On the Role of the Rank Condition in CCE Estimation of Factor-Augmented Panel Regressions *

Hande Karabiyik
Lund University

Simon Reese
Lund University

Joakim Westerlund†
Lund University
and
Centre for Economics and Financial Econometrics Research
Deakin University

February 9, 2016

Abstract

A popular approach to factor-augmented panel regressions is the common correlated effects (CCE) estimator of Pesaran (Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**, 967–1012, 2006). In fact, the approach is so popular that it has given rise to a separate CCE literature with several key publications appearing in *Journal of Econometrics*. This paper points to a problem with the CCE approach that seems to have gone largely unnoticed in the literature. The problem occurs in the empirically relevant case when the number of factors is strictly less than the number of observables used in their estimation. Specifically, the use of too many observables causes the second moment matrix of the estimated factors to become asymptotically singular, which invalidates some of the arguments commonly used to obtain asymptotic results. A new argument is therefore proposed that is shown to facilitate a straightforward asymptotic analysis.

JEL Classification: C12; C13; C33; C36.

Keywords: Factor-augmented panel regression; CCE estimation; Moore–Penrose inverse.

*The authors would like to thank Joerg Breitung and Mehdi Hosseinkouchack for many valuable comments and suggestions. Thank you also to the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship, and the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.

†Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

1 Introduction

Consider the scalar and $k \times 1$ vector of observable panel data variables $y_{i,t}$ and $\mathbf{x}_{i,t}$, where $i = 1, ..., N$ and $t = 1, ..., T$ index the cross-sectional and time series dimensions respectively. Except for some simplifications that are irrelevant for the purpose of the paper, such as the absence of all deterministic terms, the model is the same as in Pesaran (2006), and is given by

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i, \quad (1)$$

$$\mathbf{e}_i = \mathbf{F} \gamma_i + \varepsilon_i, \quad (2)$$

$$\mathbf{X}_i = \mathbf{F} \boldsymbol{\Gamma}_i' + \mathbf{V}_i, \quad (3)$$

where $\mathbf{X}_i = (\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T})'$ is $T \times k$, $\boldsymbol{\beta}$ is a $k \times 1$, $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ is a $T \times m$ matrix of unobservable common factors with γ_i and $\boldsymbol{\Gamma}_i$ being the associated vectors of factor loadings, and $\varepsilon_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$ and $\mathbf{V}_i = (\mathbf{v}_{i,1}, ..., \mathbf{v}_{i,T})'$ are $T \times 1$ and $T \times k$ matrices, respectively, of idiosyncratic errors.[1] Except for the requirement that $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ are serially uncorrelated and homoskedastic, Assumption 1 is the same as Assumptions 1–4 in Pesaran (2006).

Assumption 1.

(i) $\varepsilon_{i,t}$ is independently and identically distributed (iid) across both i and t with $E(\varepsilon_{i,t}) = 0$, $E(\varepsilon_{i,t}^2) = \sigma^2$ and $E(\varepsilon_{i,t}^4) < \infty$;

(ii) $\mathbf{v}_{i,t}$ is iid across both i and t with $E(\mathbf{v}_{i,t}) = \mathbf{0}_{k \times 1}$, $E(\mathbf{v}_{i,t}\mathbf{v}_{i,t}') = \boldsymbol{\Sigma}$ positive definite and $E(\|\mathbf{v}_{i,t}\|^4) < \infty$, where $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})}$ is the Frobenius norm of the matrix $\mathbf{A}$;

(iii) $\mathbf{f}_t$ is covariance stationary such that $E(\|\mathbf{f}_t\|^4) < \infty$ and $E(\mathbf{f}_t\mathbf{f}_t') = \boldsymbol{\Sigma}_\mathbf{f}$ is positive definite;

(iv) γ_i and $\boldsymbol{\Gamma}_i$ are iid across i, independent of $\varepsilon_{j,t}$, $\mathbf{v}_{j,t}$ and $\mathbf{f}_t$ for all i and j, have fixed means γ and $\boldsymbol{\Gamma}$, respectively, and finite variances;

(v) $\varepsilon_{i,t}$, $\mathbf{v}_{i,s}$ and $\mathbf{f}_\ell$ are mutually independent for all i, j, t, s and ℓ.

Because of the way that $\mathbf{F}$ enters both (2) and (3) the estimation of $\boldsymbol{\beta}$ is nontrivial. However, by combining (1) and (3), we have

$$\mathbf{Z}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \quad (4)$$

[1] Pesaran (2006) considers a random coefficient model in which $\boldsymbol{\beta}$ is allowed to vary across the cross-section. This model is therefore more general than the one considered here. The homogenous slope assumption is, however, just for simplicity and is by no means a restriction. Indeed, both the critique raised and the solution offered in the current paper apply even under heterogeneous slopes.

where $\mathbf{Z}_i = (\mathbf{y}_i, \mathbf{X}_i) = (\mathbf{z}_{i,1}, ..., \mathbf{z}_{i,T})'$ is $T \times (k+1)$, $\mathbf{z}_{i,t} = (y_{i,t}, \mathbf{x}'_{i,t})'$ is $(k+1) \times 1$, $\mathbf{C}_i = (\mathbf{\Gamma}'_i \boldsymbol{\beta} + \gamma_i, \mathbf{\Gamma}'_i)$ is $m \times (k+1)$, and $\mathbf{U}_i = (\mathbf{u}_{i,1}, ..., \mathbf{u}_{i,T})' = (\mathbf{V}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \mathbf{V}_i)$ is $T \times (k+1)$. Thus, (1)–(3) can be rewritten equivalently as a static factor model for $\mathbf{Z}_i$, which is convenient because it means that $\mathbf{F}$ can be estimated using existing approaches for such models. In the common correlated effects (CCE) approach of Pesaran (2006), the estimator of $\mathbf{F}$ is particularly simple, and is given by

$$\widehat{\mathbf{F}} = \overline{\mathbf{Z}} = \mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}}, \tag{5}$$

where $\overline{\mathbf{A}} = N^{-1} \sum_{i=1}^N \mathbf{A}_i$ for any $\mathbf{A}_i$. It is important to note here that under Assumption 1, $\overline{\mathbf{U}} \stackrel{p}{\to} \mathbf{0}_{T \times (k+1)}$ as $N \to \infty$, where $\stackrel{p}{\to}$ signifies convergence in probability. This means that $\widehat{\mathbf{F}}$ is consistent for the space spanned by $\mathbf{F}$. The pooled CCE estimator of $\boldsymbol{\beta}$ is the conventional pooled OLS estimator with $\widehat{\mathbf{F}}$ in place of $\mathbf{F}$;

$$\widehat{\mathbf{b}}_P = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_i, \tag{6}$$

where $\mathbf{M_A} = \mathbf{I}_T - \mathbf{P_A} = \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^+ \mathbf{A}'$ for any T-rowed matrix $\mathbf{A}$ with $(\mathbf{A}'\mathbf{A})^+$ being the Moore–Penrose (MP) inverse of $\mathbf{A}'\mathbf{A}$.[2] Because of its simplicity and generality, the CCE approach has attracted considerable attention, so much so that there is by now a separate CCE branch of the literature (see Chudik and Pesaran, 2015b, for a recent survey). This literature makes extensive use of the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$, which has been shown to be normal under a wide variety of circumstances that include lagged dependent variables, unit roots, and weakly influential factors (see, for example, Chudik et al., 2011; Chudik and Pesaran, 2015a; Kapetanios et al., 2010; Pesaran et al., 2013; Pesaran, 2007; Reese and Westerlund (2015a, 2015b); Westerlund, 2015; Westerlund and Urbain, 2015).

The current paper is about the way in which the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ has been established. A critical first step in all existing proofs is to show that the effect of the estimation of $\mathbf{F}$ is negligible, such that $\mathbf{F}$ can be treated as known in the rest of the proof. This is done by showing that $(NT)^{-1} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}}) \mathbf{X}_i$, $(NT)^{-1/2} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}}) \boldsymbol{\varepsilon}_i$ and

[2] The four defining equations of the MP inverse $\mathbf{A}^+$ of $\mathbf{A}$ are; (i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, (ii) $(\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+$, (iii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$, and (iv) $(\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}$ (see, for example, Abadir and Magnus, 2005, Exercise 10.24).

$(NT)^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}-\mathbf{M}_{\widehat{\mathbf{F}}})\mathbf{F}\gamma_i$ are negligible, such that

$$\begin{aligned}\sqrt{NT}(\widehat{\mathbf{b}}_P-\boldsymbol{\beta}) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}(\mathbf{F}\gamma_i+\boldsymbol{\varepsilon}_i) \\ &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}\boldsymbol{\varepsilon}_i+o_p(1). \end{aligned} \tag{7}$$

Let us therefore consider $\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}-\mathbf{M}_{\widehat{\mathbf{F}}}$, which can be expanded in the following way:

$$\begin{aligned}\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}-\mathbf{M}_{\widehat{\mathbf{F}}} &= \overline{\mathbf{U}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}\overline{\mathbf{U}}'+\overline{\mathbf{U}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}\overline{\mathbf{C}}'\mathbf{F}'+\mathbf{F}\overline{\mathbf{C}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}\overline{\mathbf{U}}' \\ &+ \mathbf{F}\overline{\mathbf{C}}[(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}-(\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}]\overline{\mathbf{C}}'\mathbf{F}'. \end{aligned} \tag{8}$$

Pesaran (2006, equation (21)) requires that

$$\text{rank}(\overline{\mathbf{C}}) = m \le k+1. \tag{9}$$

The fact that under this condition, $\overline{\mathbf{C}}$ is not necessarily a full rank square matrix explains the need for the MP inverse in $(\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}$. Pesaran (2006) shows that $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}-T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}\xrightarrow{p}\mathbf{0}_{(k+1)\times(k+1)}$ as N, $T\to\infty$, which is taken to imply that also the inverses are well-behaved. This same reasoning has been used in numerous other studies, such as those of Chudik et al. (2011), Chudik and Pesaran (2015a), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). However, as pointed out by Andrews (1987), this is not always true. Suppose therefore that $\mathbf{A}_n-\mathbf{A}_0\xrightarrow{p}\mathbf{0}_{r\times r}$ as $n\to\infty$ for any real $r\times r$ matrix $\mathbf{A}_0$ and any real $r\times r$ matrix sequence $\mathbf{A}_n$. Let $r_n=\text{rank}(\mathbf{A}_n)$ and $r_0=\text{rank}(\mathbf{A}_0)$. Then, according to Andrews (1987), $\mathbf{A}_n^{+}-\mathbf{A}_0^{+}\xrightarrow{p}\mathbf{0}_{r\times r}$ if and only if

$$r_n-r_0\xrightarrow{a.s.}0 \tag{10}$$

as $n\to\infty$, where $\xrightarrow{a.s.}$ signifies almost sure convergence. Therefore, only if $\text{rank}(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})-\text{rank}(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})\xrightarrow{a.s.}\mathbf{0}_{(k+1)\times(k+1)}$ as $N,T\to\infty$, can we take $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}-T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}\xrightarrow{p}\mathbf{0}_{(k+1)\times(k+1)}$ to imply that $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}-(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}\xrightarrow{p}\mathbf{0}_{(k+1)\times(k+1)}$. While this fact has not gone unnoticed in the CCE literature, it has been thought that the rank of $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$ should converge to the rank of $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$, as the former converges the latter, and that this should in turn ensure that $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}-(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}\xrightarrow{p}\mathbf{0}_{(k+1)\times(k+1)}$ regardless of whether (9) is satisfied with equality or strict inequality (see Chudik et al., 2011; Kapetanios et al., 2010; Pesaran et al., 2013). However, the rank is a discrete function and the probability that a rank change will occur before $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$ reaches its asymptotic limit in case (9) is satisfied with strict inequality is zero. That is, $P[\text{rank}(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) = \text{rank}(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})] \to 0$. The only way to ensure that $\text{rank}(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) - \text{rank}(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}) \overset{a.s.}{\to} 0_{(k+1)\times(k+1)}$ is therefore to assume that (9) is satisfied with equality.

The purpose of the present paper is in part to make the above discussion of the consequences of improper use of the MP inverse a little more precise, in part to propose a solution. In so doing, we are going to take footing in the mathematical literature on perturbed matrices. This is natural, because $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$ may be regarded as negligible perturbation of $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$. The results show that unless (9) is satisfied with equality, $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+}$ does not converge to $(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}$, but that $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{+} - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{+}$ actually diverges at a rate that is inversely related to the rate at which $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}} \overset{p}{\to} \mathbf{0}_{(k+1)\times(k+1)}$. The approach described in the last paragraph is therefore not suitable for evaluating $\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}}$, and hence also not suitable for studying the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$. An alternative approach is therefore needed. The approach taken here is very simple and builds on the literature on time series regression with regressors in different order of integration. The intuition is that when $\text{rank}(\overline{\mathbf{C}}) = m < k+1$, there is a subset of regressors in $\widehat{\mathbf{F}}$ that does not load on $\mathbf{F}$. These regressors are negligible, for $\overline{\mathbf{U}} \overset{p}{\to} \mathbf{0}_{(k+1)\times 1}$ as $N \to \infty$.

2 Some matrix perturbation theory

Consider again the $r \times r$ matrix $\mathbf{A}_0$ and the $r \times r$ matrix sequence $\mathbf{A}_n$. The following assumption is made:

$$\mathbf{A}_n = \mathbf{A}_0 + \mathbf{E}_n,$$

where $\mathbf{E}_n$ is a perturbation of $\mathbf{A}_0$. Most of the results in the existing matrix perturbation literature are derived under the assumption that both $\mathbf{E}_n$ and $\mathbf{A}_0$ are non-random matrices, and that n is fixed. It is further assumed that $\|\mathbf{A}_0^{+}\mathbf{E}_n\| < 1$, or even $\|\mathbf{A}_0^{+}\|\|\mathbf{E}_n\| < 1$ (see, for example, Stewart, 1969; 1977; Wedin, 1973). Stewart (1990) approaches the classical matrix perturbation theory from a probabilistic point of view by assuming random perturbations. In this section, we extend his results to the current more general context. The exact conditions that we will be working under are given in Assumption 2.

Assumption 2.

(i) $E(\|\mathbf{A}_0\|) < \infty$;

(ii) $\|\mathbf{E}_n\| \stackrel{p}{\to} 0$ as $n \to \infty$.

Replacing $\mathbf{A}_n$, $\mathbf{A}_0$ and $\mathbf{E}_n$ with $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$, $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$ and $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$, respectively, it is clear that Assumption 2 provides a natural starting point for our analysis. The questions are: What are the conditions under which $\mathbf{A}_n^+ - \mathbf{A}_0^+ \stackrel{p}{\to} \mathbf{0}_{r\times r}$, and what are the consequences if those conditions are not met? Before we come to the answer, however, it is useful to consider two simple examples that illustrate the nature of the problem. Suppose first that

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{E}_n = \begin{bmatrix} 0 & 0 \\ 0 & n^{-1} \end{bmatrix}.$$

Clearly,

$$\mathbf{A}_n - \mathbf{A}_0 = \mathbf{E}_n = \mathrm{diag}(0, n^{-1}) \to \mathbf{0}_{2\times 2}.$$

In spite of this, however, we have

$$\mathbf{A}_n^+ - \mathbf{A}_0^+ = \mathrm{diag}(0, n) \nrightarrow \mathbf{0}_{2\times 2}.$$

If, on the other hand,

$$\mathbf{E}_n = \begin{bmatrix} 0 & n^{-1} \\ 0 & 0 \end{bmatrix},$$

such that again $\mathbf{A}_n \to \mathbf{A}_0$, then

$$\mathbf{A}_n^+ - \mathbf{A}_0^+ = \frac{1}{1+n^{-2}} \begin{bmatrix} n^{-2} & 0 \\ n^{-1} & 0 \end{bmatrix} \to \mathbf{0}_{2\times 2}.$$

The reason for this difference in the results is that while in the second example $r_n = r_0$, in the first example $r_n > r_0$.

The above examples illustrate that problems can arise when working with the MP inverse and the rank of $\mathbf{A}_n$ is different from that of $\mathbf{A}_0$. With this in mind, we now continue to consider the results under Assumption 2. We begin by noting that since $\|\mathbf{A}_n - \mathbf{A}_0\| = \|\mathbf{E}_n\| = o_p(1)$, we have that $P(r_n \geq r_0) \to 1$ as $n \to \infty$ (see, for example, Andrews, 1987, Note 1).

Theorem 1. *Suppose that Assumption 2 is met, and that $r_n - r_0 \stackrel{a.s.}{\to} 0$ as $n \to \infty$. Then,*

$$\|\mathbf{A}_n^+ - \mathbf{A}_0^+\| = O_p(\|\mathbf{E}_n\|) = o_p(1).$$

Theorem 1 is similar to Theorem 2 of Andrews (1987), who studies the asymptotic behaviour of generalized Wald test statistics. Unfortunately, the results of Andrews (1987) cannot be used to study the consequences of a violation of $r_n - r_0 \stackrel{a.s.}{\to} 0$. Theorem 2 provides the missing piece.

Theorem 2. *Suppose that Assumption 2 is met, and that $P(r_n > r_0) \to 1$ as $n \to \infty$. Then,*

$$\|\mathbf{A}_n^+ - \mathbf{A}_0^+\| = O_p(\|\mathbf{E}_n^{-1}\|).$$

Theorem 2 implies that if $\mathbf{A}_n$ is "near" $\mathbf{A}_0$, but $r_n > r_0$, then its MP inverse can be larger and completely different from $\mathbf{A}_0^+$, and the smaller is $\mathbf{E}_n$, the worse the problem can be. The implication of this result is obvious; if we want $\mathbf{A}_n^+$ to be well-behaved in the sense that $\|\mathbf{A}_n^+ - \mathbf{A}_0^+\| = o_p(1)$, a necessary and sufficient condition is given by (10), that is, $r_n - r_0 \overset{a.s.}{\to} 0$. Unfortunately, this condition is not easily verified. Chudik et al. (2011), Kapetanios et al. (2010), and Pesaran et al. (2013) all recognize the importance of (10). They claim that if $\mathbf{E}_n \overset{p}{\to} \mathbf{0}_{r\times r}$ and $\text{rank}(\mathbf{A}_0) = r_0$, then $r_n - r_0 \overset{a.s.}{\to} 0$, which is not correct, since $\mathbf{E}_n \overset{p}{\to} \mathbf{0}_{r\times r}$ only implies $P(r_n \geq r_0) \to 1$, and not $P(r_n = r_0) \to 1$ (see Andrews, 1987, Note 1). To see this, consider the following counter-example. Let

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{E}_n = \begin{bmatrix} 0 & 0 \\ 0 & E_n \end{bmatrix},$$

where E_n is a continuous random variable such that $E_n \overset{p}{\to} 0$. Clearly,

$$\mathbf{A}_n - \mathbf{A}_0 = \mathbf{E}_n = \text{diag}(0, E_n) \overset{p}{\to} \mathbf{0}_{2\times 2}.$$

Furthermore, $r_n > r_0$ for all $n < \infty$. In fact, since E_n is a continuous random variable, we have that $P(r_n = 1) = P(E_n = 0) = 0$. This shows that $\mathbf{A}_n - \mathbf{A}_0 \overset{p}{\to} \mathbf{0}_{r\times r}$ does not imply (10). It also illustrates how the randomness of the perturbation plays an important role. A key result in this regard is that if $\mathbf{A}_n$ is a continuous random matrix that is not artificially constructed to have perfectly linearly correlated rows and/or columns, then

$$r_n - r \overset{a.s.}{\to} 0 \tag{11}$$

as $n \to \infty$ (see, for example, Feng and Zhang, 2007).[3] This means that for (10) to hold, we require

$$r_n = r_0 \tag{12}$$

for all n.

[3]Note that it is possible to obtain a reduced rank random matrix by taking the product of two lower rank matrices. For example, suppose that $\mathbf{A}_n = \mathbf{B}_n\mathbf{C}_n$, where $\mathbf{B}_n$ and $\mathbf{C}_n$ are $r \times k$ and $k \times r$, respectively. If $k < r$, then $r_n = k$ and therefore $r_n < r$.

An important consequence of Theorem 1 is that $\|\mathbf{A}_n^+\| = O_p(\|\mathbf{A}_0^+\|) = O_p(1)$ under (10). If, on the other hand, (10) is not met, then, by Theorem 2,

$$\|\mathbf{A}_n^+\| = O_p(\|\mathbf{E}_n^{-1}\|). \tag{13}$$

3 Main results

We now apply the theory developed in Section 2 to the problem described in Section 1. We begin by verifying that Assumption 2 holds with $\mathbf{A}_n$, $\mathbf{A}_0$ and $\mathbf{E}_n$ replaced by $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$, $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$ and $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})$, respectively. Clearly, $E(\|T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}\|) < \infty$. Moreover, according to (36) in Pesaran (2006), we have

$$\|T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}\| = O_p(N^{-1}) + O_p((NT)^{-1/2}),$$

showing that Assumption 2 is in fact applicable.

We now find the ranks of $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$ and $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$. We begin by noting that, in analogy to (11), we have

$$\text{rank}(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) \overset{a.s.}{\to} k+1 \tag{14}$$

as N, $T \to \infty$. Moreover, by Assumption 1 (iii), $T^{-1}\mathbf{F}'\mathbf{F}$ has full rank for all T, including $T \to \infty$. The rank of $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$ is therefore determined by the rank of $\overline{\mathbf{C}}$. Hence, in view of (9),

$$\text{rank}(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}) = m, \tag{15}$$

for all N and T, including T, $N \to \infty$.

Suppose now that (9) is satisfied with equality, such that $m = k+1$. Since in this case $\text{rank}(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}) \overset{a.s.}{\to} \text{rank}(T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})$, by Theorem 1,

$$\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\| = O_p(N^{-1}) + O_p((NT)^{-1/2}), \tag{16}$$

$$\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\| = O_p(1). \tag{17}$$

If, however, (9) is satisfied with strict inequality, such that $m < k+1$, then the rank of $T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}$ will not converge to that of $T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}$, and so, by Theorem 2,

$$\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\| = O_p(N) + O_p(\sqrt{NT}), \tag{18}$$

with $\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|$ being of the same order. This is illustrated in Figure 1, which plots the average norm of $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+$ and $(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+$ over 1,000 replications of a particular data generating process that sets $m = 2 < k + 1 = 3$ and $N = T$.[4] The two lines almost coincide and the divergence rate is $N = T$, just as expected. As an illustration of the implications of this these results, let us consider $T^{-1}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}})\mathbf{X}_i$. From Lemma 3 of Pesaran (2006), we have $\|T^{-1}\mathbf{X}_i'\overline{\mathbf{U}}\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$. In view of this, $\|\overline{\mathbf{C}}\| = O_p(1)$, $\|T^{-1}\mathbf{F}'\mathbf{X}_i\| = O_p(1)$ and (8), we have

$$
\begin{aligned}
&\|T^{-1}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}})\mathbf{X}_i\| \\
&\quad\leq \|T^{-1}\mathbf{X}_i'\overline{\mathbf{U}}\|\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|\|T^{-1}\overline{\mathbf{U}}'\mathbf{X}_i\| + 2\|T^{-1}\mathbf{X}_i'\overline{\mathbf{U}}\|\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|\|\overline{\mathbf{C}}'\|\|T^{-1}\mathbf{F}'\mathbf{X}_i\| \\
&\quad+ \|T^{-1}\mathbf{X}_i'\mathbf{F}\overline{\|}\|\mathbf{C}\|^2\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\|\|T^{-1}\mathbf{F}'\mathbf{X}_i\| \\
&\quad= [O_p(N^{-1}) + O_p((NT)^{-1/2})]O_p(\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|) \\
&\quad+ O_p(\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\|), \qquad (19)
\end{aligned}
$$

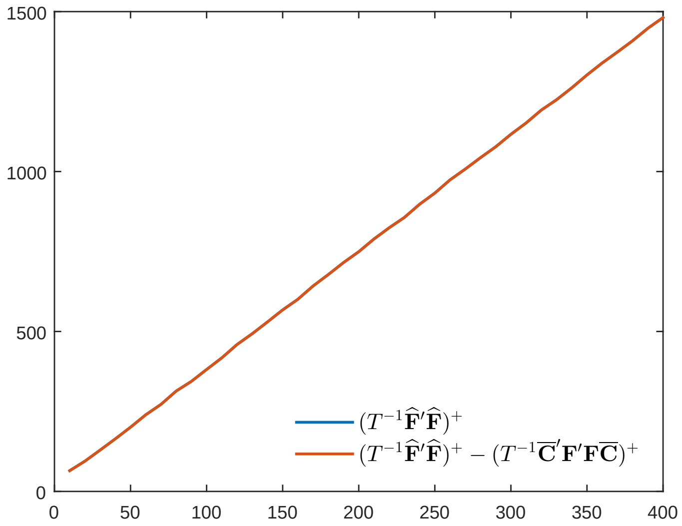
which is $O_p(N^{-1}) + O_p((NT)^{-1/2})$ if $m = k + 1$ and $O_p(N) + O_p(\sqrt{NT})$ if $m < k + 1$, where the latter is a direct contradiction when compared to the results reported by Chudik et al. (2011), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). The same steps can be used to show that $\|(NT)^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}})\boldsymbol{\varepsilon}_i\|$ and $\|(NT)^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}} - \mathbf{M}_{\widehat{\mathbf{F}}})\mathbf{F}\gamma_i\|$ are also non-negligible, unless $m = k + 1$.

[4]The process used to generate Figure 1 is a restricted version of (1)–(3) that sets $(\mathbf{f}_t', v_{i,t}', e_{i,t}) \sim N(\mathbf{0}, \mathbf{I}_5)$, $\boldsymbol{\beta} = [1,1]'$, $\gamma_i = \gamma = [1, u_1]'$ with $u_1 \sim U(1,2)$, and

$$
\mathbf{\Gamma}_i = \mathbf{\Gamma} = \begin{bmatrix} u_2 & 1 \\ u_3 & u_4 \end{bmatrix},
$$

where u_2, u_3 and u_4 are all drawn from $U(1.5, 2.5)$.

Figure 1: Divergent behaviour under $m < k+1$.



Note: The horizontal and vertical axes display $N = T$ and the average norm over $1,000$ replications, respectively.

The above results imply that many of the statements in the literature are actually yet to be proven. In what follows, we therefore propose an approach that is appropriate in general, provided that (9) is met. The idea is the same as when analysing sample second moment matrices where the elements are of different orders of magnitude (see, for example, Chang and Phillips, 1995), that is, we normalize to ensure convergence to a positive definite matrix. Let us therefore assume without loss of generality that the columns of $\overline{\mathbf{C}}$ can be organized such that $\overline{\mathbf{C}} = [\overline{\mathbf{C}}_m, \overline{\mathbf{C}}_{-m}]$, where $\overline{\mathbf{C}}_m$ is an $m \times m$ full rank matrix and $\overline{\mathbf{C}}_{-m}$ is $m \times (k+1-m)$. $\overline{\mathbf{U}}$ is partitioned conformably as $\overline{\mathbf{U}} = [\overline{\mathbf{U}}_m, \overline{\mathbf{U}}_{-m}]$. This implies

$$\widehat{\mathbf{F}} = [\mathbf{F}\overline{\mathbf{C}}_m, \mathbf{F}\overline{\mathbf{C}}_{-m}] + [\overline{\mathbf{U}}_m, \overline{\mathbf{U}}_{-m}]. \tag{20}$$

Define

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_m & \mathbf{B}_{-m} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{C}}_m^{-1} & -\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix},$$

which is of full rank under (9). Post-multiplying (20) by $\mathbf{B}$ yields

$$\widehat{\mathbf{F}}\mathbf{B} = \mathbf{F}\overline{\mathbf{C}}\mathbf{B} + \overline{\mathbf{U}}\mathbf{B} = [\mathbf{F}, \mathbf{0}_{T\times(k+1-m)}] + [\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}, \overline{\mathbf{U}}_{-m} - \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m}].$$

We now look for a conformable normalization matrix $\mathbf{D}_N$ such that

$$T^{-1}\mathbf{D}_N'\mathbf{B}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{B}\mathbf{D}_N = T^{-1}\mathbf{D}_N'\begin{bmatrix} \mathbf{F}'\mathbf{F} + (\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1} & -(\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m'(\overline{\mathbf{U}}_{-m} - \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m}) \\ -(\overline{\mathbf{U}}_{-m}' - \overline{\mathbf{C}}_{-m}'(\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m')\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1} & (\overline{\mathbf{U}}_{-m}' - \overline{\mathbf{C}}_{-m}'(\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m')(\overline{\mathbf{U}}_{-m} - \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m}) \end{bmatrix}\mathbf{D}_N$$

converges to a positive matrix. Here we make use of Lemma 1 of Pesaran (2006), which states that $\|\overline{\mathbf{U}}\| = O_p(N^{-1/2})$. Hence, while the upper left $m \times m$ block of $T^{-1}\mathbf{D}_N'\mathbf{B}'\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{B}\mathbf{D}_N$ converges to $\mathbf{\Sigma_f}$, the lower right $(k+1-m) \times (k+1-m)$ block is $O_p(N^{-1})$. This means that the required normalization matrix is given by $\mathbf{D}_N = \mathrm{diag}(\mathbf{I}_m, \sqrt{N}\mathbf{I}_{k+1-m})$. Hence, letting $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T\times(k+1-m)}]$ and $\overline{\mathbf{U}}^0 = \overline{\mathbf{U}}\mathbf{B}\mathbf{D}_N = [\overline{\mathbf{U}}_m^0, \overline{\mathbf{U}}_{-m}^0] = [\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}, \sqrt{N}(\overline{\mathbf{U}}_{-m} - \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m})]$, the resulting normalized version of $\widehat{\mathbf{F}}\mathbf{B}$ is given by

$$\widehat{\mathbf{F}}^0 = \widehat{\mathbf{F}}\mathbf{B}\mathbf{D}_N = \mathbf{F}^0 + \overline{\mathbf{U}}^0. \tag{21}$$

We now make use of $\widehat{\mathbf{F}}^0$ to establish the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$. We begin by noting that since $\mathrm{rank}(\overline{\mathbf{C}}) = m$ ($\overline{\mathbf{C}}$ has full row rank), we have $\overline{\mathbf{C}}^+ = \overline{\mathbf{C}}'(\overline{\mathbf{C}}\overline{\mathbf{C}}')^{-1}$, such that $\overline{\mathbf{C}}\overline{\mathbf{C}}^+ = (\overline{\mathbf{C}}^+)'\overline{\mathbf{C}}' = \mathbf{I}_m$ (see Abadir and Magnus, 2005, Exercise 10.31). This implies

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \widehat{\mathbf{F}}\overline{\mathbf{C}}^+\gamma_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i + \boldsymbol{\varepsilon}_i, \tag{22}$$

which can be substituted into (6), giving

$$\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}(\boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i). \tag{23}$$

Consider the numerator. Recall that $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^+\mathbf{A}'$ for any T rowed matrix $\mathbf{A}$. One of the key insights behind the approach used here is that since $\mathbf{B}\mathbf{D}_N$ is invertible, $\mathbf{P}_{\widehat{\mathbf{F}}} = \mathbf{P}_{\widehat{\mathbf{F}}^0}$. Hence, for our purposes, having $\widehat{\mathbf{F}}$ is just as good as having $\widehat{\mathbf{F}}^0$, although in practice the latter estimator is of course unobservable. The fact that $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$ implies that the second term in the numerator of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ can be written as follows:

$$\begin{aligned} &\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{P}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i\mathbf{F}'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{P}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i. \end{aligned} \tag{24}$$

We now make use of Theorem 1 to evaluate $[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+$, which is needed to be able to evaluate the last term on the right of (24). From the definition of $\widehat{\mathbf{F}}^0$,

$$T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0 = T^{-1}(\mathbf{F}^0)'\mathbf{F}^0 + T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}^0 + T^{-1}(\overline{\mathbf{U}}^0)'\mathbf{F}^0 + T^{-1}(\overline{\mathbf{U}}^0)'\overline{\mathbf{U}}^0, \tag{25}$$

where

$$T^{-1}(\mathbf{F}^0)'\mathbf{F}^0 = \begin{bmatrix} T^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{0}_{(k+1-m)\times(k+1-m)} \end{bmatrix}. \tag{26}$$

Also,

$$T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}^0 = \begin{bmatrix} T^{-1}\mathbf{F}'\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1} & \sqrt{N}T^{-1}(\mathbf{F}'\overline{\mathbf{U}}_{-m} - \mathbf{F}'\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m}) \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{0}_{(k+1-m)\times(k+1-m)} \end{bmatrix}.$$

By (A.11) in Lemma 2 of Pesaran (2006), $\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_m\|$ and $\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_{-m}\|$ are both $O_p((NT)^{-1/2})$. This implies

$$\|T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}^0\| = O_p(T^{-1/2}). \tag{27}$$

Also, by using the decomposition of $\overline{\mathbf{U}}^0$ into $\overline{\mathbf{U}}_m^0$ and $\overline{\mathbf{U}}_{-m}^0$,

$$T^{-1}(\overline{\mathbf{U}}^0)'\overline{\mathbf{U}}^0 = T^{-1}\begin{bmatrix} (\overline{\mathbf{U}}_m^0)'\overline{\mathbf{U}}_m^0 & (\overline{\mathbf{U}}_m^0)'\overline{\mathbf{U}}_{-m}^0 \\ (\overline{\mathbf{U}}_{-m}^0)'\overline{\mathbf{U}}_m^0 & (\overline{\mathbf{U}}_{-m}^0)'\overline{\mathbf{U}}_{-m}^0 \end{bmatrix}.$$

By (A.10) in Lemma 2 of Pesaran (2006), we have $\|T^{-1}\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m\| = O_p(N^{-1})$, and by further use of Assumption 1 (iv), $\|\overline{\mathbf{C}}_m\|$ and $\|\overline{\mathbf{C}}_m^{-1}\|$ are $O_p(1)$. It follows that

$$\begin{aligned} \|T^{-1}(\overline{\mathbf{U}}_m^0)'\overline{\mathbf{U}}_m^0\| &= N^{-1}\|(\overline{\mathbf{C}}_m^{-1})'(NT^{-1}\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m)\overline{\mathbf{C}}_m^{-1}\| \leq N^{-1}\|\overline{\mathbf{C}}_m^{-1}\|^2\,\|(NT^{-1}\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m)\| \\ &= O_p(N^{-1}). \end{aligned} \tag{28}$$

This implies

$$\begin{aligned} \|T^{-1}(\overline{\mathbf{U}}_m^0)'\overline{\mathbf{U}}_{-m}^0\| &= \|\sqrt{N}T^{-1}(\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_{-m} - \sqrt{N}T^{-1}(\overline{\mathbf{C}}_m^{-1})'\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m}\| \\ &\leq N^{-1/2}\|\overline{\mathbf{C}}_m^{-1}\|\,\|NT^{-1}\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_{-m}\| + N^{-1/2}\|\overline{\mathbf{C}}_m^{-1}\|^2\,\|NT^{-1}\overline{\mathbf{U}}_m'\overline{\mathbf{U}}_m\|\,\|\overline{\mathbf{C}}_{-m}\| \\ &= O_p(N^{-1/2}). \end{aligned} \tag{29}$$

It remains to consider

$$T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\overline{\mathbf{U}}_{-m}^0 = [-(\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m})', \mathbf{I}_{k+1-m}](NT^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}})\begin{bmatrix} -\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m} \\ \mathbf{I}_{k+1-m} \end{bmatrix}. \tag{30}$$

A straightforward calculation reveals that $E(\|N^{-1}T^{-1/2}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^T \mathbf{u}_{i,t}\mathbf{u}_{j,t}'\|^2) = O(1)$. This implies $\|N^{-1}T^{-1/2}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^T \mathbf{u}_{i,t}\mathbf{u}_{j,t}'\| = O_p(1)$. A similar calculation can be used to show that since the fourth moment of $\mathbf{u}_{i,t}$ is finite, we have $\|(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{u}_{i,t}\mathbf{u}_{i,t}' - \mathbf{\Sigma}_{\mathbf{u}})\| = O_p(1)$ where $\mathbf{\Sigma}_{\mathbf{u}} = E(\mathbf{u}_{i,t}\mathbf{u}_{i,t}')$. Hence,

$$\begin{aligned} NT^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{u}_{i,t}\mathbf{u}_{j,t}' \\ &= \mathbf{\Sigma}_{\mathbf{u}} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{u}_{i,t}\mathbf{u}_{i,t}' - \mathbf{\Sigma}_{\mathbf{u}}) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\mathbf{u}_{i,t}\mathbf{u}_{j,t}' \\ &= \mathbf{\Sigma}_{\mathbf{u}} + O_p((NT)^{-1/2}) + O_p(T^{-1/2}) = \mathbf{\Sigma}_{\mathbf{u}} + O_p(T^{-1/2}), \end{aligned} \tag{31}$$

which in turn implies

$$\begin{aligned} T^{-1}(\overline{\mathbf{U}}_{-m}^{0})'\overline{\mathbf{U}}_{-m}^{0} &= [-(\overline{\mathbf{C}}_{m}^{-1}\overline{\mathbf{C}}_{-m})', \mathbf{I}_{k+1-m}]\mathbf{\Sigma}_{\mathbf{u}}\begin{bmatrix} -\overline{\mathbf{C}}_{m}^{-1}\overline{\mathbf{C}}_{-m} \\ \mathbf{I}_{k+1-m} \end{bmatrix} + O_p(T^{-1/2}) \\ &= \mathbf{\Sigma}_{\mathbf{u}_{-m}^{0}} + O_p(T^{-1/2}) \end{aligned} \tag{32}$$

with an obvious definition of $\mathbf{\Sigma}_{\mathbf{u}_{-m}^{0}}$, a $(k+1-m)\times(k+1-m)$ matrix. It is important to note that this matrix is in fact positive definite. By adding the results in (28)–(32), we obtain

$$T^{-1}(\overline{\mathbf{U}}^{0})'\overline{\mathbf{U}}^{0} = \begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Sigma}_{\mathbf{u}_{-m}^{0}} \end{bmatrix} + O_p(N^{-1/2}) + O_p(T^{-1/2}), \tag{33}$$

This shows that the expression for $T^{-1}(\widehat{\mathbf{F}}^{0})'\widehat{\mathbf{F}}^{0}$ in (25) reduces to

$$T^{-1}(\widehat{\mathbf{F}}^{0})'\widehat{\mathbf{F}}^{0} = \mathbf{\Sigma}_{\mathbf{f}^{0}} + O_p(N^{-1/2}) + O_p(T^{-1/2}), \tag{34}$$

where

$$\mathbf{\Sigma}_{\mathbf{f}^{0}} = \begin{bmatrix} T^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Sigma}_{\mathbf{u}_{-m}^{0}} \end{bmatrix}.$$

Note how $T^{-1}\mathbf{F}'\mathbf{F} = \mathbf{\Sigma}_{\mathbf{f}} + O_p(T^{-1/2})$. Thus, to the stated order of approximation, $T^{-1}\mathbf{F}'\mathbf{F}$ in (34) is actually equal to $\mathbf{\Sigma}_{\mathbf{f}}$. For reasons that will soon be clear, however, the given definition of $\mathbf{\Sigma}_{\mathbf{f}^{0}}$ is more convenient. We begin by noting how $\text{rank}(T^{-1}\mathbf{F}'\mathbf{F}) \stackrel{a.s.}{\to} m$ as $T\to\infty$, which, together with the positive definiteness of $\mathbf{\Sigma}_{\mathbf{u}_{-m}^{0}}$, in turn implies $\text{rank}(\mathbf{\Sigma}_{\mathbf{f}^{0}}) \stackrel{a.s.}{\to} k+1$. But we also have $\text{rank}[T^{-1}(\widehat{\mathbf{F}}^{0})'\widehat{\mathbf{F}}^{0}] \stackrel{a.s.}{\to} k+1$ as N, $T\to\infty$, and so we obtain

$$\text{rank}[T^{-1}(\widehat{\mathbf{F}}^{0})'\widehat{\mathbf{F}}^{0}] \stackrel{a.s.}{\to} \text{rank}(\mathbf{\Sigma}_{\mathbf{f}^{0}}). \tag{35}$$

By using this result, (34) and Theorem 1, we obtain the following key result:

$$[T^{-1}(\widehat{\mathbf{F}}^{0})'\widehat{\mathbf{F}}^{0}]^{+} = \mathbf{\Sigma}_{\mathbf{f}^{0}}^{+} + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{36}$$

Lemma 1 summarizes this.

Lemma 1. *Under Assumption 1 and condition* (9), *as* $N, T \to \infty$,

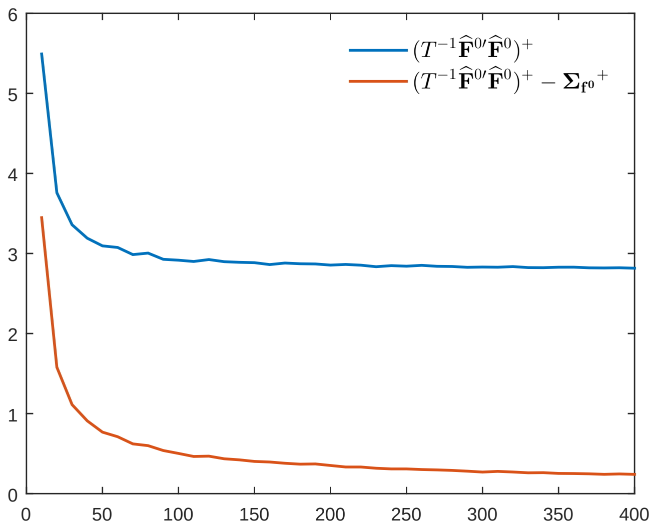
$$\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+\| = O_p(N^{-1/2}) + O_p(T^{-1/2}).$$

An important implication of Lemma 1 is that

$$\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\| = O_p(1). \tag{37}$$

Hence, in contrast to the divergence result obtained in (18) for $\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\|$ and $\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|$, $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+\|$ and $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\|$ do behave "nicely". This is illustrated in Figure 2, which plots the average of these norms over 1,000 replications of the same data generating process used to generate Figure 1. As expected, while $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+\|$ converges to zero, $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\|$ converges to a constant. Note in particular how the properties of $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\|$ and $\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+\|$ are roughly the same as those of $\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+\|$ and $\|(T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^+ - (T^{-1}\overline{\mathbf{C}}'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^+\|$ when (9) is satisfied with equality. It is therefore reasonable to expect that the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ under $m < k+1$ should not be that different from the one that applies when $m = k+1$.

Figure 2: Convergent behaviour under $m < k+1$.



Note: The horizontal and vertical axes display $N = T$ and the average norm over 1,000 replications, respectively.

Let us now consider $\mathbf{X}_i'\widehat{\mathbf{F}}^0$, which is again needed for evaluating the last term on the right of (24). By using the definition of $\mathbf{X}_i$ and $\widehat{\mathbf{F}}' = (\mathbf{B}')^{-1}\mathbf{D}_N^{-1}(\widehat{\mathbf{F}}^0)'$, we obtain the following very useful expression for $\mathbf{X}_i$:

$$
\begin{aligned}
\mathbf{X}_i &= \mathbf{F}\mathbf{\Gamma}_i' + \mathbf{V}_i = \mathbf{F}\overline{\mathbf{C}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' + \mathbf{V}_i = \widehat{\mathbf{F}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' - (\widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{C}})\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' + \mathbf{V}_i \\
&= \widehat{\mathbf{F}}^0\mathbf{D}_N^{-1}\mathbf{B}^{-1}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' - \overline{\mathbf{U}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' + \mathbf{V}_i. \qquad (38)
\end{aligned}
$$

Note in particular how $(\widehat{\mathbf{F}}^0)'\mathbf{M}_{\widehat{\mathbf{F}}^0} = \mathbf{0}_{(k+1)\times T}$, and therefore $\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0} = [\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}']\mathbf{M}_{\widehat{\mathbf{F}}}$, which in turn implies

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N [\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}']\mathbf{M}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\
&= \mathbf{R}_{0NT} - \mathbf{R}_{1NT} - \mathbf{R}_{2NT}, \qquad (39)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{R}_{0NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N [\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}']\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i, \\
\mathbf{R}_{1NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N [\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}']\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i, \\
\mathbf{R}_{2NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N [\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'](\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i.
\end{aligned}
$$

From this point on the analysis is very similar to the analyses of Chudik et al. (2011), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). The key difference is that here we make use of the fact that $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$, which allows us to proceed in roughly the same way as when using $\widehat{\mathbf{F}}$ and assuming that (9) is satisfied with equality. As an illustration, consider $T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}$ in $\mathbf{R}_{2NT}$. Similarly to (8), we have that

$$
\begin{aligned}
\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0} &= \overline{\mathbf{U}}^0[(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\overline{\mathbf{U}}^{0\prime} + \overline{\mathbf{U}}^0[(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+(\mathbf{F}^0)' + \mathbf{F}^0[(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+\overline{\mathbf{U}}^{0\prime} \\
&+ \mathbf{F}^0([(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - [(\mathbf{F}^0)'\mathbf{F}^0]^+)(\mathbf{F}^0)'.
\end{aligned}
$$

Consider the last term on the right. Note how

$$
\mathbf{\Sigma}_{\mathbf{f}^0}^+ = \begin{bmatrix} (T^{-1}\mathbf{F}'\mathbf{F})^+ & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Sigma}_{\mathbf{u}_{-m}^0}^+ \end{bmatrix} = [T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^+ + \begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Sigma}_{\mathbf{u}_{-m}^0}^+ \end{bmatrix},
$$

implying that, since the last $k+1-m$ columns of $\mathbf{F}^0$ are zero,

$$
\begin{aligned}
&\mathbf{F}^0([T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - [T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^+)(\mathbf{F}^0)' \\
&= \mathbf{F}^0([T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+)(\mathbf{F}^0)' - \mathbf{F}^0\begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Sigma}_{\mathbf{u}_{-m}^0}^+ \end{bmatrix}(\mathbf{F}^0)' \\
&= \mathbf{F}^0([T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}_{\mathbf{f}^0}^+)(\mathbf{F}^0)'. \qquad (40)
\end{aligned}
$$

Hence,

$$\begin{aligned}
T(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}) &= \overline{\mathbf{U}}^0[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+}\overline{\mathbf{U}}^{0\prime} + \overline{\mathbf{U}}[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+}(\mathbf{F}^0)' + \mathbf{F}^0[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+}\overline{U}^{0\prime} \\
&+ \mathbf{F}^0([T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+} - \boldsymbol{\Sigma}_{\mathbf{f}^0}^{+})(\mathbf{F}^0)' \\
&= I + II + III + IV. \qquad (41)
\end{aligned}$$

Using (34), we have

$$\begin{aligned}
I &= \overline{\mathbf{U}}_m^0(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\overline{\mathbf{U}}_m^{0\prime} + \overline{\mathbf{U}}_{-m}^0(T^{-1}\overline{\mathbf{U}}_{-m}^{0\prime}\overline{\mathbf{U}}_{-m}^0)^{-1}\overline{\mathbf{U}}_{-m}^{0\prime} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \overline{\mathbf{U}}_m^0(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\overline{\mathbf{U}}_m^{0\prime} + T\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0} + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned}$$

and

$$II = \overline{\mathbf{U}}_m^0(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' + \overline{\mathbf{U}}_{-m}^0(T^{-1}\overline{\mathbf{U}}_{-m}^{0\prime}\overline{\mathbf{U}}_{-m}^0)^{-1}\mathbf{0} + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

with III being the transpose of II. Disregarding from dominated terms, these results imply

$$\begin{aligned}
&T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}} \\
&= T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}_m^0[T^{-1}\mathbf{F}'\mathbf{F}]^{-1}T^{-1}\overline{\mathbf{U}}_m^{0\prime}\overline{\mathbf{U}} + T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}} + T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}_m^0[T^{-1}\mathbf{F}'\mathbf{F}]^{-1}T^{-1}\mathbf{F}'\overline{\mathbf{U}} \\
&+ T^{-1}\overline{\mathbf{U}}'\mathbf{F}[T^{-1}\mathbf{F}'\mathbf{F}]^{+}T^{-1}\overline{\mathbf{U}}_m^{0\prime}\overline{\mathbf{U}} \\
&+ T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0([T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+} - \boldsymbol{\Sigma}_{\mathbf{f}^0}^{+})T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}. \qquad (42)
\end{aligned}$$

Note now that $\overline{\mathbf{U}}_m^0 = \overline{\mathbf{U}}\mathbf{B}_m$, where $\mathbf{B}_m$ contains the first m columns of $\mathbf{B}$. By (A.10) and (A.11) in Lemma 2 of Pesaran (2006), we have $\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\| = O_p(N^{-1})$ and $\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\| = O_p((NT)^{-1/2})$. Furthermore, $\|\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}\| \leq \|\overline{\mathbf{U}}'\overline{\mathbf{U}}\|$. Making use of these results, Lemma 1 and (37), we can show that

$$\begin{aligned}
&\|T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}} - T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}\| \\
&\leq \|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|^2\|\mathbf{B}_m\|^2\|[T^{-1}\mathbf{F}'\mathbf{F}]^{+}\| \\
&+ 2\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|\|[T^{-1}\mathbf{F}'\mathbf{F}]^{+}\|\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\| + \|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|^2\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^{+} - \boldsymbol{\Sigma}_{\mathbf{f}^0}^{+}\| \\
&= O_p(N^{-2}) + O_p(N^{-3/2}T^{-1/2}) + O_p(N^{-1}T^{-3/2}), \qquad (43)
\end{aligned}$$

so that

$$T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}} = T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}} + O_p(N^{-2}) + O_p(N^{-3/2}T^{-1/2}) + O_p(N^{-1}T^{-3/2}). \qquad (44)$$

which is of order $O_p(N^{-1})$ due to the first term on the RHS. In fact, the same kind of result is obtained for the remaining part of $\|\mathbf{R}_{2NT}\|$. However, since the arguments used in showing this result are very similar to those used in (43), we put the details in Appendix and just provide here the final result.

Lemma 2. *Under Assumption 1 and condition* (9)*, as* $N, T \to \infty$ *with* $T/N \to \tau < \infty$*,*

(i) $\mathbf{R}_{2NT} = \sqrt{T}N^{-1/2}\overline{\mathbf{d}}_1 + O_p(N^{-1}) + O_p(T^{-1/2})$;

(ii) $\|\mathbf{R}_{1NT}\| = O_p(T^{-1/2})$;

(iii) $\mathbf{R}_{0NT} = \sqrt{T}N^{-1/2}(\overline{\mathbf{b}}_1 - \overline{\mathbf{b}}_2) + O_p(N^{-1/2}) + O_p(T^{-1/2})$,

where

$$\begin{aligned}
\overline{\mathbf{b}}_1 &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}[\boldsymbol{\beta}, \mathbf{I}_k]\overline{\mathbf{C}}^{+}\gamma_i, \\
\overline{\mathbf{b}}_2 &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\mathbf{\Sigma}_{\mathbf{u}}\overline{\mathbf{C}}^{+}\gamma_i, \\
\overline{\mathbf{d}}_1 &= T^{-1}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i - T^{-1}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i.
\end{aligned}$$

Next up is $(NT)^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_i$, the first term in the numerator of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$, which in view of $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$ and (38) can be written as follows:

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\boldsymbol{\varepsilon}_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\boldsymbol{\varepsilon}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}']\mathbf{M}_{\widehat{\mathbf{F}}^0}\boldsymbol{\varepsilon}_i \\
&= \mathbf{Q}_{0NT} - \mathbf{Q}_{1NT} - \mathbf{Q}_{2NT}, \qquad (45)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Q}_{0NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}']\boldsymbol{\varepsilon}_i, \\
\mathbf{Q}_{1NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}']\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i, \\
\mathbf{Q}_{2NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\mathbf{V}_i' - \mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'](\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i.
\end{aligned}$$

Again, once $\widehat{\mathbf{F}}$ has been replaced by $\widehat{\mathbf{F}}^0$, the evaluation of $\mathbf{Q}_{0NT}$, $\mathbf{Q}_{1NT}$ and $\mathbf{Q}_{2NT}$ is the same as when using $\widehat{\mathbf{F}}$ and assuming that (9) is satisfied with equality.

Lemma 3. *Under the conditions of Lemma 2,*

(i) $\mathbf{Q}_{2NT} = \sqrt{T}N^{-1/2}\overline{\mathbf{d}}_2 + O_p(N^{-1}) + O_p(T^{-1/2}) + O_p(\sqrt{N}T^{-1})$;

(ii) $\|\mathbf{Q}_{1NT}\| = O_p(T^{-1/2})$;

(iii) $\mathbf{Q}_{0NT} = (NT)^{-1/2}\sum_{i=1}^{N}\mathbf{V}_i'\varepsilon_i - \sqrt{T}N^{-1/2}\overline{\mathbf{b}}_3 + O_p(N^{-1/2}) + O_p(T^{-1/2})$,

where

$$
\begin{aligned}
\overline{\mathbf{b}}_3 &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\sigma^2[1, \mathbf{0}_{k\times 1}']', \\
\overline{\mathbf{d}}_2 &= T^{-1}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\varepsilon_i - T^{-1}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\varepsilon_i.
\end{aligned}
$$

In view of (39), Lemmas 2 and 3 imply

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}(\varepsilon_i - \overline{\mathbf{U}\mathbf{C}}^{+}\gamma_i) \\
&\quad = \tfrac{1}{\sqrt{NT}}\textstyle\sum_{i=1}^{N}\mathbf{V}_i'\varepsilon_i + \sqrt{T}N^{-1/2}(\overline{\mathbf{b}} - \overline{\mathbf{d}}) + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{N}T^{-1}), \qquad (46)
\end{aligned}
$$

where $\overline{\mathbf{b}} = \overline{\mathbf{b}}_1 - \overline{\mathbf{b}}_2 - \overline{\mathbf{b}}_3$ and $\overline{\mathbf{d}} = \overline{\mathbf{d}}_1 + \overline{\mathbf{d}}_2$. Moreover, since the fourth order moments of $\mathbf{v}_{i,t}$ and $\varepsilon_{i,t}$ are bounded by Assumption 1, by a central limit law for iid processes,

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\varepsilon_i \xrightarrow{d} N(\mathbf{0}_{m\times 1}, \sigma^2\mathbf{\Sigma}) \qquad (47)
$$

as N, $T \to \infty$. This completes the analysis of the numerator of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$. What is missing now is the denominator.

Lemma 4. *Under conditions of Lemma 2,*

$$
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i = \mathbf{\Sigma} + O_p(N^{-1}) + O_p(T^{-1/2}).
$$

The results in Lemmas 2–4 can be inserted together with (48) into (23), leading to the following asymptotic distribution for $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$:

$$
\begin{aligned}
\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta}) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}(\varepsilon_i - \overline{\mathbf{U}\mathbf{C}}^{+}\gamma_i) \\
&= \mathbf{\Sigma}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\varepsilon_i + \sqrt{T}N^{-1/2}(\overline{\mathbf{b}} - \overline{\mathbf{d}})\right) + o_p(1) \qquad (48) \\
&\xrightarrow{d} N(\mathbf{0}_{m\times 1}, \sigma^2\mathbf{\Sigma}^{-1}) + \lim_{N,T\to\infty}\sqrt{T}N^{-1/2}\mathbf{\Sigma}^{-1}\overline{\mathbf{b}}, \qquad (49)
\end{aligned}
$$

as N, $T \to \infty$ with $T/N \to \tau < \infty$ and $\sqrt{N}/T \to 0$.

Note that the bias term $\overline{\mathbf{d}}$ is absent in the last line. This is due to the fact that the estimation error of $\hat{\mathbf{F}}$, which is simply $\overline{\mathbf{U}}$, disappears asymptotically. The rotated and rescaled estimation error $\overline{\mathbf{U}}^0$ must hence be defined to be $\mathbf{0}$ in the limit in order to be consistent with the characteristics of $\overline{\mathbf{U}}$. This naturally implies that $\overline{\mathbf{U}}^0_{-m}$ is discontinuous; it converges towards the limiting distribution of $\sqrt{N}\overline{\mathbf{U}}\mathbf{B}_{-m}$ but drops to zero asymptotically. Hence, the bias term involving $\overline{\mathbf{U}}^0_{-m}$ is zero asymptotically while being present for all finite N, T. In the latter case, the distribution of $\widehat{\mathbf{b}}_P$ can be given a more convenient representation. Using the finite-sample expressions for $\mathbf{R}_{0NT}$ and $\mathbf{Q}_{0NT}$ instead of their limit, we can write

$$\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta}) \quad = \quad \boldsymbol{\Sigma}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\overline{\mathbf{U}}^0_{-m}}\varepsilon_i + \overline{\mathbf{b}}^*\right), \tag{50}$$

where

$$\overline{\mathbf{b}}^* = T^{-1}\sum_{i=1}^{N}(\mathbf{V}_i'\mathbf{M}_{\overline{\mathbf{U}}^0_{-m}}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{M}_{\overline{\mathbf{U}}^0_{-m}}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{M}_{\overline{\mathbf{U}}^0_{-m}}\varepsilon_i) \tag{51}$$

converges to the equivalent of $\mathbf{b}$ after projecting $\varepsilon_i, \mathbf{V}_i$ and $\overline{\mathbf{U}}$ away from $\overline{\mathbf{U}}^0_{-m}$. An implication of this result is that $\|\mathbf{b}^*\| \leq \|\mathbf{b}\|$ if $m+1 > r$ which means that the inclusion of additional variables potentially reduces the bias relative to the case where $m+1 = r$. However, this result has to be contrasted to the risk of overfitting $\hat{\mathbf{F}}$ when including more covariates into the analysis. Lastly, note that $\overline{\mathbf{d}}$ disappears from (49) completely if $m+1 = r$ since the transformed estimation error $\overline{\mathbf{U}}^0$ consists only of the first m columns $\overline{\mathbf{U}}^0_m$.

The fact that $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ is asymptotically normal is in agreement with the existing results in the literature. However, unlike the proofs of these other results, the proof of (49) is valid regardless of whether (9) is satisfied with equality or strict inequality. The main contribution here is therefore the method of proof, which can be used to study the asymptotic distribution of $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ under more general conditions, or to study the asymptotic properties of other CCE estimators, such as the mean group estimator of Pesaran (2006). The proof also demonstrates how the bulk of the existing results developed for the case when $T/N \to 0$ (see, for example, Pesaran, 2006; Chudik et al., 2011; Kapetanios et al., 2010; Pesaran et al., 2013) can be extended to the more general $T/N \to \tau < \infty$ case considered here. The fact that $\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta})$ is biased in the latter case is in agreement with the results reported by Bai (2009), and Westerlund and Urbain (2015). If, however, the more restrictive condition is met, such that $T/N \to 0$, then the asymptotic distribution reported in (49) reduces to

$$\sqrt{NT}(\widehat{\mathbf{b}}_P - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}_{m\times 1}, \sigma^2 \boldsymbol{\Sigma}^{-1}),$$

which under Assumption 1 is the same as in Theorem 4 of Pesaran (2006).

4 Conclusion

The CCE approach of Pesaran (2006) has attracted considerable interest in the literature on factor-augmented panel regressions. In fact, the approach has in a few years established itself as the workhorse of the literature with a large number of theoretical extensions and countless applications. In the present paper we point to a problem with the CCE approach that seems to have gone largely unnoticed in this literature. The problem occurs in the empirically relevant case when $m < k + 1$. Specifically, the use of too many observables causes the second moment matrix of the estimated factors to become asymptotically singular, which in turn invalidates some of the arguments commonly used to establish asymptotic theory. Hence, the bulk of existing theories is actually yet to be proven. A new method of proof is therefore proposed that is shown to alleviate the singularity problem, leading to a straightforward asymptotic analysis. The model that we consider is chosen with an eye on transparency of the derivations. However, the new method can be easily extended to cover also more general models, as well as other CCE-based estimators.

References

Abadir, K. M., and J. R. Magnus (2005). *Matrix algebra*, volume 1. Cambridge University Press.

Andrews, D. W. K. (1987). Asymptotic Results for Generalized Wald Tests. *Econometric Theory* **3**, 348–358.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica* **77**, 1229–1279.

Chang, Y., and P. C. B. Phillips (1995). Time Series Regression with Mixtures of Integrated Processes. *Econometric Theory* **10**, 1033–1094.

Chudik, A., and M. H. Pesaran (2011). Econometric Analysis of High Dimensional VARs Featuring a Dominant Unit. *Econometric Reviews* **32**, 592–649.

Chudik, A. M., M. H. Pesaran, and E. Tosetti (2011). Weak and Strong Cross Section Dependence and Estimation of Large Panels. *Econometrics Journal* **14**, C45–C90.

Chudik, A. M., and M. H. Pesaran (2015a). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *Journal of Econometrics* **188**, 393–420.

Chudik, A. M., and M. H. Pesaran (2015b). Large Panel Data Models with Cross-Sectional Dependence: A Survey. In B. H. Baltagi (Ed.), *The Oxford Handbook on Panel Data*. Oxford University Press.

Feng, X., and Z. Zhang (2007). The Rank of a Random Matrix. *Applied Mathematics and Computation* **185**, 689–694.

Horn, R. A., and C. R. Johnson (1985). *Matrix Analysis*. Cambridge University Press. New York.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with Non-Stationary Multifactor Error Structures. *Journal of Econometrics*, **160**, 326–348.

Noble, B. (1976). Methods for Computing the Moore-Penrose Generalized Inverse, and Related Matters. In M. Z. Nashed (Ed.), *Generalized Inverses and Applications*. Academic Press.

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* **74**, 967–1012.

Pesaran, H. M. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.

Pesaran, H. M., L. V. Smith, and T. Yamagata (2013). Panel Unit Root Tests in the Presence of a Multifactor Error Structure. *Journal of Econometrics* **175**, 94–115.

Reese, S. and J. Westerlund (2015a). Estimation of Factor-Augmented Panel Regressions with Weakly Influential Factors. Forthcoming in *Econometric Reviews*.

Reese, S. and J. Westerlund (2015b). PANICCA – PANIC on Cross-Section Averages. Forthcoming in *Journal of Applied Econometrics*.

Stewart, G. W. (1969). On the Continuity of the Generalized Inverse. *SIAM Journal of Applied Mathematics* **17**, 33–45.

Stewart, G. W. (1977). On the Perturbation of Pseudo-Inverses, Projections and Linear Least Squares Problems. *SIAM Review* **19**, 634–662.

Stewart, G. W. (1990). Stochastic Perturbation Theory. *SIAM review* **32**, 579–610.

Wedin, P.-Å. (1973). Perturbation Theory for Pseudo-Inverses. *BIT* **13**, 217–232.

Westerlund, J. (2015). The Effect of Recursive Detrending on Panel Unit Root Tests. *Journal of Econometrics* **185**, 452–467.

Westerlund, J, and J.-P. Urbain (2015). Cross-Sectional Averages versus Principal Components. *Journal of Econometrics* **185**, 372–377.

Appendix: Proofs

Some notation. Let $\mathbf{A}$ denote a generic square matrix of dimension $r \times r$. Let $\lambda_1(\mathbf{A}) \geq ... \geq \lambda_r(\mathbf{A})$ be the ordered eigenvalues of $\mathbf{A}'\mathbf{A}$, and let $\rho(\mathbf{A}) = \max_{j=1,...,r} |\lambda_j(\mathbf{A})| = \lambda_1(\mathbf{A})$ be the spectral radius of $\mathbf{A}$. Denote by $\sigma_j(\mathbf{A}) = \sqrt{\lambda_j(\mathbf{A})}$ the j-th singular value of $\mathbf{A}$. The spectral norm of $\mathbf{A}$ is given by $\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A})} = \sqrt{\lambda_1(\mathbf{A})} = \sigma_1(\mathbf{A})$. Note how $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|$ (see Horn and Johnson, 1985, Corollary 5.6.35) and $\|\mathbf{A}\| \leq \sqrt{r}\|\mathbf{A}\|_2$.

Proof of Theorem 1.

Note that $\mathbf{A}_n$ that has a dimension of $r \times r$. Now, suppose without loss of generality, that $\text{rank}(\mathbf{A}_n) = r_n = r$ for all n.[5] Hence, $r_n = r = r_0$. We begin by noting that by Theorem 4.1 of Noble (1976),

$$\mathbf{A}_n^+ - \mathbf{A}_0^+ = -\mathbf{A}_n^+\mathbf{E}_n\mathbf{A}_0^+ + \mathbf{A}_n^+(\mathbf{A}_n^+)'\mathbf{E}_n'(\mathbf{I}_r - \mathbf{A}_0\mathbf{A}_0^+) + (\mathbf{I}_r - \mathbf{A}_n^+\mathbf{A}_n)\mathbf{E}_n'(\mathbf{A}_0^+)'\mathbf{A}_0^+. \tag{A1}$$

Since $r_n = r$, we have $\mathbf{A}_n^+ = (\mathbf{A}_n'\mathbf{A}_n)^{-1}\mathbf{A}_n'$, which in turn implies $(\mathbf{I}_r - \mathbf{A}_n^+\mathbf{A}_n) = \mathbf{0}_{r\times r}$. Also, $\mathbf{I}_r - \mathbf{A}_0\mathbf{A}_0^+ = \mathbf{I}_r - \mathbf{A}_0(\mathbf{A}_0'\mathbf{A}_0)^{-1}\mathbf{A}_0'$, which is symmetric, idempotent (see Abadir and Magnus, 2005, Exercise 10.26), and therefore $\|\mathbf{I}_r - \mathbf{A}_0\mathbf{A}_0^+\|_2 = 1$. By using this, and the fact that $\|\mathbf{A}'\|_2 = \|\mathbf{A}\|_2$ for any matrix $\mathbf{A}$, we obtain

$$\begin{aligned} \|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 &\leq \|\mathbf{E}_n\|_2\|\mathbf{A}_n^+\|_2(\|\mathbf{A}_0^+\|_2 + \|\mathbf{A}_n^+\|_2\|\mathbf{I}_r - \mathbf{A}_0\mathbf{A}_0^+\|_2) \\ &= \|\mathbf{E}_n\|_2\|\mathbf{A}_n^+\|_2(\|\mathbf{A}_0^+\|_2 + \|\mathbf{A}_n^+\|_2). \end{aligned} \tag{A2}$$

Consider $\mathbf{A}_n$. By the singular value decomposition (SVD), $\mathbf{A}_n = \mathbf{V}_n\mathbf{\Sigma}_n\mathbf{U}_n'$, where the $r \times r$ matrix $\mathbf{\Sigma}_n$ is a diagonal matrix that contains the singular values of $\mathbf{A}_n$ on its diagonal, such that $\mathbf{\Sigma}_n = \text{diag}[\sigma_1(\mathbf{A}_n), ..., \sigma_r(\mathbf{A}_n)]$ with $\sigma_1(\mathbf{A}_n) \geq \ldots \geq \sigma_r(\mathbf{A}_n)$, and where $\mathbf{V}_n$ and $\mathbf{U}_n$ are both $r \times r$ matrices that contain the left and right singular vectors of $\mathbf{A}_n$, respectively. Note that $\mathbf{V}_n$ and $\mathbf{U}_n$ are orthogonal matrices, that is, $\mathbf{V}_n'\mathbf{V}_n = \mathbf{V}_n\mathbf{V}_n' = \mathbf{I}_r$ and $\mathbf{U}_n'\mathbf{U}_n = \mathbf{U}_n\mathbf{U}_n' = \mathbf{I}_r$. By the properties of the MP inverse (see Abadir and Magnus, 2005, Exercise 10.29), $\mathbf{A}_n^+ = (\mathbf{V}_n\mathbf{\Sigma}_n\mathbf{U}_n')^+ = \mathbf{U}_n\mathbf{\Sigma}_n^+\mathbf{V}_n'$, where $\mathbf{\Sigma}_n^+ = \text{diag}[\sigma_1(\mathbf{A}_n)^{-1}, ..., \sigma_r(\mathbf{A}_n)^{-1}]$. Now, by applying Weyl's theorem (see Horn and Johnson, 1985, Theorem 4.3.1) to $\mathbf{A}_0 = \mathbf{A}_n - \mathbf{E}_n$, we obtain $\sigma_p(\mathbf{A}_0) \leq$

[5]Without loss of generality holds because one can decompose any sequence $\mathbf{A}_n$ into a finite number of different subsequences, each of which has the property that all its terms have the same rank. Once the proof of the theorem has been established for any such subsequence, it is easy to see that it holds for the sequence as a whole (see Andrews, 1987, page 354).

$\sigma_p(\mathbf{A}_n) + \sigma_1(-\mathbf{E}_n)$ for all $p = 1, ..., r$. Hence, letting $p = r$, we have

$$\sigma_r(\mathbf{A}_0) \leq \sigma_r(\mathbf{A}_n) + \sigma_1(-\mathbf{E}_n). \tag{A3}$$

But $\sigma_1(-\mathbf{E}_n) = -\sigma_1(\mathbf{E}_n) = -\|\mathbf{E}_n\|_2$, $\|\mathbf{A}_n^+\|_2 = \sigma_r(\mathbf{A}_n)^{-1}$ and $\|\mathbf{A}_0^+\|_2 = \sigma_r(\mathbf{A}_0)^{-1}$, giving

$$\|\mathbf{A}_n^+\|_2^{-1} \geq \|\mathbf{A}_0^+\|_2^{-1} - \|\mathbf{E}_n\|_2 = \|\mathbf{A}_0^+\|_2^{-1}(1 - \|\mathbf{A}_0^+\|_2\|\mathbf{E}_n\|_2),$$

or

$$\|\mathbf{A}_n^+\|_2 \leq \|\mathbf{A}_0^+\|_2(1 - \|\mathbf{A}_0^+\|_2\|\mathbf{E}_n\|_2)^{-1}, \tag{A4}$$

where the denominator converges to one as $\|\mathbf{E}_n\|_2$ goes to zero since $\|\mathbf{A}_0^+\|_2 = \sigma_r(\mathbf{A}_0)^{-1} < \infty$. Direct insertion into (A2) now yields

$$\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 \leq \|\mathbf{E}_n\|_2 \frac{\|\mathbf{A}_n^+\|_2^2}{1 - \|\mathbf{A}_0^+\|_2\|\mathbf{E}_n\|_2}\left(1 + \frac{1}{1 - \|\mathbf{A}_0^+\|_2\|\mathbf{E}_n\|_2}\right) = o_p(1). \tag{A5}$$

The proof is completed by noting that $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\| \leq \sqrt{r}\|\mathbf{A}\|_2$ for square matrix $\mathbf{A}$, which means that $O_p(\|\mathbf{A}\|_2) = O_p(\|\mathbf{A}\|)$. ■

Proof of Theorem 2.

Similarly to the proof of Theorem 1, assuming $r_n = r$ for all n, we have $r_n = r > r_0$. According to the SVD, $\mathbf{A}_n = \mathbf{V}_n\mathbf{\Sigma}_n\mathbf{U}_n'$, where $\mathbf{V}_n$, $\mathbf{\Sigma}_n$ and $\mathbf{U}_n$ are as in Proof of Theorem 1. Consider the following "regularized" version of $\mathbf{A}_n$: $\widetilde{\mathbf{A}}_n = \mathbf{V}_n\widetilde{\mathbf{\Sigma}}_n\mathbf{U}_n'$, where

$$\widetilde{\Sigma}_n = \begin{bmatrix} \mathbf{\Delta}_{r_0} & \mathbf{0}_{r_0 \times (r-r_0)} \\ \mathbf{0}_{(r-r_0)\times r_0} & \mathbf{0}_{(r-r_0)\times(r-r_0)} \end{bmatrix},$$

and $\mathbf{\Delta}_{r_0} = \text{diag}[\sigma_1(\mathbf{A}_n), ..., \sigma_{r_0}(\mathbf{A}_n)]$. In this notation,

$$\mathbf{A}_n^+ - \mathbf{A}_0^+ = \mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+ - (\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+), \tag{A6}$$

Consider $\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+$. Note that $\text{rank}(\widetilde{\mathbf{A}}_n) = r_0$. According to Theorem 2 (b) of Andrews (1987), if (i) $\|\widetilde{\mathbf{A}}_n - \mathbf{A}_0\| = o_p(1)$ as $n \to \infty$, and (ii) $\|\widetilde{\mathbf{A}}_n^+\| = O_p(1)$, then $\|\widetilde{\mathbf{A}}_n^+ - \mathbf{A}_0^+\| = o_p(1)$. Condition (i) is easy to verify. Indeed, since $\mathbf{A}_n - \mathbf{A}_0 = \mathbf{E}_n$ and $\|\mathbf{E}_n\|_2 = o_p(1)$ under Assumption 2 and $\text{rank}(\widetilde{\mathbf{A}}_n) = r_0$, we have that $\|\widetilde{\mathbf{A}}_n - \mathbf{A}_0\| = o_p(1)$ (see Andrews, 1987, page 355). Consider (ii). As in Proof of Theorem 1, $\widetilde{\mathbf{A}}_n^+ = \mathbf{U}_n\widetilde{\mathbf{\Sigma}}_n^+\mathbf{V}_n'$, where

$$\widetilde{\mathbf{\Sigma}}_n^+ = \begin{bmatrix} \mathbf{\Delta}_{r_0}^{-1} & 0_{r_0 \times (r-r_0)} \\ 0_{(r-r_0)\times r_0} & 0_{(r-r_0)\times(r-r_0)} \end{bmatrix},$$

with $\mathbf{\Delta}_{r_0}^{-1} = \text{diag}[\sigma_1(\mathbf{A}_n)^{-1}, ..., \sigma_{r_0}(\mathbf{A}_n)^{-1}]$. By using this and the orthogonality of $\mathbf{V}_n$ and $\mathbf{U}_n$, we obtain

$$\begin{aligned} \|\widetilde{\mathbf{A}}_n^+\|^2 &= \text{tr}[(\widetilde{\mathbf{A}}_n^+)'\widetilde{\mathbf{A}}_n^+] = \text{tr}(\mathbf{V}_n\widetilde{\mathbf{\Sigma}}_n^+\mathbf{U}_n'\mathbf{U}_n\widetilde{\mathbf{\Sigma}}_n^+\mathbf{V}_n') = \text{tr}(\widetilde{\mathbf{\Sigma}}_n^+\widetilde{\mathbf{\Sigma}}_n^+) \\ &= \sum_{j=1}^{r_0}\sigma_j(\mathbf{A}_n)^{-2} = \sum_{j=1}^{r_0}\lambda_j(\mathbf{A}_n)^{-1} = O_p(1), \end{aligned} \tag{A7}$$

as required for (ii). It follows that $\|\widetilde{\mathbf{A}}_n^+ - \mathbf{A}_0^+\| = o_p(1)$. From (A6), $\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 = \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+ - (\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+)\|_2$, implying

$$\|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 - \|\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+\|_2 \leq \|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 \leq \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 + \|\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+\|_2.$$

But $\|\mathbf{A}_0^+ - \widetilde{\mathbf{A}}_n^+\|_2 = o_p(1)$, and therefore

$$\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 = \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 + o_p(1). \tag{A8}$$

In what remains we evaluate $\|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2$. We begin by noting that

$$\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+ = \mathbf{U}_n(\mathbf{\Sigma}_n^+ - \widetilde{\mathbf{\Sigma}}_n^+)\mathbf{V}_n' = \mathbf{U}_n \begin{bmatrix} \mathbf{0}_{r_0\times r_0} & \mathbf{0}_{r_0\times(r-r_0)} \\ \mathbf{0}_{(r-r_0)\times r_0} & \mathbf{\nabla}_{r-r_0}^{-1} \end{bmatrix} \mathbf{V}_n',$$

where $\mathbf{\nabla}_{r-r_0}^{-1} = \text{diag}[\sigma_{r_0+1}(\mathbf{A}_n)^{-1}, ..., \sigma_r(\mathbf{A}_n)^{-1}]$. Since $\sigma_{r_0+1}(\mathbf{A}_n)^{-1} \leq ... \leq \sigma_r(\mathbf{A}_n)^{-1}$, we obtain

$$\|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 = \sigma_r(\mathbf{A}_n)^{-1}. \tag{A9}$$

By Weyl's theorem, $\sigma_p(\mathbf{A}_0) + \sigma_r(\mathbf{E}_n) \leq \sigma_p(\mathbf{A}_n) \leq \sigma_p(\mathbf{A}_0) + \sigma_1(\mathbf{E}_n)$ for $p = r_0 + 1, ..., r$, which reduces to $\sigma_r(\mathbf{E}_n) \leq \sigma_p(\mathbf{A}_n) \leq \sigma_1(\mathbf{E}_n)$, as $\sigma_{r_0+1}(\mathbf{A}_0) = ... = \sigma_r(\mathbf{A}_0) = 0$. This implies $\sigma_1(\mathbf{E}_n)^{-1} \leq \sigma_p(\mathbf{A}_n)^{-1} \leq \sigma_r(\mathbf{E}_n)^{-1}$, and therefore

$$\sigma_1(\mathbf{E}_n)^{-1} \leq \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 \leq \sigma_r(\mathbf{E}_n)^{-1}.$$

But $\sigma_1(\mathbf{E}_n) = \|\mathbf{E}_n\|_2$ and $\sigma_r(\mathbf{E}_n)^{-1} = \|\mathbf{E}_n^{-1}\|_2$, implying

$$\|\mathbf{E}_n\|_2^{-1} \leq \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 \leq \|\mathbf{E}_n^{-1}\|_2. \tag{A10}$$

We can therefore show that

$$\|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 = O_p(\|\mathbf{E}_n^{-1}\|_2), \tag{A11}$$

and so, by adding the results,

$$\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2 = \|\mathbf{A}_n^+ - \widetilde{\mathbf{A}}_n^+\|_2 + o_p(1) = O_p(\|\mathbf{E}_n^{-1}\|_2). \tag{A12}$$

The proof is made complete by noting that the order of $\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|_2$ is equal to that of $\|\mathbf{A}_n^+ - \mathbf{A}_0^+\|$. ∎

Proof of Lemma 2.

Consider (i). Using (42), we get

$$\begin{aligned}
& T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\hat{\mathbf{F}}^0})\overline{\mathbf{U}} \\
= \;& T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}_m^0(T^{-1}\mathbf{F}'\mathbf{F})^{-1}T^{-1}\overline{\mathbf{U}}_m^{0\prime}\overline{\mathbf{U}} + T^{-1}\mathbf{V}_i\mathbf{P}_{\overline{\mathbf{U}}_m^0}\overline{\mathbf{U}} \\
+ \;& T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}_m^0(T^{-1}\mathbf{F}'\mathbf{F})^{-1}T^{-1}\mathbf{F}'\overline{\mathbf{U}} + T^{-1}\mathbf{V}_i'\mathbf{F}(T^{-1}\mathbf{F}'\mathbf{F})^{-1}T^{-1}\overline{\mathbf{U}}_m^{0\prime}\overline{\mathbf{U}} \\
+ \;& T^{-1}\mathbf{V}_i'\mathbf{F}^0[(T^{-1}\hat{\mathbf{F}}^{0\prime}\hat{\mathbf{F}}^{0\prime})^+ - \boldsymbol{\Sigma}_{\mathbf{f}^0}^+]T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}. \qquad \text{(A13)}
\end{aligned}$$

Now recall that $\overline{\mathbf{U}}_m^0 = \overline{\mathbf{U}}\mathbf{B}_m$. Furthermore, $\|\mathbf{V}_i\mathbf{P}_{\overline{\mathbf{U}}_m^0}\overline{\mathbf{U}}\| \leq \|\mathbf{V}_i'\overline{\mathbf{U}}\|$. These results can be used together with $\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\| = O_p(T^{-1/2})$ and $\|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ by (A.12) and (A.13) in Lemma 2 of Pesaran (2006) to obtain

$$\begin{aligned}
& \|T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\| \\
\leq \;& \|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|^2\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\| + T^{-1}\mathbf{V}_i\mathbf{P}_{\overline{\mathbf{U}}_m^0}\overline{\mathbf{U}}\| \\
+ \;& T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\| + \|T^{-1}\mathbf{V}_i'\mathbf{F}\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|\mathbf{B}_m\|\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\| \\
+ \;& \|T^{-1}\mathbf{V}_i'\mathbf{F}^0\|\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \boldsymbol{\Sigma}_{\mathbf{f}^0}^+\|\|T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}\| \\
= \;& O_p(N^{-1}) + O_p((NT)^{-1/2}). \qquad \text{(A14)}
\end{aligned}$$

In view of (43) and (39) we get

$$\begin{aligned}
\mathbf{R}_{2NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\
&= \sqrt{T}N^{-1/2}\sum_{i=1}^N T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\
&- \sqrt{T}N^{-1/2}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\
&= \sqrt{T}N^{-1/2}[\overline{\mathbf{d}}_1 + O_p(N^{-1}) + O_p(T^{-1/2}) + O_p(\sqrt{N}T^{-3/2})], \qquad \text{(A15)}
\end{aligned}$$

where

$$\overline{\mathbf{d}}_1 = T^{-1}\sum_{i=1}^N \mathbf{V}_i'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - T^{-1}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i. \qquad \text{(A16)}$$

Note that $N/T\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}$ is $O_p(1)$ as a direct consequence of Lemma 2 in Pesaran (2006). Additionally, it can be shown that even $T^{-1}\sum_{i=1}^N \mathbf{V}_i'\mathbf{P}_{\overline{\mathbf{U}}_{-m}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i$ is $O_p(1)$. Thus, provided that $T/N = O(1)$, $\mathbf{R}_{2NT}$ does not converge to zero if $r < m+1$.

As for (ii), by using again $\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\| = O_p(T^{-1/2})$, $\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\| = O_p((NT)^{-1/2})$ and $\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\| = O_p(1)$, we obtain

$$\begin{aligned}
\|T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\| &\leq \|T^{-1}\mathbf{V}_i'\mathbf{F}^0\|\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\|\|T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}\| \\
&= O_p(T^{-1/2})O_p((NT)^{-1/2}) = O_p(N^{-1/2}T^{-1}), \\
\|T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\| &\leq \|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|^2\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\| = O_p((NT)^{-1}).
\end{aligned}$$

Hence,

$$\begin{aligned}
\|\mathbf{R}_{1NT}\| &\leq \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i\right\| + \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i\right\| \\
&\leq \sqrt{T}N^{-1/2}\sum_{i=1}^{N}\|T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i\| + \sqrt{T}N^{-1/2}\sum_{i=1}^{N}\|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i\| \\
&= \sqrt{T}N^{-1/2}O_p(\sqrt{N}T^{-1}) + \sqrt{T}N^{-1/2}O_p(T^{-1}) = O_p(T^{-1/2}). \qquad \text{(A17)}
\end{aligned}$$

Consider (iii), where

$$\mathbf{R}_{0NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i. \qquad \text{(A18)}$$

Making use of (31) and assuming that $T/N = O(1)$, the second term on the right simplifies to

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i &= \sqrt{T}N^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}NT^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i \\
&= \sqrt{T}N^{-1/2}\overline{\mathbf{b}}_2 + O_p(T^{-1/2}), \qquad \text{(A19)}
\end{aligned}$$

where

$$\overline{\mathbf{b}}_2 = \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\mathbf{\Sigma}_{\mathbf{u}}\overline{\mathbf{C}}^{+}\gamma_i.$$

For the first term on the right of (A18), we use

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\gamma_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{j=1}^{N}\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta} + \frac{1}{N}\sum_{j=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_j, \frac{1}{N}\sum_{j=1}^{N}\mathbf{V}_i'\mathbf{V}_j\right]\overline{\mathbf{C}}^{+}\gamma_i \\
&= \sqrt{T}N^{-1/2}\frac{1}{N}\sum_{i=1}^{N}[T^{-1}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\beta} + T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i, T^{-1}\mathbf{V}_i'\mathbf{V}_i]\overline{\mathbf{C}}^{+}\gamma_i \\
&+ \frac{1}{N^{3/2}\sqrt{T}}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta} + \mathbf{V}_i'\boldsymbol{\varepsilon}_j, \mathbf{V}_i'\mathbf{V}_j]\overline{\mathbf{C}}^{+}\gamma_i. \qquad \text{(A20)}
\end{aligned}$$

By using the fact that the fourth moments of $\mathbf{v}_{i,t}$ and $\varepsilon_{i,t}$ are bounded, we can show that $E(\|T^{-1/2}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\|^2) = O(1)$. By using this, $T^{-1}\mathbf{V}_i'\mathbf{V}_i = \mathbf{\Sigma} + O_p(T^{-1/2})$ (see (31)) and $T/N = O(1)$, we obtain

$$\sqrt{T}N^{-1/2}\frac{1}{N}\sum_{i=1}^{N}[T^{-1}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\beta} + T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i, T^{-1}\mathbf{V}_i'\mathbf{V}_i]\overline{\mathbf{C}}^{+}\gamma_i = \sqrt{T}N^{-1/2}\overline{\mathbf{b}}_1 + O_p(T^{-1/2}), \qquad \text{(A21)}$$

where

$$\overline{\mathbf{b}}_1 = \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}[\boldsymbol{\beta}, \mathbf{I}_k]\overline{\mathbf{C}}^+\gamma_i.$$

As for the second term on the right-hand side of (A20), tedious but straightforward calculations reveal that $E(\|N^{-1}T^{-1/2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_j\overline{\mathbf{C}}^+\gamma_i\|^2)$, $E(\|N^{-1}T^{-1/2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{V}_i'\mathbf{V}_j\overline{\mathbf{C}}^+\gamma_i\|^2)$ and $E(\|N^{-1}T^{-1/2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}\overline{\mathbf{C}}^+\gamma_i\|^2)$ are all $O(1)$. Hence,

$$\left\|\frac{1}{N^{3/2}\sqrt{T}}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta} + \mathbf{V}_i'\boldsymbol{\varepsilon}_j, \mathbf{V}_i'\mathbf{V}_j]\overline{\mathbf{C}}^+\gamma_i\right\| = O_p(N^{-1/2}), \tag{A22}$$

implying that

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i &= \sqrt{T}N^{-1/2}\frac{1}{N}\sum_{i=1}^{N}[T^{-1}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\beta} + T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i, T^{-1}\mathbf{V}_i'\mathbf{V}_i]\overline{\mathbf{C}}^+\gamma_i + O_p(N^{-1/2}) \\ &= \sqrt{T}N^{-1/2}\overline{\mathbf{b}}_1 + O_p(N^{-1/2}) + O_p(T^{-1/2}). \qquad \text{(A23)}\end{aligned}$$

The results in (A19) and (A23) imply that (A18) reduces to

$$\begin{aligned}\mathbf{R}_{0NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\gamma_i \\ &= \sqrt{T}N^{-1/2}(\overline{\mathbf{b}}_1 - \overline{\mathbf{b}}_2) + O_p(N^{-1/2}) + O_p(T^{-1/2}). \qquad \text{(A24)}\end{aligned}$$

This establishes (iii), and hence the proof of the lemma is complete. ∎

Proof of Lemma 3.

We begin by establishing (i). Note that $T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i$ can be decomposed in the same fashion as (42) and (A13). According to (A.13) in Lemma 2 of Pesaran (2006), $\|T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$, a result that holds for $\|T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\boldsymbol{\varepsilon}_i\|$ as well. Furthermore,it is easy to show that $\|T^{-1}(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i\| = O_p(T^{-1/2})$. By using this, Lemma 1, (37), and the known orders of $\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|$, $\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|$, $\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\|$ and $\|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\|$, we obtain

$$\begin{aligned}&\|T^{-1}\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i - T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\boldsymbol{\varepsilon}_i\| \\ &= \|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|^2\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i\| + \|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\mathbf{F}'\boldsymbol{\varepsilon}_i\| \\ &+ \|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|\mathbf{B}_m\|\|T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i\| \\ &+ \|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \mathbf{\Sigma}^+_{\mathbf{f}^0}\|\|T^{-1}(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i\| \\ &= O_p(N^{-2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-3/2}), \qquad \text{(A25)}\end{aligned}$$

and

$$
\begin{aligned}
& \|T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i - T^{-1}\overline{\mathbf{V}}_i'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\boldsymbol{\varepsilon}_i\| \\
= \;& \|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|^2\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i\| + \|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\|\|\mathbf{B}_m\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|T^{-1}\mathbf{F}'\boldsymbol{\varepsilon}_i\| \\
+ \;& \|T^{-1}\mathbf{V}_i'\mathbf{F}\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|\mathbf{B}_m\|\|T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i\| \\
+ \;& \|T^{-1}\mathbf{V}_i'\mathbf{F}^0\|\|[T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0]^+ - \boldsymbol{\Sigma}_{\mathbf{f}^0}^+\|\|T^{-1}(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i\| \\
= \;& O_p(N^{-2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}). \qquad \text{(A26)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbf{Q}_{2NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i \\
&= \sqrt{T}N^{-1/2}\sum_{i=1}^N T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i \\
&- \sqrt{T}N^{-1/2}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i \\
&= \sqrt{T}N^{-1/2}[\overline{\mathbf{d}}_2 + O_p(N^{-1}) + O_p(T^{-1/2}) + O_p(\sqrt{N}T^{-1})], \qquad \text{(A27)}
\end{aligned}
$$

where

$$
\overline{\mathbf{d}}_2 = T^{-1}\sum_{i=1}^N \mathbf{V}_i'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\boldsymbol{\varepsilon}_i - T^{-1}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\overline{\mathbf{U}}^0_{-m}}\boldsymbol{\varepsilon}_i. \qquad \text{(A28)}
$$

Again, given that $T/N = O(1)$, this expression only converges to zero if $r = m+1$.

Next, consider (ii). We begin by noting how

$$
\begin{aligned}
\|T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\| &\leq \|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\|\|T^{-1}(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i\| \\
&= O_p((NT)^{-1/2})O_p(T^{-1/2}) = O_p(N^{-1/2}T^{-1}),
\end{aligned}
$$

implying

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\| &\leq \sqrt{T}N^{-1/2}\sum_{i=1}^N \|T^{-1}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\| \\
&= \sqrt{T}N^{-1/2}O_p(\sqrt{N}T^{-1}) = O_p(T^{-1/2}). \qquad \text{(A29)}
\end{aligned}
$$

Since $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') = \sigma^2\mathbf{I}_T$ and

$$
\|T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\| \leq \|T^{-1}\mathbf{V}_i'\mathbf{F}^0\|^2\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\| = O_p(T^{-1}),
$$

we also have

$$E\left(\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\|^2\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{tr}[E(\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_j)]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}[E(T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i)] = O(T^{-1}). \quad \text{(A30)}$$

We can therefore show that

$$\|\mathbf{Q}_{1NT}\| \leq \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\| + \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\|$$

$$\leq \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\| + \sqrt{T}N^{-1/2}\sum_{i=1}^{N}\|T^{-1}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\|$$

$$= O_p(T^{-1}) + O_p(T^{-1/2}) = O_p(T^{-1/2}). \quad \text{(A31)}$$

Finally, consider (iii). Note how

$$\frac{1}{T}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{U}_j'\boldsymbol{\varepsilon}_i$$

$$= \begin{bmatrix} (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\beta}'\mathbf{V}_j'\boldsymbol{\varepsilon}_i + (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_i \\ (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{V}_j'\boldsymbol{\varepsilon}_i \end{bmatrix}. \quad \text{(A32)}$$

By applying the same argument used earlier for showing $\|N^{-1}T^{-1/2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\| = O_p(1)$, we obtain $\|(NT)^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_i\| = O_p(T^{-1/2})$. This implies

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_i = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_i$$

$$= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + O_p(T^{-1/2})$$

$$= \sigma^2\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+ + O_p(T^{-1/2}),$$

where the last result is due to $T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i = \sigma^2 + O_p(T^{-1/2})$, which holds because $TE[(T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i - \sigma^2)^2] = O_p(1)$. But $\|(NT)^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{V}_j'\boldsymbol{\varepsilon}_i\|$ and $\|(NT)^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\boldsymbol{\beta}'\mathbf{V}_j'\boldsymbol{\varepsilon}_i\|$ are $O_p(T^{-1/2})$ too, and therefore

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{V}_j'\boldsymbol{\varepsilon}_i\right\| \leq \left\|\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right\| + \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}_i(\overline{\mathbf{C}}')^+\mathbf{V}_j'\boldsymbol{\varepsilon}_i\right\|$$

$$= O_p(\sqrt{NT}) + O_p(T^{-1/2}) = O_p(T^{-1/2}),$$

with $\|(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\boldsymbol{\beta}'\mathbf{V}_j'\boldsymbol{\varepsilon}_i\|$ being of the same order. This means that

$$\frac{1}{T}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\sigma^2\begin{bmatrix}1\\ \mathbf{0}_{k\times 1}\end{bmatrix} + O_p(T^{-1/2}). \tag{A33}$$

Therefore,

$$\begin{aligned}\mathbf{Q}_{0NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i - \sqrt{T}N^{-1/2}\frac{1}{T}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i - \sqrt{T}N^{-1/2}\overline{\mathbf{b}}_3 + O_p(N^{-1/2}) + O_p(T^{-1/2}), \qquad \text{(A34)}\end{aligned}$$

where

$$\overline{\mathbf{b}}_3 = \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\overline{\mathbf{C}}')^{+}\sigma^2[1,\mathbf{0}_{k\times 1}']'.$$

■

Proof of Lemma 4.

Direct substitution from (38) gives

$$\begin{aligned}T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\mathbf{X}_i \\ &= T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^{+})'\overline{\mathbf{U}}'\mathbf{M}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\mathbf{\Gamma}_i' - T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^{+})'\overline{\mathbf{U}}'\mathbf{M}_{\widehat{\mathbf{F}}^0}\mathbf{V}_i \\ &- T^{-1}\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\mathbf{\Gamma}_i' + T^{-1}\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\mathbf{V}_i. \qquad \text{(A35)}\end{aligned}$$

Consider

$$\begin{aligned}&\|T^{-1}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{X}_i\| \\ &= \|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^{+})'\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\overline{\mathbf{U}}\overline{\mathbf{C}}^{+}\mathbf{\Gamma}_i'\| + 2\|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^{+})'\overline{\mathbf{U}}'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{V}_i\| \\ &+ \|T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{V}_i\|. \qquad \text{(A36)}\end{aligned}$$

All terms here are known from before, except $\|T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{V}_i\|$, which is of the same order as $\|T^{-1}\mathbf{V}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\boldsymbol{\varepsilon}_i\|$. Insertion and simplification yields

$$\|T^{-1}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{X}_i\| = O_p(N^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}). \tag{A37}$$

Note that the terms that were explicitly stated in Lemmas 2 and 3 are of no importance here since the expression we consider is multiplied by the factor $(NT)^{-1}$. This causes the projections on the space spanned by $\overline{\mathbf{U}}_{-m}^{0}$, which previously remained as a bias term, to be $o_p(1)$ for all N, T. By further use of the known orders of $\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|$, $\|T^{-1}(\mathbf{F}^0)'\mathbf{V}_i\|$ and $\|[T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{-1}\|$, we can show that $\|T^{-1}\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\| = O_p((NT)^{-1})$, $\|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\| = O_p(N^{-1/2}T^{-1})$ and $\|T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\| = O_p(T^{-1})$, which in turn implies

$$\begin{aligned}\|T^{-1}\mathbf{X}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{X}_i\| &= \|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i'\| + 2\|T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\| \\ &+ \|T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\| = O_p(T^{-1}). \qquad \text{(A38)}\end{aligned}$$

By using these results and

$$\begin{aligned}T^{-1}\mathbf{X}_i'\mathbf{X}_i &= T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' - T^{-1}\mathbf{\Gamma}_i(\overline{\mathbf{C}}^+)'\overline{\mathbf{U}}'\mathbf{V}_i - T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^+\mathbf{\Gamma}_i' + T^{-1}\mathbf{V}_i'\mathbf{V}_i \\ &= T^{-1}\mathbf{V}_i'\mathbf{V}_i + O_p(N^{-1}) + O_p((NT)^{-1/2}) \\ &= \mathbf{\Sigma} + O_p(N^{-1}) + O_p(T^{-1/2}), \qquad \text{(A39)}\end{aligned}$$

we obtain

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i &= \frac{1}{N}\sum_{i=1}^{N}T^{-1}\mathbf{X}_i'\mathbf{X}_i - \frac{1}{N}\sum_{i=1}^{N}T^{-1}\mathbf{X}_i'\mathbf{P}_{\mathbf{F}^0}\mathbf{X}_i - \frac{1}{N}\sum_{i=1}^{N}T^{-1}\mathbf{X}_i'(\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0})\mathbf{X}_i \\ &= \mathbf{\Sigma} + O_p(N^{-1}) + O_p(T^{-1/2}). \qquad \text{(A40)}\end{aligned}$$

■