by $\sum_{s=-\infty}^{\infty} E(u_{it}u'_{it-s})$. In case of the PC estimator we can also relax the assumption that ϵ_i and ε_i are iid across i, but then at the cost of some extra moment conditions, see Bai (2009, Section 4) and Greenaway-McGrevy *et al.* (2010, Assumption A). The assumption that the cross-section dependence of x_i has a common factor structure is similarly not necessary, at least not for the PC estimator, see remark 5 to Theorem 1. Again depending on the estimator used, the requirement that y_i and x_i are stationarity can similarly be relaxed in various ways, see for example Bai *et al.* (2009) in case of the PC estimator and Kapetanios *et al.* (2010) in case of the CA estimator. Finally, while we only consider non-random loadings, the results of Section 3 still hold when λ_i and Λ_i are random, provided that they are independent of ϵ_{it}, ε_{it} and F_t, and satisfies some moment conditions.

2. Assumption 2 (a) ensures that $\bar{A}$, $\bar{B}$ and $\bar{\Sigma}_{\varepsilon}$ are invertible, which in turn implies that $\bar{D}$ and $\bar{\Sigma}_u$ are invertible. To appreciate this, note that $C_i = (\lambda_i, \Lambda'_i)P'$, where

$$P = \begin{bmatrix} 1 & \beta' \\ 0 & I_m \end{bmatrix}$$

has rank $m + 1$. Thus, since $\Sigma_{ui} = P \text{diag}(\sigma_{\epsilon i}^2, \Sigma_{\varepsilon i})P'$ and $\bar{\Sigma}_{\varepsilon}$ is of full rank, $\bar{\Sigma}_u$ must be of full rank too. Finally, $\bar{D}$ and $\bar{\Sigma}_{\varepsilon}$ are full rank by Assumption 1. Assumption 2 (b) ensures that the $r \times r$ matrix $\bar{C}\bar{C}'$ is invertible.

3. The assumption that e_i and x_i depend on the same set of factors is not a restriction. Suppose for example that the factors to e_i and x_i do not have any elements in common. In order to capture this we introduce the $r \times r$ orthogonal matrix $J = (J_e, J_x)$, which is such that $J'J = JJ' = I_r$. The component matrices J_e and J_x, which are of dimension $r \times r_e$ and $r \times (r - r_e)$, respectively, are such that $J'_x J_e = 0$, $J'_e \Lambda_i = 0$ and $J'_x \lambda_i = 0$. The

6

matrix J allows us to rotate F as $FJ = (FJ_e, FJ_x) = (F_e, F_x)$. Thus, defining $J'_e\lambda_i = \lambda_{ei}$ and $J'_x\Lambda_i = \Lambda_{xi}$, we have $e_i = F\lambda_i + \epsilon_i = FJJ'\lambda_i + \epsilon_i = F_e\lambda_{ei} + \epsilon_i$ and similarly $x_i = F_x\Lambda_{xi} + \varepsilon_i$.

3   Asymptotic results

As usual, β, λ_i, F, ϵ_i, Λ_i and ε_i are treated as unknown, and the interest lies with the estimation of β, λ_i, Λ_i and F. Initially, however, we will assume that C_i is known, which allows us to focus on the estimation on F, and then we show how to proceed in the more general case.

3.1 C_i known

The first estimator of F_t that we consider is the infeasible PC (IPC) estimator, which in the current setting is given by

$$\hat{F}_t^{IPC} = \left(\sum_{i=1}^N C_iC_i'\right)^{-1}\sum_{i=1}^N C_iz_{it}$$

In order to obtain the limiting distribution of this estimator, and also that of the infeasible CA (ICA) estimator, we make use of Assumptions 1 and 2, which ensure that a law of large numbers and a central limit theorem for heterogeneous processes apply. It follows that

$$\sqrt{N}(\hat{F}_t^{IPC} - F_t) = \left(\frac{1}{N}\sum_{i=1}^N C_iC_i'\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N C_iu_{it} \to_d N(0, \Omega_{IPC}) \tag{5}$$

as $N \to \infty$, where

$$\Omega_{IPC} = \lim_{N\to\infty}\bar{\Omega}_{IPC} = \lim_{N\to\infty}\bar{A}^{-1}\bar{B}\bar{A}^{-1}.$$

The second estimator the we consider, the ICA estimator, is defined as

$$\hat{F}_t^{ICA} = (\bar{C}\bar{C}')^{-1}\bar{C}\bar{z}_t,$$

where $\bar{z}_t$ is the cross-sectional average of z_{it}. The asymptotic distribution of $\sqrt{N}(\hat{F}_t^{ICA} - F_t)$ is also normal, but the covariance matrix is now given by

$$\Omega_{ICA} = \lim_{N\to\infty}\bar{\Omega}_{ICA} = \lim_{N\to\infty}(\bar{C}\bar{C}')^{-1}\bar{C}\bar{\Sigma}_u\bar{C}'(\bar{C}\bar{C}')^{-1}.$$

The purpose of the current paper is to analyze the relative efficiency of the PC and CA estimators, and also that of the associated factor augmented estimators of β. Foreshadowing

this analysis, it is useful to view $\hat{F}_t^{IPC}$ and $\hat{F}_t^{ICA}$ as instrumental variables estimators, having the general form $(\sum_{i=1}^N Z_i C_i')^{-1} \sum_{i=1}^N Z_i z_{it}$, with Z_i being the instrument. The choice $Z_i = C_i$ leads to the PC estimator, while $Z_i = \bar{C}$ leads to the CA estimator. Intuitively, since C_i is independent of u_{it}, the PC estimator should be more efficient. The next proposition shows that this is indeed the case.

Proposition 1. *Under Assumptions 1 and 2, $\bar{\Omega}_{ICA} - \bar{\Omega}_{IPC} \geq 0$.*

Thus, the PC estimator is generally superior in terms of efficiency. The only exception is when $\Sigma_{\varepsilon i}$, Λ_i and λ_i are equal across i, in which case the two estimators are equally efficient. Note also that Proposition 1 implies $\Omega_{ICA} - \Omega_{IPC} \geq 0$.

The next step in our analysis is to make use of Proposition 1 when analyzing the efficiency of estimators of β. In so doing, it is convenient to define the factor projection matrix

$$M_F = I_T - F(F'F)^{-1}F',$$

suggesting that the infeasible LS (ILS) estimator of β can be written as

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

The asymptotic distribution of this estimator can be readily deduced by noting that $M_F x_i = M_F \varepsilon_i$, suggesting

$$\sqrt{NT}(\hat{\beta}_{ILS} - \beta) = \left(\frac{1}{NT} \sum_{i=1}^N \varepsilon_i' M_F \varepsilon_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \varepsilon_i' M_F \varepsilon_i \rightarrow_d N(0, \Omega_{ILS}) \qquad (6)$$

as $N, T \to \infty$, where

$$\Omega_{ILS} = \lim_{N \to \infty} \bar{\Sigma}_\varepsilon^{-1} \bar{D} \bar{\Sigma}_\varepsilon^{-1}.$$

Hence, $\hat{\beta}_{ILS}$ is not only $\sqrt{NT}$-consistent for β but also asymptotically normal.[2]

---

[2]In this paper we focus on the comparison between the CA and PC estimators, and do not spend time on issues such as implementation and inference, for which there are results available already (see for example Pesaran, 2004; Bai, 2009). However, it should be mentioned that in order to make inference based on the asymptotic normality results provided herein, typically all that is needed is an estimator of the asymptotic variance. For example, in order to make inference using $\hat{\beta}_{ILS}$ an estimator of Ω_{ILS} is needed, which can be obtained by simply replacing $\Sigma_{\varepsilon i}$ and D_i in $\bar{\Sigma}_\varepsilon$ and $\bar{D}$ with $\frac{1}{T} x_i' M_{\hat{F}^{IPC}} x_i$ and $\frac{1}{T}(y_i - x_i \hat{\beta}_{ILS} - \hat{F}^{IPC} \lambda_i)'(y_i - x_i \hat{\beta}_{ILS} - \hat{F}^{IPC} \lambda_i)$, respectively. The alternative would be to use $\hat{F}^{ICA}$ instead of $\hat{F}^{IPC}$.

Let us now consider replacing the true factors by their IPC and ICA estimators. In the next theorem we report asymptotic distributions of the resulting ILS estimators of β, henceforth denoted $\hat{\beta}_{IPC}$ and $\hat{\beta}_{ICA}$, respectively. Note that the ICA estimator can be seen as an infeasible version of the CCE estimator of Pesaran (2006).

Theorem 1. *Under Assumptions 1 and 2, as* N, $T \to \infty$ *with* $\frac{\sqrt{T}}{N} \to 0$,

$$\sqrt{NT}(\hat{\beta}_{IPC} - \hat{\beta}_{ILS}) = \bar{\Sigma}_\varepsilon^{-1}\sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC}) + o_p(1),$$
$$\sqrt{NT}(\hat{\beta}_{ICA} - \hat{\beta}_{ILS}) = \bar{\Sigma}_\varepsilon^{-1}\sqrt{T}N^{-1/2}(B_{1ICA} - B_{2ICA}) + o_p(1),$$

where

$$B_{1IPC} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \Lambda_i \bar{\Omega}_{IPC}\lambda_i,$$

$$B_{2IPC} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \Sigma_{\varepsilon i}((\beta\beta' + I_m)\Lambda_i + \beta\lambda_i')\bar{A}^{-1}\lambda_i,$$

$$B_{1ICA} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \Lambda_i \bar{\Omega}_{ICA}\lambda_i,$$

$$B_{2ICA} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \Sigma_{\varepsilon i}((\beta\beta' + I_m)\bar{\Lambda} + \beta\bar{\lambda}')(\bar{C}\bar{C}')^{-1}\lambda_i.$$

Remarks.

1. The fact that the estimators are biased in the current setup might come as a surprise. Indeed, since C_i is known, it should be possible to just add it as a regressor when estimating β. Since the regressors in a regression of y_i on x_i and C_i are exogenous, there should not be any bias. However, this is not how the estimation is carried out here. In particular, since the data are projected upon the estimated factors, the knowledge of C_i is not really used in this step of the estimation procedure, and it is this practice that leads to bias. The reason for doing the estimation in this way is that it works also when C_i is unknown.

2. Consider the IPC estimator. According to Theorem 1,

$$\sqrt{NT}(\hat{\beta}_{IPC} - \beta) = \sqrt{NT}(\hat{\beta}_{ILS} - \beta) + \sqrt{NT}(\hat{\beta}_{IPC} - \hat{\beta}_{ILS})$$
$$= \sqrt{NT}(\hat{\beta}_{ILS} - \beta) + \bar{\Sigma}_\varepsilon^{-1}\sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC}) + o_p(1),$$

suggesting that if we assume that $\bar{\Sigma}_\varepsilon \to \Sigma_\varepsilon$, $\frac{T}{N} \to \tau > 0$ and $\frac{\sqrt{T}}{N} \to 0$ as $N, T \to \infty$,

$$\sqrt{NT}(\hat{\beta}_{IPC} - \beta) \to_d N(0, \Omega_{ILS}) + \Sigma_\varepsilon^{-1}\sqrt{\tau}(B_{1IPC} - B_{2IPC}).$$

A similar result applies to the ICA estimator. Thus, if $\tau > 0$, while consistent, because of the correlation between e_i and x_i, the asymptotic distributions of the estimators are biased. On the other hand, if $\tau = 0$, then the bias is negligible and therefore $\sqrt{NT}(\hat{\beta}_{IPC} - \hat{\beta}_{ICA}) = o_p(1)$. Another possibility is if $\lambda_i = 0$, in which case the correlation between e_i and x_i is zero and therefore the bias disappears.

3. The bias in the case when $\frac{T}{N} \to \tau > 0$ depends critically on the choice of estimator of F_t. Unfortunately, in general one cannot say anything about the size and direction of this bias. However, suppose that $m = r = 1$, and that $\Sigma_{\varepsilon i} = \Sigma_\varepsilon$, $\Lambda_i = \Lambda$ and $\lambda_i = \lambda$. Then the difference in absolute bias of the ICA and IPC estimators is given by

$$\bar{\Sigma}_\varepsilon^{-1}\sqrt{\tau}(|B_{1ICA} - B_{2ICA}| - |B_{1IPC} - B_{2IPC}|) = \Sigma_\varepsilon^{-1}\sqrt{\tau}|\Lambda||\lambda|(\Omega_{ICA} - \Omega_{IPC}) = 0,$$

where the last equality follows from the fact that $\Omega_{ICA} = \Omega_{IPC}$ whenever $\Sigma_{\varepsilon i}$, Λ_i and λ_i are equal across i. In this case it is therefore clear that the two estimators as equally biased.

4. Proposition 1 and Theorem 1 make use of Assumption 2 requiring that $\bar{A}$, $\bar{B}$, $\bar{D}$, $\bar{\Sigma}_\varepsilon$ and $\bar{\Sigma}_u$ all have full rank and that $\text{rk}(\bar{C}) = r$. However, careful inspection of the results of these propositions reveal that the two estimators are actually based on different moment conditions, and that it is only when taken together that Assumption 2 is required. For example, consider Proposition 1. While the result for the PC estimator requires that $\bar{A}$, $\bar{B}$ and $\bar{D}$ have full rank, the corresponding result for the CA estimator requires that $\bar{D}$ and $\bar{\Sigma}_u$ have full rank and that $\text{rk}(\bar{C}) = r$.

5. In this paper F is estimated from z_i. This is not necessary. Bai (2009) considers a similar model, but where F is estimated by applying PC to $y_i - x_i\hat{\beta}_{LS}$, where $\hat{\beta}_{LS}$ is obtained by applying LS to (1). This estimator can then be used in a second step to obtain the associated IPC estimator of β. These steps can then be repeated until convergence. Bai (2009, Corollary 1) shows that in the special case when $\sigma_{\varepsilon i}^2 = \sigma_\varepsilon^2$ for all i, the resulting iterated IPC estimator of β is $\sqrt{NT}$-consistent and asymptotically normal as $N, T \to \infty$

with $\frac{T}{N} \to 0$.[3] The variance is in our notation given by $\sigma_\varepsilon^2 D_0^{-1}$, where

$$D_0 = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \left[x_i' M_F x_i - \sum_{j=1}^{N} x_i' M_F x_j \lambda_i' \left(\sum_{j=1}^{N} \lambda_i \lambda_i' \right)^{-1} \lambda_j \right],$$

which under (3) reduces to

$$D_0 = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' M_F \varepsilon_i = \lim_{N\to\infty} \bar{\Sigma}_\varepsilon.$$

Thus, since $\Omega_{ILS} = \sigma_\varepsilon^2 \lim_{N\to\infty} \bar{\Sigma}_\varepsilon^{-1}$ in this case, we have that under the additional requirement that $\frac{\sqrt{T}}{N} \to 0$ the PC and CA estimators considered here have the same asymptotic distributions as the one considered by Bai (2009). Another possibility is to estimate F from (3). However, in this case there must be no factors that are unique to e_i.

6. One advantage of using the approach of Bai (2009) is that the cross-section dependence in x_i does not need to have a factor structure. In fact, the only necessary requirement in this case is that x_i is independent of ϵ_i, see Bai (2009, Assumption D). However, it is important to note that if x_i does not have a factor structure, then the above results are no longer valid. Specifically, since in this case $x_i' M_F x_i \neq \varepsilon_i' M_F \varepsilon_i$, $\Sigma_{\varepsilon i}$ should be replaced by $\text{var}(x_i' M_F x_i)$, which is then assumed to be nonsingular, see Bai (2009, Assumption A) and Greenaway-McGrevy et al. (2010, Assumption B).[4]

7. Bai (2009) and Greenaway-McGrevy et al. (2010) relax the common factor assumption for x_i. However, since no alternative structure is assumed, the results for their PC estimators are not as clear as ours with the limiting distributions being expressed in terms of the parameters of the model.[5] They are also not as clear as ours when it comes to the restrictions placed on the relative expansion rate of N and T. For example, while Bai (2009, Theorem 3) only assumes that $\frac{T}{N} \to \tau > 0$, as Theorem 1 makes clear, in order to ensure that the various sample moments converge to their population counterparts, we also need $\frac{\sqrt{T}}{N} \to 0$. Thus, as always, there is a trade-off here between generality and clearness of results.

---

[3]Thus, while the first-step LS estimator of β is inconsistent, the iterated estimator is not.

[4]But while quite general when it comes to the allowable cross-sectional dependencies in x_i, the Bai (2009) approach also has some drawbacks. One drawback is that because $y_i - x_i \hat{\beta}_{LS}$ is used in place of e_i to estimate F, the second-step estimator may not be able to identify β (Greenaway-McGrevy et al., 2010). This can for example happen when there are common factors that are unique to x_i.

[5]Compare to Bai (2009, Theorem 3) and Greenaway-McGrevy et al. (2010, Theorem 1).

8. Pesaran (2006) relaxes Assumption 2.[6] However, this greater generality comes at the expense of additional assumptions regarding β and γ_i, which then have to admit a random coefficient representation. To illustrate this suppose that instead of (1) we have $y_i = x_i \beta_i + e_i$, where $\beta_i = \beta + v_i$, $v_i \sim \text{iid}(0, \Sigma_v)$ and $\Sigma_v > 0$. A similar assumption is assumed to hold for λ_i. Specifically, we assume that $\lambda_i = \lambda + w_i$, where w_i is again iid with mean zero and positive definite covariance matrix. v_i and w_i are mutually independent. It can be shown that

$$
\begin{aligned}
\sqrt{N}(\hat{\beta}_{ICA} - \beta) &= \left(\frac{1}{NT} \sum_{i=1}^{N} x_i' M_{\hat{F}^{ICA}} x_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i' M_{\hat{F}^{ICA}} (x_i v_i + F\lambda_i + \epsilon_i) \\
&= \left(\frac{1}{NT} \sum_{i=1}^{N} x_i' M_{\hat{F}^{ICA}} x_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i' M_{\hat{F}^{ICA}} (x_i v_i + F\lambda_i) + o_p(1).
\end{aligned}
$$

If Assumption 2 holds, then one can replace $M_{\hat{F}^{ICA}}$ with M_F, and therefore it is not difficult to see that

$$
\begin{aligned}
\sqrt{N}(\hat{\beta}_{ICA} - \beta) &= \left(\frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' M_F \varepsilon_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon_i' M_F \varepsilon_i v_i + o_p(1) \\
&= \bar{\Sigma}_\varepsilon^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\varepsilon i} v_i + o_p(1) \\
&\to_d \left(\lim_{N \to \infty} \bar{\Sigma}_\varepsilon^{-1} \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\varepsilon i} \Sigma_v \Sigma_{\varepsilon i} \bar{\Sigma}_\varepsilon^{-1} \right)^{1/2} N(0, I_m)
\end{aligned}
$$

as $T \to \infty$ and then $N \to \infty$. Thus, in this case the normality assumption is a direct consequence of the assumed randomness of β_i. A similar result applies in case Assumption 2 fails but then the normality stems from the assumed randomness of both β_i and λ_i. Thus, while Assumption 2 can be relaxed, this does not come free of charge, and, depending on the application, it is unclear which assumption is most restrictive. Note also that if β_i is allowed to vary, the rate of convergence is reduced, from $\sqrt{NT}$ to $\sqrt{N}$.

3.2 C_i unknown

Relaxing the assumption of known C_i has at least two important consequences. First, since F_t and C_i are no longer separately identifiable, F_t can only be estimated up to a matrix rotation.

---

[6]More precisely, Pesaran (2006) relaxes the assumption that the limit of the cross-sectional average of (λ_i, Λ_i') has rank $r \leq m + 1$.

Hence, instead of estimating F_t as in the previous section, now the best that we can hope for is consistent estimation of the space spanned by F_t. Second, the number of common factors, r, is unknown.

The feasible PC (FPC) estimator of F, denoted $\hat{F}^{FPC}$, is $\sqrt{T}$ times the matrix consisting of the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix zz'. The feasible CA (FCA) estimator is just the cross-sectional average of z_{it}, that is, $\hat{F}_t^{FCA} = \bar{z}_t$. To capture the fact that F_t, $\hat{F}_t^{FPC}$ and $\hat{F}_t^{FCA}$ may be of different dimension, we introduce the $r \times k$ matrix $H = \frac{1}{NT}C'CF'\hat{F}^{FPC}V^{-1}$ and the $r \times (m+1)$ matrix $G = \bar{C}$, where V is a $k \times k$ diagonal matrix containing the k largest eigenvalues of $\frac{1}{NT}zz'$ in descending order. These matrices will be use to rotate F, suggesting that something has to be assumed regarding their rank. Given Assumption 2 (b) the rank of G is clearly $r \leq m+1$. As for H, we make the following assumption.

Assumption 3. $\mathrm{rk}(H) = r \leq k$.

The asymptotic distributions of the feasible factor estimators are given in Proposition 2.

Proposition 2. *Under Assumptions 1–3, as N, T $\rightarrow \infty$ with $\frac{\sqrt{N}}{T} \rightarrow 0$,*

$$\sqrt{N}(\hat{F}_t^{FPC} - H'F_t) \quad \rightarrow_d \quad N(0, \Omega_{FPC}),$$
$$\sqrt{N}(\hat{F}_t^{FCA} - G'F_t) \quad \rightarrow_d \quad N(0, \Omega_{FCA}),$$

where

$$\Omega_{FPC} = \lim_{N,T\rightarrow\infty} H'\bar{\Omega}_{IPC}H,$$
$$\Omega_{FCA} = \lim_{N,T\rightarrow\infty} G'\bar{\Omega}_{ICA}G.$$

As expected, we see that the Ω_{FPC} and Ω_{FCA} depend on the rotation matrices H and G, thus making it impossible to compare the relative variance of the estimators. However, while C_i' and F_t are not identifiable, their product is. Proposition 3 therefore provides the limiting distributions of the PC and CA estimators of the common component, $C_i'F_t$. The PC and CA estimators of C_i are obtained from a LS regression of z_{it} onto the corresponding factor estimate.

Proposition 3. *Under Assumptions 1–3, as N, T → ∞,*

$$(N^{-1}C_i'\Omega_{IPC}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t) \quad \rightarrow_d \quad N(0,1),$$

$$(N^{-1}C_i'\Omega_{ICA}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}((\hat{C}_i^{FCA})'\hat{F}_t^{FCA} - C_i'F_t) \quad \rightarrow_d \quad N(0,1).$$

In contrast to Proposition 2, Proposition 3 holds regardless of the relative expansion rate of N and T. However, the results simplify if either $\frac{N}{T} \to 0$ or $\frac{T}{N} \to 0$. On the one hand, if $\frac{N}{T} \to 0$, then

$$(N^{-1}C_i'\Omega_{IPC}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t)$$

$$= (C_i'\Omega_{IPC}C_i + NT^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}\sqrt{N}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t)$$

$$= (C_i'\Omega_{IPC}C_i)^{-1/2}\sqrt{N}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t) + o_p(1).$$

suggesting that $(\hat{C}_i^{FPC})'\hat{F}_t^{FPC}$ is $\sqrt{N}$-consistent for $C_i'F_t$ and that the asymptotic variance is given by $C_i'\Omega_{IPC}C_i$. The rate of consistency for $(\hat{C}_i^{FCA})'\hat{F}_t^{FCA}$ is the same. However, since the variance of this estimator is given by $C_i'\Omega_{ICA}C_i$, the relative efficiency depends on the choice of estimator of F_t. In particular, since $C_i'\Omega_{ICA}C_i - C_i'\Omega_{IPC}C_i = C_i'(\Omega_{ICA} - \Omega_{IPC})C_i \geq 0$, the PC estimator of the common component is as least as efficient as the CA estimator. On the other hand, if $\frac{T}{N} \to 0$, then

$$(N^{-1}C_i'\Omega_{IPC}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t)$$

$$= (F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{-1/2}\sqrt{T}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t) + o_p(1).$$

The same result applies $(\hat{C}_i^{FCA})'\hat{F}_t^{FCA}$. Hence, in contrast to the case when $\frac{N}{T} \to 0$, since the variances are now the same, in this case the two estimators are equally efficient.

Proposition 4 provides the relevant asymptotic theory for the feasible estimators of β.

Proposition 4. *Under Assumptions 1–3, as N, T → ∞ with $\frac{\sqrt{T}}{N} \to 0$ and $\frac{\sqrt{N}}{T} \to 0$,*

$$\sqrt{NT}(\hat{\beta}_{FPC} - \hat{\beta}_{ILS}) = \sqrt{NT}(\hat{\beta}_{IPC} - \hat{\beta}_{ILS}) + o_p(1),$$

$$\sqrt{NT}(\hat{\beta}_{FCA} - \hat{\beta}_{ILS}) = \sqrt{NT}(\hat{\beta}_{ICA} - \hat{\beta}_{ILS}) + o_p(1).$$

Remarks.

1. Proposition 4 requires that $\frac{\sqrt{T}}{N}$ and $\frac{\sqrt{N}}{T}$ should both go to zero, which is satisfied if $\sqrt{T} < N < T^2$. The asymptotic distributions of $\sqrt{NT}(\hat{\beta}_{FPC} - \beta)$ and $\sqrt{NT}(\hat{\beta}_{FCA} - \beta)$ in this case can be deduced in the same way as in remark 1 to Theorem 1. They are normal with variance Ω_{ILS}. To also ensure that the mean is zero we need $\frac{T}{N} \to 0$, which together with the requirement that $\frac{\sqrt{T}}{N} \to 0$ implies that for valid inference based on the FPC and FCA estimators of β we need $T < N < T^2$.

2. As already mentioned the FPC and FCA estimators of the common component are equally efficient if $\frac{T}{N} \to 0$. Coincidentally, this is also the requirement for $\sqrt{NT}(\hat{\beta}_{FPC} - \beta)$ and $\sqrt{NT}(\hat{\beta}_{FCA} - \beta)$ to be asymptotically unbiased.

3. The fact that the FPC estimator is biased when $\frac{T}{N} \to \tau > 0$ is in agreement with the results reported by Bai (2009, Theorem 3) for his residual-based PC estimator, and as such it is not that surprising. The fact that also the FCA estimator is biased is, on the other hand, new to this paper. Proposition 4 can therefore be seen as a generalization of the results of Pesaran (2004) to the case when the assumption that $\frac{T}{N} \to 0$ is violated.

4. An obvious solution to the problem with bias in case $\frac{T}{N} \to \tau > 0$ is to use bias correction. In case of PC estimation, this means using $\hat{\beta}_{FPC} - \frac{1}{N}\hat{\Sigma}_\varepsilon^{-1}(\hat{B}_{1IPC} - \hat{B}_{2IPC})$ in place of $\hat{\beta}_{FPC}$, where $\hat{\Sigma}_\varepsilon = \frac{1}{NT}\sum_{i=1}^{N} x_i' M_{\hat{F}^{FPC}} x_i$ and $\hat{B}_{1IPC}$ and $\hat{B}_{2IPC}$ are B_{1IPC} and B_{2IPC}, respectively, with β, $\Sigma_{\varepsilon i}$ and $\sigma_{\varepsilon i}^2$ replaced by $\hat{\beta}_{FPC}$, $\hat{\Sigma}_{\varepsilon i} = \frac{1}{T}x_i' M_{\hat{F}^{FPC}} x_i$ and $\hat{\sigma}_{\varepsilon i}^2 = \frac{1}{T}(y_i - x_i\hat{\beta}_{FPC} - \hat{F}^{FPC}\hat{\lambda}_i^{FPC})'(y_i - x_i\hat{\beta}_{FPC} - \hat{F}^{FPC}\hat{\lambda}_i^{FPC})$, respectively. The estimators $\hat{\lambda}_i^{FPC}$ and $\hat{\Lambda}_i^{FPC}$ of λ_i and Λ_i, respectively, are obtained by simply picking the appropriate elements in $\hat{C}_i^{FPC}$.

5. Assumptions 2 (a) and 3 clarify an often overlooked fact in applied work, namely, that the problem of selecting k in PC estimation is the same as the problem of selecting $m + 1$ in CA estimation. According to these assumptions the assumed number of common factors must be at least as large as the true number, r. In case of the PC estimator this means that one has to put k large enough in the estimation of the factors, whereas in case of the CA estimator it means that the number of regressors must be large enough. Thus, the results reported in Propositions 2–4 for the PC estimator only require $r \leq k$ and the results for the CA estimator only require $r \leq m + 1$. Kapetanios and Pesaran (2006) claim that unlike the CA approach the implementation of the PC procedure requires the determination of the number of factors. But this ignores the fact that the

implementation of the CA approach requires setting $m = r - 1$, which is just as problematic as trying to set $k = r$. In fact, to the extent that more factors also require more regressors, one can even claim that the CA method is more demanding than PC.

4   Monte Carlo simulations

A small-scale simulation study was conducted to assess the accuracy of our theoretical results in small samples. The data generating process is a restricted version of the one given in (4), and sets $m = 1$, $r = 2$, $\beta = 1$ and $(F'_t, \epsilon_{it}, \varepsilon_{it})' \sim N(0, \text{diag}(I_2, \sigma^2_{\epsilon}, \Sigma_{\varepsilon i}))$. Also, $\Sigma_{\varepsilon i} \sim a + bU(0, 1)$, where we use a and b to control the level and heterogeneity of $\Sigma_{\varepsilon i}$. To ensure that C_i is positive definite, we further set $\Lambda_i = \Lambda = (0, 1)$ and use c to determine the heterogeneity of $\lambda_i \sim (1, 1)' + cU(0, 1)$. The data are generated for 3,000 panels with either $N = T$ or $N = T^{4/3}$ and $T = 25, 50, 100, 200$. If $N = T$, the CA and PC estimators should be biased, whereas if $N = T^{4/3}$, they should be asymptotically unbiased.

We focus on the estimation of β. Specifically, in order to verify Theorem 1 and Proposition 5, we report the mean and standard deviation of $\sqrt{NT}(\hat{\beta}_{ICA} - \hat{\beta}_{ILS})$, $\sqrt{NT}(\hat{\beta}_{IPC} - \hat{\beta}_{ILS})$, $\sqrt{NT}(\hat{\beta}_{FCA} - \hat{\beta}_{ILS})$ and $\sqrt{NT}(\hat{\beta}_{FPC} - \hat{\beta}_{ILS})$. The theoretical bias obtained from evaluating the formula given in Theorem 1 is also reported. As a measure of the relevance of these results for inference, we also report the size of a double-sided t-test for the null hypothesis of $\beta = 1$ when the 5% critical value of 1.96 is used. The feasible estimators are implemented using the true number of factors and their standard errors are computed using the standard LS formula.

The results reported in Tables 1 and 2 are generally in agreement with theory and can be summarized as follows:

- While the estimators are on average equally biased when $b = c = 0$ such that $\Sigma_{\varepsilon i} = \Sigma_{\varepsilon}$ and $\lambda_i = \lambda$, this is not the case when b and c are set differently. In particular, based on the results reported here, the two CA estimators consistently outperform their PC counterparts. To take an extreme example, consider the case when $N = T$, $a = 3$ and $c = 4$, in which the theoretical PC bias according to Theorem 1 is roughly -1.5.[7] As the table makes clear, this is a very accurate prediction of the empirical IPC and FPC bias. The CA, on the other hand, seems to perform very well with a bias close to zero.

---

[7]The bias results in this case is of similar magnitude as the ones reported by Greenaway-McGrevy *et al.* (2010).

- The bias is generally very close to what is predicted by our asymptotic theory. The only exception is for the FPC estimator, which tends to be more biased than expected, especially when $b = c = 0$. However, the discrepancy vanishes quickly as T increases, suggesting that the bias in this case is merely a small-sample effect. We also see that the extent of this effect depends on a and σ_ϵ^2, which determine the variance of the idiosyncratic component relative to that of the common component. As expected, the bias is increasing in a and σ_ϵ^2.

- The size of the bias depend on whether $N = T$ or $N = T^{4/3}$. If $N = T$, except possibly for FPC estimator when $b = c = 0$, we see that the bias is flat in T. By contrast, if $N = T^{4/3}$, then the bias is clearly decreasing in T. This is in agreement with Theorem 1 showing that the bias is $O(\sqrt{T/N})$, which is $O(1)$ when $N = T$ and $O(T^{-1/6})$ when $N = T^{4/3}$. The effect is especially pronounced for the PC estimators being severely biased in the case when $N = T$.

- As expected, regardless of whether $N = T$ or $N = T^{4/3}$, the standard deviation of the simulated bias is decreasing in T. We also see that the rate at which the bias vanishes is roughly of the theoretically predicted rate of $\frac{1}{T}$.

- The size of the t-tests depend critically on the extent of bias. In particular, we see that there is a clear tendency for the size distortions to increase with the size of the bias. The PC estimators are much more biased than the CA estimators, which translates into relatively large size distortions. In fact, based on the results reported here, it is not unusual for these tests to have sizes that are in excess of 50% when they should be 5%. However, as expected the size distortions are substantially reduced when $N = T^{4/3}$ and as T increases.

5 Conclusions and implications for empirical work

The results obtained here are interesting in their own right but also because of the implications they have for applied work. The fact that the CA estimator of the factor-augmented regression, which is the CCE of Pesaran (2006), is in general less biased than PC is, for example, very useful. Specifically, since the variances of the two estimators are the same, it means that in practice one is unlikely to do better than when using the relatively simple CA

approach.

In this paper we have assumed that the true number of factors, r, is known, which is not the case in practice. However, it is reasonable to assume that r is bounded by a sufficiently large integer value, r_{max} say. One way to implement the CA and PC estimators is to simply set $r_{max} = k = m + 1$, thereby allowing true number of factors to be anything between zero and r_{max}. Thus, in this case we are not interested in the number of factors but just want to make sure that the effect of the cross-section dependence has been accounted for. Of course, with r_{max} large, for the CA estimator this approach is not really feasible. As an alternative in such cases one may chose r using some model selection criteria, such as an information criterion. This approach has been shown to work in the context of a conventional common factor model (Bai and Ng, 2002) and it is expected to work well also in the current regression setting. Note also that all results reported in this paper are based on the assumption that $r \leq m + 1$ in case of the CA estimator and $r \leq k$ in case of the PC estimator. If these assumptions are violated, which in practice means setting r_{max} too low, then the results reported here are no longer valid.

Chudik *et al.* (2010) introduce the notions of weak, semi-strong and strong factor structures and prove that these do not affect the consistency of the FCA estimator, a result that is partly confirmed by our simulations. The FPC estimator, on the other hand, seems to be severely small-sample biased even if the cross-section dependence is absent. The implication is that if the objective is just to mop up the effect of a relatively weak common component, then the FCA estimator seems like the best choice.

The relative magnitude of N and T has important implications for performance, not only in small samples but also asymptotically. Thus, given the size of the potential bias, unless $N > T$, some kind of bias-correction might be appropriate (see Bai, 2009). A natural candidate towards this end consists of subtracting from FCA and FPC an estimator of the bias given in Theorem 1. For example, instead of $\hat{\beta}_{FCA}$ we use $\hat{\beta}_{FCA} - N^{-1}\bar{\Sigma}_{\varepsilon}^{-1}(B_{1ICA} - B_{2ICA})$, where the unknown coefficients in $\bar{\Sigma}_{\varepsilon}$, B_{1ICA} and B_{2ICA} should be replaced by estimates. In case of the FPC estimator, given its small-sample bias, this is probably not enough. Similarly, even if $N > T$, for the FPC estimator to work properly one also needs N and T relatively large.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**, 135–173.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77**, 1229–1279.

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

Bai, J., C. Kao and S. Ng (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics* **149**, 82–99.

Chudik, A., M. H. Pesaran and E. Tosetti (2010). Weak and strong cross section dependence and estimation of large panels. Forthcoming in *Econometrics Journal*.

Coakley, J., A. Fuertes and R. P. Smith (2002). A principal components approach to cross-section dependence in panels. Unpublished manuscript.

Greenaway-McGrevy, R., C. Han and D. Sul (2010). Asymptotic distribution of factor augmented estimators for panel regression. Forthcoming in *Journal of Econometrics*.

Kapetanios, G., and M. H. Pesaran (2006). Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. CESifo Working Paper Series 1416.

Kapetanios, G., and M. H. Pesaran (2010). Panels with nonstationary multifactor error structures. Forthcoming in *Journal of Econometrics*.

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

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

Appendix: Proofs

Proof of Proposition 1.

Define the $N(m+1) \times N(m+1)$ covariance matrix $\Sigma_u = \mathrm{diag}(\Sigma_{u1}, ..., \Sigma_{uN})$. Thus, letting $J = \iota_N \otimes I_{m+1}$, where $\iota_N = (1, ..., 1)'$ is a N-dimensional vector of ones, $\bar{A}$, $\bar{B}$, $\bar{C}$ and $\bar{\Sigma}_u$ can be rewritten as

$$
\begin{aligned}
\bar{A} &= \frac{1}{N} C'C, \\
\bar{B} &= \frac{1}{N} C'\Sigma_u C, \\
\bar{C} &= \frac{1}{N} C'J, \\
\bar{\Sigma}_u &= \frac{1}{N} J'\Sigma_u J.
\end{aligned}
$$

In this notation,

$$
\begin{aligned}
\bar{\Omega}_{IPC} &= N(C'C)^{-1} C'\Sigma_u C (C'C)^{-1}, \\
\bar{\Omega}_{ICA} &= N(C'JJ'C)^{-1} C'JJ'\Sigma_u JJ'C (C'JJ'C)^{-1}.
\end{aligned}
$$

We now show that $\bar{\Omega}_{ICA} - \bar{\Omega}_{IPC}$ is positive semidefinite, which holds if $\bar{R} = \bar{\Omega}_{IPC}^{-1} - \bar{\Omega}_{CA}^{-1}$ is positive semidefinite. Thus, letting

$$
\begin{aligned}
V &= C'C, \\
S &= C'\Sigma_u C, \\
W &= C'(CC')^{-} JJ'(CC')^{-} C,
\end{aligned}
$$

we have

$$
\begin{aligned}
VWSWV &= (C'C)\,(C'(CC')^{-} JJ'(CC')^{-} C)\,(C'\Sigma_u C)\,(C'(CC')^{-} JJ'(CC')^{-} C)\,(C'C) \\
&= (C'C)C'(CC')^{-} JJ'\Sigma_u JJ'(CC')^{-} C(C'C) = C'JJ'\Sigma_u JJ'C, \\
VWV &= (C'C)\,(C'(CC')^{-} JJ'(CC')^{-} C)\,(C'C) = C'JJ'C,
\end{aligned}
$$

which in turn implies

$$
\begin{aligned}
\bar{R} &= \frac{1}{N}(C'C)(C'\Sigma_u C)^{-1}(C'C) - \frac{1}{N} C'JJ'C(C'JJ'\Sigma_u JJ'C)^{-1} C'JJ'C \\
&= \frac{1}{N} VS^{-1}V - \frac{1}{N} VWV(VWSWV)^{-1} VWV.
\end{aligned}
$$

Moreover, since $C'\Sigma_u C$ is positive definite, there exists a nonsingular $(m+1) \times (m+1)$ matrix L such that $(C'\Sigma_u C)^{-1} = L'L$. Let $H = LV$ and $M_G = I_{m+1} - G(G'G)^{-1}G'$, where $G = (L')^{-1}WV$. It follows that

$$
\begin{aligned}
\bar{R} &= \frac{1}{N}VL'LV - \frac{1}{N}VWV(VWL^{-1}(L')^{-1}WV)^{-1}VWV \\
&= \frac{1}{N}H'H - \frac{1}{N}H'G(G'G)^{-1}G'H = H'M_G H.
\end{aligned}
$$

But M_G is symmetric and idempotent, as $M_G' = M_G$ and $M_G M_G = M_G$, suggesting that for any $(m+1)$-dimensional vector v,

$$
v'\bar{R}v = v'H'M_G H v = w'M_G w \geq 0,
$$

where $w = Hv$ and the inequality holds because M_G is positive semidefinite. Therefore, $\bar{R}$ is positive semidefinite. ∎

Proof of Theorem 1.

We begin by considering the PC estimator. Write

$$
y_i = x_i\beta + \hat{F}^{IPC}\lambda_i - d\lambda_i + \epsilon_i,
$$

where $d = \hat{F}^{IPC} - F$. The least squares estimator of β is given by

$$
\hat{\beta}_{IPC} = \left(\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}x_i \right)^{-1} \sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}y_i,
$$

suggesting that

$$
\begin{aligned}
\sqrt{NT}(\hat{\beta}_{IPC} - \beta) &= \left(\frac{1}{NT}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}x_i \right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}\epsilon_i \\
&\quad - \left(\frac{1}{NT}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}x_i \right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}d\lambda_i.
\end{aligned}
$$

We begin by considering the second term on the right-hand side. Clearly,

$$
\begin{aligned}
-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}d\lambda_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} x_i'M_{\hat{F}^{IPC}}F\lambda_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \Lambda_i F'M_{\hat{F}^{IPC}}F\lambda_i \\
&\quad + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \epsilon_i'M_{\hat{F}^{IPC}}F\lambda_i = R_1 + R_2,
\end{aligned}
$$

with R_1 and R_2 implicitly defined.

Consider R_1, where $F'M_{\hat{F}^{IPC}}F$ can be expanded as

$$F'M_{\hat{F}^{IPC}}F = d'M_{\hat{F}^{IPC}}d = d'M_Fd - d'(M_F - M_{\hat{F}^{IPC}})d$$

From the definitions of M_F and $M_{\hat{F}^{IPC}}$,

$$
\begin{aligned}
d'(M_F - M_{\hat{F}^{IPC}})d &= d'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'd + d'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}F'd \\
&+ d'F((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'd + d'F[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}]F'd.
\end{aligned}
$$

Write

$$
\begin{aligned}
((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1} &= ((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}((\hat{F}^{IPC})'\hat{F}^{IPC} - F'F)(F'F)^{-1} \\
&= ((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}(d'\hat{F}^{IPC} + F'd)(F'F)^{-1},
\end{aligned}
$$

where

$$\sqrt{N}T^{-1/2}F'd = \frac{\sqrt{N}}{\sqrt{T}}\sum_{t=1}^{T}F_td_t' = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}F_tu_{it}'C_i'\bar{A}^{-1} = O_p(1),$$

with $d_t = \hat{F}_t^{IPC} - F_t$. Similarly, since

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}C_iu_{it}u_{jt}'C_j' \to_p \bar{B}$$

as $T \to \infty$, we have

$$
\begin{aligned}
\sqrt{N}T^{-1/2}d'\hat{F}^{IPC} &= \frac{\sqrt{N}}{\sqrt{T}}\sum_{t=1}^{T}d_t(\hat{F}_t^{IPC})' = \bar{A}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}C_iu_{it}(\hat{F}_t^{IPC})' \\
&= \bar{A}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}C_iu_{it}F_t' + \bar{A}^{-1}\frac{1}{N^{3/2}\sqrt{T}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}C_iu_{it}u_{jt}'C_j'\bar{A}^{-1} \\
&= \bar{A}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}C_iu_{it}F_t' + \frac{\sqrt{T}}{\sqrt{N}}\bar{A}^{-1}\bar{B}\bar{A}^{-1} + o_p(1) \\
&= O_p(1) + O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
T||((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}|| &\leq ||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||T^{-1}||d'\hat{F}^{IPC} + F'd|| \\
&\times ||(T^{-1}F'F)^{-1}|| \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right).
\end{aligned}
$$

But we also have

$$NT^{-1}d'd = \frac{N}{T}\sum_{t=1}^{T}d_t d_t' = \bar{A}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}C_i\frac{1}{T}\sum_{t=1}^{T}u_{it}u_{jt}'C_j'\bar{A}^{-1} \to_p \bar{A}^{-1}\bar{B}\bar{A}^{-1} = \bar{\Omega}_{IPC}$$

as $T \to \infty$, from which it follows that

$$
\begin{aligned}
||T^{-1}d'(M_F - M_{\hat{F}^{IPC}})d|| &\leq ||T^{-1}d'd||\,||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||\,||T^{-1}d'd|| \\
&+ ||T^{-1}d'd||\,||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||\,||T^{-1}F'd|| \\
&+ ||T^{-1}d'F||\,||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||\,||T^{-1}d'd|| \\
&+ ||T^{-1}d'F||\,||T||\,((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}||\,||T^{-1}F'd|| \\
&= O_p\left(\frac{1}{N^2}\right) + O_p\left(\frac{1}{N^{3/2}\sqrt{T}}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'(M_F - M_{\hat{F}^{IPC}})d\lambda_i\right\| &\leq \sqrt{NT}||T^{-1}d'(M_F - M_{\hat{F}^{IPC}})d||\frac{1}{N}\sum_{i=1}^{N}||\Lambda_i||\,||\lambda_i|| \\
&= O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right).
\end{aligned}
$$

Moreover, since $d'M_F d = d'd - d'F(F'F)^{-1}F'd$ with

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'F(F'F)^{-1}F'd\lambda_i\right\| &\leq \frac{1}{\sqrt{NT}}||\sqrt{N}T^{-1/2}d'F||\,||(T^{-1}F'F)^{-1}|| \\
&\times ||\sqrt{N}T^{-1/2}F'd||\frac{1}{N}\sum_{i=1}^{N}||\Lambda_i||\,||\lambda_i|| = O_p\left(\frac{1}{\sqrt{NT}}\right),
\end{aligned}
$$

we obtain

$$
\begin{aligned}
R_1 &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'M_F d\lambda_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'(M_F - M_{\hat{F}^{IPC}})d\lambda_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'M_F d\lambda_i + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Lambda_i d'd\lambda_i + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right).
\end{aligned}
$$

As for the first term on the right-hand side, we have

$$\frac{1}{T}\sum_{i=1}^{N}\Lambda_i d'd\lambda_i = \frac{1}{N}\sum_{i=1}^{N}\Lambda_i(NT^{-1}d'd)\lambda_i \to_p B_{1IPC}$$

as $N, T \to \infty$, where

$$B_{1IPC} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\bar{\Omega}_{IPC}\lambda_i,$$

23

which in turn suggests

$$R_1 - \sqrt{T}N^{-1/2}B_{1IPC} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right).$$

Next, consider R_2. Clearly,

$$
\begin{aligned}
R_2 &= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'(M_F - M_{\hat{F}^{IPC}})F\lambda_i \\
&= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}F'F\lambda_i \\
&\quad - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'F((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'F[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}]F'F\lambda_i.
\end{aligned}
$$

Since $d_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d_s$ is a scalar, the order of the first term can be inferred as follows:

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}d_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}\sum_{s=1}^{T}d_sF_s'\lambda_i \\
&= \sqrt{T}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}d_t'(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d_s\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}F_s'\lambda_i \\
&\leq \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}||d_t||^2||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}|| \\
&\quad \times \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}F_s'\lambda_i\right\|^2\right)^{1/2} = O_p\left(\frac{\sqrt{T}}{N}\right).
\end{aligned}
$$

Similarly, since $F_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i$ is scalar, the third term can be written as

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'F((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}F_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'F\lambda_i \\
&= \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}F_t'(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}\sqrt{N}T^{-1/2}d'F\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it} \\
&\leq \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}||F_t||^2\right)^{1/2}||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||\,||\sqrt{N}T^{-1/2}d'F|| \\
&\quad \times \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}\right\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

The corresponding result for the fourth term is given by

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}T^{-1/2}\varepsilon_i'FT[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}-(F'F)^{-1}]T^{-1}F'F\lambda_i$$

$$= \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}F_t'T[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}-(F'F)^{-1}]T^{-1}F'F\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}$$

$$\leq \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}||F_t||^2\right)^{1/2}T||((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}-(F'F)^{-1}||\,||T^{-1}F'F||$$

$$\times \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\varepsilon_{it}\right\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{\sqrt{T}}{N}\right).$$

As for the second term, by using Taylor expansion of $(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}$ around $(T^{-1}F'F)^{-1}$,

$$\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i'd(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}T^{-1}F'F\lambda_i = \frac{1}{T}\sum_{i=1}^{N}\varepsilon_i'd\lambda_i + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Moreover, since $E(\varepsilon_{it}u_{it}') = \Sigma_{\varepsilon i}(\beta, I_m)$ and $(\beta, I_m)C_i' = ((\beta\beta' + I_m)\Lambda_i + \beta\lambda_i')$, we obtain

$$\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i'd\lambda_i = \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}d_t'\lambda_i = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}u_{jt}'C_j'\bar{A}^{-1}\lambda_i \to_p B_{2IPC}$$

as $N, T \to \infty$, where

$$B_{2IPC} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Sigma_{\varepsilon i}((\beta\beta' + I_m)\Lambda_i + \beta\lambda_i)\bar{A}^{-1}\lambda_i',$$

suggesting

$$R_2 + \sqrt{T}N^{-1/2}B_{2IPC} = O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{\sqrt{T}}{N}\right).$$

Thus, by adding the results,

$$R_1 + R_2 - \sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC}) = -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}d\lambda_i - \sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC})$$

$$= O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{\sqrt{T}}{N}\right).$$

Next, consider

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}\epsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_F\epsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'(M_F - M_{\hat{F}^{IPC}})\epsilon_i$$

where

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'(M_F - M_{\hat{F}^{IPC}})\epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'\epsilon_i$$

$$+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}F'\epsilon_i$$

$$+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'F((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'\epsilon_i$$

$$+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'F[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}]F'\epsilon_i.$$

The order of these terms can be obtained by using the same steps as when analyzing R_2. The order of the first term is given by

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d'\epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}d_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} \sum_{s=1}^{T} d_s\epsilon_{is}$$

$$= \sqrt{T}\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} d_t'(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}d_s \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{it}\epsilon_{is}$$

$$\leq \sqrt{T}\frac{1}{T} \sum_{t=1}^{T} ||d_t||^2||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||$$

$$\times \left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left|\left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{it}\epsilon_{is} \right|\right|^2 \right)^{1/2} = O_p\left(\frac{\sqrt{T}}{N}\right).$$

Similarly,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i'd((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}F'\epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}d_t'((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} \sum_{s=1}^{T} F_s\epsilon_{is}$$

$$= \frac{1}{T} \sum_{t=1}^{T} d_t'(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{T} F_s x_{it}\epsilon_{is}$$

$$\leq \left(\frac{1}{T} \sum_{t=1}^{T} ||d_t||^2 \right)^{1/2} ||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}||$$

$$\times \left(\frac{1}{T} \sum_{t=1}^{T} \left|\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{T} F_s x_{it}\epsilon_{is} \right|\right|^2 \right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right),$$

with the third term being of the same order. The order of the fourth term is given by

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} T^{-1}x_i'FT[((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}]T^{-1/2}F'\epsilon_i$$

$$\leq \sqrt{N} \left(\frac{1}{N} \sum_{i=1}^{N} ||T^{-1}x_i'F||^2 \right)^{1/2} T||((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}|| \left(\frac{1}{N} \sum_{i=1}^{N} ||T^{-1/2}F'\epsilon_i||^2 \right)^{1/2}$$

$$= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

26

It follows that

$$\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'(M_F - M_{\hat{F}^{IPC}})\epsilon_i\right\| = O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which in turn implies

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}\epsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_F\epsilon_i + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Next, consider

$$\frac{1}{NT}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}x_i = \frac{1}{NT}\sum_{i=1}^{N}x_i'M_F x_i - \frac{1}{NT}\sum_{i=1}^{N}x_i'(M_F - M_{\hat{F}^{IPC}})x_i.$$

We have $NT^{-1}||d||^2 = O_p(1)$, implying $\sqrt{N}T^{-1/2}||d|| = O_p(1)$. Similarly, since $T^{-1}||F||^2 = T^{-1}\sum_{t=1}^{T}||F_t||^2 = O_p(1)$, we have $T^{-1/2}||F|| = O_p(1)$. Hence,

$$\begin{aligned}||M_F - M_{\hat{F}^{IPC}}|| &= T^{-1}||d||^2||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}|| + 2T^{-1}||d||\,||F||\,||(T^{-1}(\hat{F}^{IPC})'\hat{F}^{IPC})^{-1}|| \\ &+ T^{-1}||F||^2 T||((\hat{F}^{IPC})'\hat{F}^{IPC})^{-1} - (F'F)^{-1}|| = O_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

implying

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}x_i'(M_F - M_{\hat{F}^{IPC}})x_i\right\| \leq ||M_F - M_{\hat{F}^{IPC}}||\frac{1}{NT}\sum_{i=1}^{N}||x_i||^2 = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Therefore,

$$\frac{1}{NT}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}x_i = \frac{1}{NT}\sum_{i=1}^{N}x_i'M_F x_i + O_p\left(\frac{1}{\sqrt{N}}\right).$$

By adding all the results,

$$\begin{aligned}\sqrt{NT}(\hat{\beta}_{IPC} - \beta) &= \left(\frac{1}{NT}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}x_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}\epsilon_i \\ &- \left(\frac{1}{NT}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}x_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{IPC}}d\lambda_i \\ &= \left(\frac{1}{NT}\sum_{i=1}^{N}x_i'M_F x_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_F\epsilon_i + \bar{\Sigma}_\epsilon^{-1}\sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC}) \\ &+ O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right) \\ &= \sqrt{NT}(\hat{\beta}_{ILS} - \beta) + \bar{\Sigma}_\epsilon^{-1}\sqrt{T}N^{-1/2}(B_{1IPC} - B_{2IPC}) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &+ O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right),\end{aligned}$$

which completes the proof for the PC estimator. The proof for the CA estimator is almost identical, and the details are therefore omitted. The only difference is the bias terms, B_{1IPC} and B_{2IPC}. Indeed, redefining $d = \hat{F}_{CA} - F$, we have

$$B_{1ICA} = \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Lambda_i (NT^{-1} d'd)\lambda_i = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Lambda_i \bar{\Omega}_{ICA}\lambda_i.$$

Similarly, since $(\beta, I_m)\bar{C}' = ((\beta\beta' + I_m)\bar{\Lambda} + \beta\bar{\lambda}')$, we obtain

$$
\begin{aligned}
\frac{1}{T} \sum_{i=1}^{N} \varepsilon_i' d\lambda_i &= \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} d_t' \lambda_i = \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \bar{u}_t' \bar{C}' (\bar{C}\bar{C}')^{-1} \lambda_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} u_{jt}' \bar{C}' (\bar{C}\bar{C}')^{-1} \lambda_i \to_p B_{2ICA}
\end{aligned}
$$

as $N, T \to \infty$, where

$$B_{2ICA} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\varepsilon i} ((\beta\beta' + I_m)\bar{\Lambda} + \beta\bar{\lambda}')(\bar{C}\bar{C}')^{-1}\lambda_i.$$

This establishes the proof. ∎

Proof of Proposition 2.

We begin by considering the PC estimator. As in Bai (2003, page 158), if we denote by V the $k \times k$ diagonal matrix consisting of the first k eigenvalues of $\frac{1}{NT}zz'$ in descending order, then, by the definition of eigenvalues and eigenvectors, $\hat{F}^{FPC} = \frac{1}{NT}zz'\hat{F}^{FPC}V^{-1}$. Thus, letting $H = \frac{1}{NT}C'CF'\hat{F}^{FPC}V^{-1}$, we have

$$
\begin{aligned}
\hat{F}^{FPC} - FH &= \frac{1}{NT}zz'\hat{F}^{FPC}V^{-1} - \frac{1}{NT}FC'CF'\hat{F}^{FPC}V^{-1} \\
&= \frac{1}{NT}(zz' - FC'CF')\hat{F}^{FPC}V^{-1} \\
&= \frac{1}{NT}(uu' + uCF' + FC'u')\hat{F}^{FPC}V^{-1},
\end{aligned}
$$

or, in vector notation,

$$\hat{F}_t^{FPC} - H'F_t = \frac{1}{NT}V^{-1}(\hat{F}^{FPC})'(uu_t' + FC'u_t + uCF_t).$$

Now, by using Lemma A.2 of Bai (2003),

$$
\begin{aligned}
\frac{1}{NT}(\hat{F}^{FPC})'uu_t' &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right), \\
\frac{1}{NT}(\hat{F}^{FPC})'uCF_t &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}
$$

Hence, since $V = O_p(1)$,

$$
\begin{aligned}
\sqrt{N}(\hat{F}_t^{FPC} - H'F_t) &= \sqrt{N}\frac{1}{NT}V^{-1}(\hat{F}^{FPC})'FC'u_t + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{N}}{T}\right) \\
&= \sqrt{N}H'(C'C)^{-1}C'u_t + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{N}}{T}\right) \\
&\to_p N(0, \Omega_{FPC})
\end{aligned}
$$

as $N, T \to \infty$ with $\frac{\sqrt{N}}{T} \to 0$, where the second equality holds because, via the definition of H, $\frac{1}{NT}V^{-1}(\hat{F}^{FPC})'F = H'(C'C)^{-1}$.

Let us now consider the CA estimator, in which case $\hat{F}_t^{FCA} = \bar{z}_t$. Because $G = \bar{C}$, we have that $\bar{z}_t = \bar{C}'F_t + \bar{u}_t = G'F_t + \bar{u}_t$. Hence, since

$$
G'\bar{\Omega}_{ICA}G = \bar{C}'(\bar{C}\bar{C}')^{-1}\bar{C}\bar{\Sigma}_u\bar{C}'(\bar{C}\bar{C}')^{-1}\bar{C} = \bar{C}'(\bar{C}^-)'\bar{C}^-\bar{C}\bar{\Sigma}_u\bar{C}'(\bar{C}^-)'\bar{C}^-\bar{C} = \bar{\Sigma}_u,
$$

we can show that

$$
\sqrt{N}(\hat{F}_t^{FCA} - G'F_t) = \sqrt{N}\bar{u}_t \to_d N(0, G'\bar{\Omega}_{ICA}G)
$$

as $N, T \to \infty$, and so the proof is complete. $\blacksquare$

Proof of Proposition 3.

We begin by considering the PC estimator. Note that

$$
\begin{aligned}
T^{-1}(\hat{F}^{FPC})'z_i &= T^{-1}(\hat{F}^{FPC})'FC_i + T^{-1}(\hat{F}^{FPC})'u_i \\
&= T^{-1}H'F'FC_i + T^{-1}(\hat{F}^{FPC} - FH)'FC_i + T^{-1}(\hat{F}^{FPC})'u_i \\
&= T^{-1}H'F'FC_i + T^{-1}H'F'u_i + T^{-1}(\hat{F}^{FPC} - FH)'FC_i + T^{-1}(\hat{F}^{FPC} - FH)'u_i
\end{aligned}
$$

where, by Lemmas B.1 and B.2 of Bai (2003),

$$
T^{-1}(\hat{F}^{FPC} - FH)'FC_i = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),
$$

with $T^{-1}(\hat{F}^{FPC} - FH)'u_i$ having the same order. Moreover, by using the same argument as in the proof of Theorem 1,

$$
(T^{-1}(\hat{F}^{FPC})'\hat{F}^{FPC})^{-1} - (T^{-1}H'F'FH)^{-1} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right).
$$

It follows that

$$
\begin{aligned}
\hat{C}_i^{FPC} &= ((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}(\hat{F}^{FPC})'z_i = H^-C_i + (T^{-1}H'F'FH)^{-1}T^{-1}H'F'u_i \\
&+ O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

which can be used, together with Proposition 2, to show that

$$
\begin{aligned}
(\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t &= C_i'(H')^-(\hat{F}_t^{FPC} - H'F_t) + (\hat{C}_i^{FPC} - H^-C_i)'\hat{F}_t^{FPC} \\
&= C_i'(H')^-(\hat{F}_t^{FPC} - H'F_t) + (\hat{C}_i^{FPC} - H^+C_i)'H'F_t \\
&+ (\hat{C}_i^{FPC} - H^-C_i)'(\hat{F}_t^{FPC} - H'F_t) \\
&= C_i'(H')^-H'(C'C)^{-1}C'u_t + u_i'FH(H'F'FH)^{-1}H'F_t \\
&+ O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) \\
&= C_i'(C'C)^{-1}C'u_t + u_i'F(F'F)^{-1}F_t + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right) \\
&+ O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

where the last equality uses that

$$
\begin{aligned}
u_i'FH(H'F'FH)^{-1}H'F_t &= u_i'F(H'F'F)^-(H'F'F)H(H'F'FH)^{-1}(H'F'FH)(H'F'FH)^{-1} \\
&\times H'(F'FH)(F'FH)^-F_t = u_i'F(H'F'F)^-(H'F'FH)(F'FH)^-F_t \\
&= u_i'F(F'F)^{-1}F_t.
\end{aligned}
$$

As for the first term on the right-hand side, we have

$$
\sqrt{N}C_i'(C'C)^{-1}C'u_t = C_i'(N^{-1}C'C)^{-1}N^{-1/2}C'u_t \to_d N(0, C_i'\Omega_{IPC}C_i)
$$

as $N \to \infty$. Moreover, since

$$
\begin{aligned}
TE(u_i'F(F'F)^{-1}(F'F)^{-1}F'u_i) &= T\sum_{t=1}^{T}\sum_{s=1}^{T}E(u_{it}F_t'(F'F)^{-1}(F'F)^{-1}F_s u_{is}') \\
&= T\sum_{t=1}^{T}E(F_t'(F'F)^{-1}(F'F)^{-1}F_t)E(u_{it}u_{it}') \\
&= TE\left[\operatorname{tr}\left(\sum_{s=1}^{T}F_tF_t'(F'F)^{-1}(F'F)^{-1}\right)\right]\Sigma_{ui} \\
&= E[\operatorname{tr}((T^{-1}F'F)^{-1})]\Sigma_{ui} \to \operatorname{tr}(\Sigma_F^{-1})\Sigma_{ui}
\end{aligned}
$$

as $T \to \infty$, we get

$$
\sqrt{T}u_i'F(F'F)^{-1}F_t = T^{-1/2}u_i'F(T^{-1}F'F)^{-1}F_t \to_d N(0, F_t'\Sigma_F^{-1}F_t\Sigma_{ui}).
$$

The fact that $N^{-1/2}C'u_t$ and $T^{-1/2}u_i'F$ are based on different summations implies that their asymptotic distributions are independent. Hence,

$$(N^{-1}C_i'\Omega_{IPC}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{1/2}((\hat{C}_i^{FPC})'\hat{F}_t^{FPC} - C_i'F_t) \to_d N(0,1)$$

as $N, T \to \infty$.

In case of the CA estimator, we simply replace H with G, giving

$$
\begin{aligned}
(\hat{C}_i^{FCA})'\hat{F}_t^{FCA} - C_i'F_t &= C_i'(G^-)'(\hat{F}_t^{FCA} - G'F_t) + (\hat{C}_i^{FCA} - G^-C_i)'G'F_t + o_p(1) \\
&= C_i'(\bar{C}^-)'\bar{u}_t + u_i'F(F'F)^{-1}F_t + o_p(1) \\
&= C_i'(\bar{C}\bar{C}')^{-1}\bar{C}\bar{u}_t + u_i'F(F'F)^{-1}F_t + o_p(1),
\end{aligned}
$$

suggesting

$$(N^{-1}C_i'\Omega_{ICA}C_i + T^{-1}F_t'\Sigma_F^{-1}F_t\Sigma_{ui})^{1/2}((\hat{C}_i^{FCA})'\hat{F}_t^{FCA} - C_i'F_t) \to_d N(0,1)$$

as $N, T \to \infty$.

Proof of Proposition 4.

When analyzing the PC estimator it is convenient to redefine $d = \hat{F}^{FPC} - FH$ and $d_t = \hat{F}_t^{FPC} - H'F_t$, such that

$$H'F'M_{\hat{F}^{FPC}}FH = d'M_{\hat{F}^{FPC}}d = d'd - d'F(F'F)^{-1}F'd - d'(M_F - M_{\hat{F}^{FPC}})d.$$

Moreover, by using the results provided in the proof of Proposition 2, we have

$$\hat{F}_t^{FPC} - H'F_t = H'(\hat{F}_t^{IPC} - F_t) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right),$$

where the order terms are such that they do not affect the analysis of R_1 and R_2. Hence, we may use the same steps as in the proof of Theorem 1 to show that

$$
\begin{aligned}
R_1 &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \Lambda_i(H^-)'H'F'M_{\hat{F}^{FPC}}FHH^-\lambda_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \Lambda_i(H^-)'d'dH^-\lambda_i + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right),
\end{aligned}
$$

where, by proposition 4, $NT^{-1}d'd \to_p H'\bar{\Omega}_{IPC}H$ as $T \to \infty$, suggesting that

$$\frac{1}{T}\sum_{i=1}^N \Lambda_i(H^-)'d'dH^-\lambda_i \to_p B_{1IPC}$$

as $N, T \to \infty$. Hence, just as in the proof of Theorem 1,

$$R_1 - \sqrt{T}N^{-1/2}B_{1IPC} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{N}\right).$$

As for R_2, note that

$$
\begin{aligned}
M_F - M_{\hat{F}^{FPC}} &= d((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}d' + d((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}H'F' + FH((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}d' \\
&\quad + FH[((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1} - (H'F'FH)^{-1}]H'F',
\end{aligned}
$$

where the second term is also the leading term. Hence, by again using the same steps as before,

$$
\begin{aligned}
R_2 &= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'(M_F - M_{\hat{F}^{FPC}})F\lambda_i \\
&= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_i'd((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}H'F'F\lambda_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right),
\end{aligned}
$$

where, via Taylor expansion of the type $(T^{-1}(\hat{F}^{FPC})'\hat{F}^{FPC})^{-1} = (T^{-1}H'F'FH)^{-1} + o_p(1)$,

$$
\begin{aligned}
\frac{1}{T}\sum_{i=1}^{N}\varepsilon_i'd(T^{-1}(\hat{F}^{FPC})'\hat{F}^{FPC})^{-1}T^{-1}H'F'FHH^-\lambda_i &= \frac{1}{T}\sum_{i=1}^{N}\varepsilon_i'dH^-\lambda_i + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}d_t'H^-\lambda_i + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}u_{jt}'C_j'\bar{A}^{-1}HH^-\lambda_i + O_p\left(\frac{1}{\sqrt{N}}\right) \to_p B_{2IPC}
\end{aligned}
$$

as $N, T \to \infty$, suggesting

$$R_2 + \sqrt{T}N^{-1/2}B_{2IPC} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right).$$

Hence, so far the results are exactly the same as in Theorem 1.

Next, consider the normalized sums of $x_i'M_{\hat{F}^{FPC}}\epsilon_i$ and $x_i'M_{\hat{F}^{FPC}}x_i$. The analysis of the latter sum is unaffected by the fact that we have replaced $\hat{F}^{IPC}$ by $\hat{F}^{FPC}$. However, this is not the case for the first sum, which is now given by

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_{\hat{F}^{FPC}}\epsilon_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_F\epsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'(M_F - M_{\hat{F}^{FPC}})\epsilon_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'M_F\epsilon_i \\
&\quad - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}x_i'FH[((\hat{F}^{FPC})'\hat{F}^{FPC})^{-1} - (H'F'FH)^{-1}]H'F'\epsilon_i \\
&\quad + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right),
\end{aligned}
$$

where

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} T^{-1} x_i' F H T [((\hat{F}^{FPC})' \hat{F}^{FPC})^{-1} - (H'F'FH)^{-1}] T^{-1/2} H'F' \epsilon_i$$

$$\leq \quad \sqrt{N} \left(\frac{1}{N} \sum_{i=1}^{N} ||T^{-1} x_i' F||^2 \right)^{1/2} ||H||^2 T ||((\hat{F}^{FPC})' \hat{F}^{FPC})^{-1} - (H'F'FH)^{-1}||$$

$$\times \quad \left(\frac{1}{N} \sum_{i=1}^{N} ||T^{-1/2} F' \epsilon_i||^2 \right)^{1/2}.$$

By using the same argument as in the proof of Theorem 1,

$$T ||((\hat{F}^{FPC})' \hat{F}^{FPC})^{-1} - (H'F'FH)^{-1}|| = O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{N} \right) + O_p \left(\frac{1}{T} \right),$$

where the last order term is new. Thus, by direct substitution,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i' M_{\hat{F}^{FPC}} \epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} x_i' M_F \epsilon_i + O_p \left(\frac{1}{\sqrt{N}} \right) + O_p \left(\frac{\sqrt{T}}{N} \right) + O_p \left(\frac{\sqrt{N}}{T} \right),$$

which in turn implies

$$\sqrt{NT}(\hat{\beta}_{FPC} - \beta) = \sqrt{NT}(\hat{\beta}_{ILS} - \beta) + \bar{\Sigma}_\varepsilon^{-1} \sqrt{T} N^{-1/2} (B_{1IPC} - B_{2IPC}) + O_p \left(\frac{1}{\sqrt{N}} \right)$$

$$+ \quad O_p \left(\frac{1}{\sqrt{T}} \right) + O_p \left(\frac{\sqrt{T}}{N} \right) + O_p \left(\frac{\sqrt{N}}{T} \right).$$

This completes the proof for the PC estimator. The proof for the CA estimator is entirely analogous. ∎

Table 1: Simulation results when $N = T$.

T	a, σ_ϵ^2	Mean of bias						Standard deviation of bias						Size at the 5% level				
		IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	ICA	FCA	ILS
		$b = c = 0$																
25	1.0	−0.003	−4.554	0.000	−0.003	−0.003	0.000	0.238	1.507	0.000	0.238	0.238	0.000	6.8	93.5	6.8	6.8	6.3
50	1.0	−0.004	−2.591	0.000	−0.004	−0.004	0.000	0.155	1.438	0.000	0.155	0.155	0.000	5.9	62.0	5.9	5.9	5.6
100	1.0	0.000	−0.841	0.000	0.000	0.000	0.000	0.103	0.268	0.000	0.103	0.103	0.000	5.3	15.4	5.3	5.3	5.1
200	1.0	−0.001	−0.356	0.000	−0.001	−0.001	0.000	0.071	0.096	0.000	0.071	0.071	0.000	5.5	7.0	5.5	5.5	5.4
25	0.3	−0.003	−1.271	0.000	−0.003	−0.003	0.000	0.222	0.948	0.000	0.222	0.222	0.000	6.8	25.5	6.8	6.8	6.3
50	0.3	−0.003	−0.468	0.000	−0.003	−0.003	0.000	0.149	0.240	0.000	0.149	0.149	0.000	6.0	9.0	6.0	6.0	5.6
100	0.3	0.000	−0.205	0.000	0.000	0.000	0.000	0.101	0.115	0.000	0.101	0.101	0.000	5.3	5.7	5.3	5.3	5.1
200	0.3	−0.001	−0.097	0.000	−0.001	−0.001	0.000	0.070	0.072	0.000	0.070	0.070	0.000	5.4	5.6	5.4	5.4	5.4
25	3.0	−0.004	−3.437	0.000	−0.004	−0.004	0.000	0.270	0.733	0.000	0.270	0.270	0.000	6.9	87.5	6.9	6.9	6.3
50	3.0	−0.005	−4.754	0.000	−0.005	−0.005	0.000	0.168	0.793	0.000	0.168	0.168	0.000	5.7	99.1	5.7	5.7	5.6
100	3.0	0.001	−4.790	0.000	0.001	0.001	0.000	0.109	0.601	0.000	0.109	0.109	0.000	5.2	95.1	5.2	5.2	5.1
200	3.0	−0.001	−1.454	0.000	−0.001	−0.001	0.000	0.073	0.332	0.000	0.073	0.073	0.000	5.5	32.5	5.5	5.5	5.4
		$b = 0, c = 4$																
25	1.0	−1.694	−1.830	−1.604	0.004	0.004	0.000	0.453	0.497	0.000	0.423	0.423	0.000	39.5	44.4	7.7	7.7	6.3
50	1.0	−1.668	−1.734	−1.566	−0.011	−0.011	0.000	0.316	0.330	0.000	0.290	0.290	0.000	39.8	42.0	5.7	5.7	5.6
100	1.0	−1.664	−1.705	−1.537	0.005	0.005	0.000	0.226	0.231	0.000	0.188	0.188	0.000	40.4	41.6	4.7	4.7	5.1
200	1.0	−1.667	−1.687	−1.567	0.002	0.002	0.000	0.158	0.160	0.000	0.133	0.133	0.000	39.4	40.1	5.2	5.2	5.4
25	0.3	−1.706	−1.747	−1.604	0.005	0.005	0.000	0.457	0.470	0.000	0.425	0.425	0.000	39.7	41.1	7.5	7.5	6.3
50	0.3	−1.674	−1.694	−1.566	−0.011	−0.011	0.000	0.317	0.321	0.000	0.292	0.292	0.000	40.1	40.4	5.8	5.8	5.6
100	0.3	−1.668	−1.680	−1.537	0.005	0.005	0.000	0.226	0.228	0.000	0.188	0.188	0.000	40.5	40.8	4.7	4.7	5.1
200	0.3	−1.669	−1.675	−1.567	0.002	0.002	0.000	0.158	0.158	0.000	0.133	0.133	0.000	39.6	39.7	5.1	5.1	5.4
25	3.0	−1.664	−2.076	−1.604	0.002	0.002	0.000	0.447	0.592	0.000	0.418	0.418	0.000	38.2	52.6	7.8	7.8	6.3
50	3.0	−1.651	−1.847	−1.566	−0.012	−0.012	0.000	0.314	0.358	0.000	0.287	0.287	0.000	39.2	46.2	5.8	5.8	5.6
100	3.0	−1.653	−1.775	−1.537	0.005	0.005	0.000	0.225	0.243	0.000	0.188	0.188	0.000	39.8	43.8	4.7	4.7	5.1
200	3.0	−1.662	−1.721	−1.567	0.002	0.002	0.000	0.157	0.164	0.000	0.133	0.133	0.000	39.3	41.0	5.2	5.2	5.4

Continued overleaf

Table 1: Continued.

		Mean of bias						Standard deviation of bias						Size at the 5% level				
T	a, σ_ϵ^2	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	ICA	FCA	ILS
		$b = 1, c = 4$																
25	1.0	-1.664	-1.828	-1.607	0.041	0.041	0.037	0.423	0.483	0.000	0.403	0.403	0.000	49.8	56.0	7.9	7.9	6.4
50	1.0	-1.657	-1.739	-1.597	-0.034	-0.034	-0.023	0.295	0.315	0.000	0.280	0.280	0.000	52.5	55.9	5.8	5.8	5.6
100	1.0	-1.606	-1.657	-1.522	0.011	0.011	0.006	0.207	0.215	0.000	0.178	0.178	0.000	50.7	52.8	4.7	4.7	5.0
200	1.0	-1.583	-1.606	-1.517	0.004	0.004	0.002	0.143	0.146	0.000	0.125	0.125	0.000	49.5	50.5	5.2	5.2	5.2
25	0.3	-1.653	-1.727	-1.610	0.080	0.080	0.073	0.395	0.429	0.000	0.388	0.388	0.000	68.0	70.8	7.6	7.6	6.6
50	0.3	-1.652	-1.711	-1.638	-0.063	-0.063	-0.052	0.266	0.285	0.000	0.269	0.269	0.000	89.0	90.4	6.2	6.2	5.5
100	0.3	-1.562	-1.586	-1.510	0.016	0.016	0.010	0.189	0.194	0.000	0.169	0.169	0.000	71.6	73.0	4.8	4.8	4.8
200	0.3	-1.512	-1.522	-1.473	0.006	0.006	0.004	0.129	0.130	0.000	0.118	0.118	0.000	69.2	69.9	5.2	5.2	5.1
25	3.0	-1.649	-2.086	-1.605	0.016	0.016	0.016	0.434	0.595	0.000	0.408	0.408	0.000	41.5	57.3	7.7	7.7	6.2
50	3.0	-1.644	-1.853	-1.580	-0.021	-0.021	-0.010	0.304	0.354	0.000	0.282	0.282	0.000	43.9	51.7	5.6	5.6	5.6
100	3.0	-1.627	-1.758	-1.530	0.008	0.008	0.002	0.217	0.237	0.000	0.183	0.183	0.000	43.7	48.5	4.7	4.7	4.9
200	3.0	-1.625	-1.688	-1.546	0.003	0.003	0.001	0.151	0.158	0.000	0.129	0.129	0.000	42.6	45.5	5.1	5.1	5.2

Notes: The data are generated from equation (4) with $m = 1$, $r = 2$, $\beta = 1$, $\Lambda_i = \Lambda = (0, 1)$, $\lambda_i \sim (1, 1)' + cU(0, 1)$ and $(F_t', \epsilon_{it}, \varepsilon_{it})' \sim N(0, \mathrm{diag}(I_2, \sigma_\epsilon^2, \Sigma_{\varepsilon i}))$, where $\Sigma_{\varepsilon i} \sim a + bU(0, 1)$. The mean and standard deviation are computed for $\sqrt{NT}(\hat\beta_{ICA} - \hat\beta_{ILS})$, $\sqrt{NT}(\hat\beta_{IPC} - \hat\beta_{ILS})$, $\sqrt{NT}(\hat\beta_{FCA} - \hat\beta_{ILS})$ and $\sqrt{NT}(\hat\beta_{FPC} - \hat\beta_{ILS})$. The size results are for a double-sided t-test of the null hypothesis of $\beta = 1$. "Theory" refers to the asymptotic results reported in Theorem 1 and Proposition 5.

35

Table 2: Simulation results when $N = T^{4/3}$.

T	a, σ^2_ϵ	Mean of bias						Standard deviation of bias						Size at the 5% level				
		IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	ICA	FCA	ILS
		$b = c = 0$																
25	1.0	0.000	−2.491	0.000	0.000	0.000	0.000	0.142	1.431	0.000	0.142	0.142	0.000	6.6	58.0	6.6	6.6	6.5
50	1.0	−0.001	−0.786	0.000	−0.001	−0.001	0.000	0.079	0.290	0.000	0.079	0.079	0.000	5.5	13.4	5.5	5.5	5.1
100	1.0	0.001	−0.305	0.000	0.001	0.001	0.000	0.050	0.087	0.000	0.050	0.050	0.000	5.3	5.9	5.3	5.3	5.1
200	1.0	−0.001	−0.132	0.000	−0.001	−0.001	0.000	0.030	0.036	0.000	0.030	0.030	0.000	5.7	5.5	5.7	5.7	5.6
25	0.3	−0.001	−0.550	0.000	−0.001	−0.001	0.000	0.131	0.306	0.000	0.131	0.131	0.000	6.7	9.9	6.7	6.7	6.5
50	0.3	0.000	−0.211	0.000	0.000	0.000	0.000	0.076	0.104	0.000	0.076	0.076	0.000	5.6	6.4	5.6	5.6	5.1
100	0.3	0.001	−0.087	0.000	0.001	0.001	0.000	0.049	0.053	0.000	0.049	0.049	0.000	5.3	5.4	5.3	5.3	5.1
200	0.3	−0.001	−0.039	0.000	−0.001	−0.001	0.000	0.030	0.030	0.000	0.030	0.030	0.000	5.6	5.4	5.6	5.6	5.6
25	3.0	0.000	−3.698	0.000	0.000	0.000	0.000	0.167	0.920	0.000	0.167	0.167	0.000	6.6	90.1	6.6	6.6	6.5
50	3.0	−0.001	−3.304	0.000	−0.001	−0.001	0.000	0.088	1.364	0.000	0.088	0.088	0.000	5.4	80.0	5.4	5.4	5.1
100	3.0	0.001	−1.044	0.000	0.001	0.001	0.000	0.052	0.265	0.000	0.052	0.052	0.000	5.3	19.2	5.3	5.3	5.1
200	3.0	−0.001	−0.413	0.000	−0.001	−0.001	0.000	0.031	0.071	0.000	0.031	0.031	0.000	5.6	7.7	5.6	5.6	5.6
		$b = 0, c = 4$																
25	1.0	−0.977	−1.046	−0.913	−0.001	−0.001	0.000	0.274	0.295	0.000	0.239	0.239	0.000	16.7	18.1	6.7	6.7	6.5
50	1.0	−0.881	−0.910	−0.823	−0.001	−0.001	0.000	0.169	0.176	0.000	0.150	0.150	0.000	14.9	15.4	5.5	5.5	5.1
100	1.0	−0.791	−0.800	−0.738	0.001	0.001	0.000	0.106	0.107	0.000	0.097	0.097	0.000	12.8	12.9	5.2	5.2	5.1
200	1.0	−0.698	−0.703	−0.649	−0.001	−0.001	0.000	0.067	0.068	0.000	0.059	0.059	0.000	11.3	11.4	5.3	5.3	5.6
25	0.3	−0.980	−1.001	−0.913	−0.001	−0.001	0.000	0.273	0.279	0.000	0.237	0.237	0.000	16.7	17.2	6.6	6.6	6.5
50	0.3	−0.882	−0.890	−0.823	−0.001	−0.001	0.000	0.169	0.171	0.000	0.150	0.150	0.000	14.8	14.9	5.5	5.5	5.1
100	0.3	−0.791	−0.794	−0.738	0.001	0.001	0.000	0.106	0.106	0.000	0.096	0.096	0.000	12.8	12.8	5.2	5.2	5.1
200	0.3	−0.698	−0.700	−0.649	−0.001	−0.001	0.000	0.067	0.067	0.000	0.058	0.058	0.000	11.3	11.3	5.4	5.4	5.6
25	3.0	−0.969	−1.177	−0.913	0.000	0.000	0.000	0.276	0.345	0.000	0.245	0.245	0.000	16.5	21.0	6.7	6.7	6.5
50	3.0	−0.877	−0.965	−0.823	−0.001	−0.001	0.000	0.170	0.190	0.000	0.152	0.152	0.000	14.6	16.6	5.5	5.5	5.1
100	3.0	−0.790	−0.819	−0.738	0.001	0.001	0.000	0.106	0.110	0.000	0.097	0.097	0.000	12.8	13.4	5.3	5.3	5.1
200	3.0	−0.698	−0.713	−0.649	−0.001	−0.001	0.000	0.067	0.069	0.000	0.059	0.059	0.000	11.3	11.7	5.4	5.4	5.6

Continued overleaf

Table 2: Continued.

		Mean of bias						Standard deviation of bias						Size at the 5% level				
T	a, σ_ϵ^2	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	Theory	ICA	FCA	Theory	IPC	FPC	ICA	FCA	ILS
								$b = 1, c = 4$										
25	1.0	−0.944	−1.024	−0.904	0.004	0.004	0.004	0.251	0.279	0.000	0.227	0.227	0.000	20.9	23.3	6.3	6.3	6.2
50	1.0	−0.844	−0.878	−0.806	0.020	0.020	0.021	0.154	0.163	0.000	0.143	0.143	0.000	18.1	19.3	5.6	5.6	5.2
100	1.0	−0.763	−0.774	−0.728	−0.014	−0.014	−0.015	0.096	0.097	0.000	0.092	0.092	0.000	16.1	16.4	5.3	5.3	5.1
200	1.0	−0.671	−0.677	−0.638	−0.006	−0.006	−0.006	0.061	0.062	0.000	0.056	0.056	0.000	13.9	14.0	5.3	5.3	5.7
25	0.3	−0.919	−0.952	−0.895	0.007	0.007	0.008	0.229	0.243	0.000	0.214	0.214	0.000	30.6	32.5	6.9	6.9	5.9
50	0.3	−0.797	−0.818	−0.783	0.048	0.048	0.049	0.132	0.139	0.000	0.131	0.131	0.000	37.5	38.5	5.7	5.7	5.0
100	0.3	−0.739	−0.744	−0.720	−0.027	−0.027	−0.028	0.086	0.087	0.000	0.088	0.088	0.000	23.2	23.4	5.3	5.3	5.2
200	0.3	−0.648	−0.650	−0.629	−0.011	−0.011	−0.011	0.055	0.056	0.000	0.053	0.053	0.000	19.4	19.5	5.5	5.5	5.4
25	3.0	−0.954	−1.173	−0.909	0.002	0.002	0.002	0.267	0.342	0.000	0.239	0.239	0.000	17.7	23.4	6.5	6.5	6.4
50	3.0	−0.861	−0.954	−0.816	0.008	0.008	0.009	0.164	0.185	0.000	0.149	0.149	0.000	16.0	18.1	5.5	5.5	5.3
100	3.0	−0.777	−0.808	−0.734	−0.005	−0.005	−0.007	0.102	0.106	0.000	0.095	0.095	0.000	14.2	14.6	5.4	5.4	5.0
200	3.0	−0.686	−0.702	−0.644	−0.003	−0.003	−0.002	0.065	0.066	0.000	0.058	0.058	0.000	12.3	12.7	5.4	5.4	5.5

Notes: See Table 1.